Control Channel Design for Many-Antenna MU-MIMO

Clayton Shepard, Abeer Javed, and Lin Zhong
Department of Electrical and Computer Engineering
Rice University, Houston, TX
{cws, abeer.javed, lzhong}@rice.edu

ABSTRACT

Many-antenna MU-MIMO faces a critical, previously unaddressed challenge: it lacks a practical control channel. At the heart of this challenge is that the potential range of MU-MIMO beamforming systems scales with up to the square of the number of base-station antennas once they have channel state information (CSI), whereas the range of traditional control channel operations remains constant since they take place before or during CSI acquisition. This range gap between no-CSI and CSI modes presents a critical challenge to the efficiency and feasibility of many-antenna base stations, as their operational range is limited to the no-CSI mode.

We present a novel control channel design for many-antenna MU-MIMO, *Faros*, that allows the number of base-station antennas to scale up to 100s in practice. *Faros* leverages a combination of open-loop beamforming and coding gains to bridge the range gap between the CSI and no-CSI modes. Not only does *Faros* provide an elegant and efficient control channel for many-antenna MU-MIMO, but on a more fundamental level it exposes flexible, fine-grained, control over space, time, and code resources, which enables previously impossible optimizations. We implement our design on the Argos many-antenna base station and evaluate its performance in bridging the range gap, synchronization, and paging. With 108 antennas, *Faros* can provide over 40 dB of gain, which enables it to function reliably at over 250 meters outdoors with less than 100 μW of transmit power per antenna, 10 mW total, at 2.4 GHz.

1. INTRODUCTION

Many-antenna MU-MIMO is a rapidly growing research field, which has recently shown promise of commercialization [1, 2]. However, there are still many system challenges facing the creation of practical many-antenna base stations. Perhaps the most critical issue is the lack of an efficient and reliable control channel in current architectures. This channel is required for basic network operations such as time-frequency synchronization, association, channel state information (CSI) collection, random access, and paging, which take place *before* a MIMO channel is established. Today, wireless systems realize the control channel using a single high-power antenna, or simple diversity schemes, but these methods rapidly become very inefficient as the number of base-station antennas (M) increases.

MobiCom'15, September 7–11, 2015, Paris, France.
ACM ISBN 978-1-4503-3543-0/15/09.
DOI: http://dx.doi.org/10.1145/2789168.2790120.

All MIMO base stations have two modes: the *no-CSI mode* that takes place before the base station knows the CSI for the active users, and the *CSI mode* that provides a much more efficient MIMO channel. In order for the base station to collect CSI, it must establish time-frequency synchronization with the users and receive uplink pilots from them; furthermore, once a user becomes inactive, the base station must be able to notify the user of an incoming transmission, i.e. page the user, prompting it to send a pilot. All of these operations are part of the control channel, which is traditionally sent entirely over the no-CSI mode.

In MIMO systems the CSI mode has a gain of up to M^2 higher than the no-CSI mode (see §3). When M is small, as in current systems, one can easily overcome this gain gap by simply using a lower modulation rate or a coding gain in the no-CSI mode. However, as M increases, this gap quickly becomes large and problematic. In existing systems *all* control channel operations are performed in the no-CSI mode and sent omnidirectionally to the entire coverage area. Thus, the base station's operational range is limited by the no-CSI mode, which is significantly shorter than that of the CSI mode. One naïve solution is to use higher transmission power in the no-CSI mode. This will lead to more expensive hardware, e.g., power amplifier, possible violation of FCC regulations, and increased inter-cell interference.

We present *Faros*,[1] a novel control channel design that addresses the above gain gap for base stations or access points with multiple antennas. *Faros* leverages two key insights. (i) First, *as much of the control channel as possible should be sent over the CSI mode.* We find that the only control channel operations that must use the no-CSI mode are time-frequency synchronization, association, CSI collection, paging, and random access, which are required to establish the CSI mode. By implementing the remaining control channel operations over the CSI mode, we substantially increase their efficiency, as well as avoid the aforementioned gain gap. (ii) Our second key insight is that *synchronization and association are not time-critical.* That is, synchronization is valid for 100s of ms and association only happens once; thus by reducing the frequency of synchronization *Faros* is able to substantially reduce the channel overhead of these operations in the no-CSI mode, at the cost of slightly increased association latency at the cell edges.

Guided by these insights, *Faros* leverages open-loop beamforming and coding gains to ensure that many-antenna base stations can achieve their full potential range (see §4 and §5). Through open-loop beamforming, *Faros* is able to use the full diversity, power, and beamforming gains from all of the antennas on the base station, which enables it to

[1] $\Phi\acute{\alpha}\rho o\sigma$, or *Faros*, means "beacon" or "lighthouse" in Greek. The rotation of a lighthouse's strong beam of light is analogous to the beamsweeping employed by *Faros*.

scale with M, the number of base-station antennas. Because open-loop beamforming is never as performant as its MU-MIMO counterpart, closed-loop beamforming, *Faros* employs coding gains to further increase the range and to ensure that synchronization and paging are reliable even at the cell edges. To be as efficient as possible, *Faros* only performs these essential tasks and communication outside of the CSI mode, which offers much higher spectral capacity. Specifically, *Faros* uses open-loop beamforming to sweep extra-long synchronization sequences across the coverage area. This synchronization sequence not only enables users to establish time-frequency synchronization with the base station, but also encodes the base-station ID, and optionally user IDs for paging. *Faros* can dynamically configure important parameters, such as the beam patterns, sweep rate, and sequence length, to match the required gain for full coverage of the desired area. Furthermore, by increasing open-loop beamforming and coding gains in no-CSI mode while reducing the modulation rate or number of users served in CSI mode, *Faros* can be used to extend the range of the base station in remote areas.

We implement *Faros* on Argos, a many-antenna MU-MIMO base station over a 2.4 GHz channel, with 108 antennas (see §6) and evaluate the real-world performance and overhead of the implementation (see §7). Measurements show that our implementation provides over a 40 dB gain compared to traditional control channel operations. Anecdotally, this enables us to provide reliable synchronization to mobile users at over 250 meters with less than 100 μW of power per base-station antenna, or 10 mW of total power, using only standard low-gain 3 dBi omnidirectional antennas. Our design facilitates collecting high resolution channel measurements in highly mobile environments, with less than 0.5% channel overhead. To reduce the overhead of paging delay, we additionally implement a simple paging scheme that leverages the users last known location for directing the paging signal, which reduces paging delay by 400%.

In designing and implementing *Faros*, we do not invent any new physical layer techniques. Rather, *Faros* contributes a novel synthesis of known methods, such as beamforming, coding, and synchronization, to achieve a very practical and flexible control channel that bridges the gain gap with extremely low overhead. To the best of our knowledge, *Faros* is the first reported control channel design for many-antenna MU-MIMO that can effectively bridge the gain gap.

2. BACKGROUND

2.1 Beamforming and MU-MIMO

Beamforming utilizes multiple antennas transmitting at the same frequency to realize directional transmission. Constructive and destructive interference of the signals from multiple antennas causes the signal strength received to vary spatially, leading to a *beam pattern*. This beam pattern can be altered by changing the beamforming *weights* applied to each antenna, effectively altering the amplitude and phase of the signal sent from that antenna. *Open-loop* beamforming uses precomputed beamforming weights (beamweights), such as DFT weights [3], to steer the beam in a desired spatial direction, without knowledge of the users' locations. *Closed-loop* or *adaptive* beamforming employs channel state information (CSI) to calculate the beamweights that maximize the signal strength at intended users and minimize the interference at unintended ones.

Multi-user multiple-input multiple-output (MU-MIMO) base stations leverage multiple antennas, each with its own radio, to serve multiple users simultaneously on the same time-frequency-code resource, typically through closed-loop beamforming. For simplicity, we use the term antenna to include both the radio and antenna. It is well-known that the spectral and energy efficiency of MU-MIMO systems grow with the number of base-station antennas (M) and the number of concurrent users (K), given $M \geq K$.

Many-antenna MU-MIMO: In light of this, several strong theoretical analyses have advocated a very large number of base-station antennas [4–6], commonly referred as *massive* MIMO and widely considered one of the few candidate technologies for 5G cellular networks [1, 7, 8]. We use the term *many-antenna* to refer to base stations that have many more antennas than users, but are not necessarily "massive". There have been a number of real-world many-antenna prototypes recently reported, including [2, 9–14], as well as efforts towards commercialization and standardization [1, 15]. The succinct background of many-antenna MU-MIMO relevant to this work is: (*i*) Efficient massive channel estimation requires uplink pilots that are used to infer the downlink CSI via TDD reciprocity. (*ii*) Since channel estimates are only ephemerally accurate, downlink beamforming must happen immediately after channel estimation. As a result, an efficient many-antenna MU-MIMO transmission frame structure needs four parts, as depicted by Figure 2.

2.2 Control Channel

In wireless systems, the control channel performs operations required to setup data communication. This includes synchronization, gain control, association, timing advance, random access, paging, setting modulation rates, gain control, scheduling and more. Additionally, in MIMO systems, the control channel must coordinate the collection of CSI across many antennas from multiple users efficiently. This paper focus on the control channel operations required to establish the MIMO channel, which are *synchronization*, *association*, *CSI collection*, *random access*, and *paging*, as *Faros* performs the remaining control channel operations over the more efficient MIMO channel using existing techniques.

Synchronization: Since nodes in wireless networks do not share oscillators, their time-frequency reference is subject to drift. Thus all high-performance digital wireless communication schemes require tight time-frequency synchronization. In existing systems users establish time-frequency synchronization in four steps: (*i*) First, they auto-correlate the received signal with itself for frame detection and coarse timing. (*ii*) Then, they perform automatic gain control (AGC) to ensure the received signal is within their ADC's dynamic range. (*iii*) Next, they perform a cross-correlation with a pre-known sequence to achieve fine-grained time synchronization. (*iv*) Finally, they leverage the distortion within the known signal, i.e. phase shift, to recover the frequency offset and establish frequency synchronization.

For example, in 802.11 the user continuously performs an auto-correlation to detect the short training sequence (STS) at the start of a packet, which triggers AGC, then performs a cross-correlation on the following LTS for time synchronization. Similarly, in LTE, the user continuously performs an auto-correlation to detect the cyclic prefix of each sym-

bol, then performs a cross-correlation on the PSS and SSS for time synchronization. Typically reference symbols are transmitted throughout the frame in order to maintain this synchronization, as well as compensate for other channel effects. For example, 802.11 dedicates four subcarriers to pilots, and LTE sends reference symbols in a checkerboard-like pattern that are close enough together in time and frequency to continuously correct for drift.

Association: Before a user can transmit or receive data, it must first identify the nearby base stations, select one, then connect to it. To facilitate this association procedure, base stations typically transmit a unique identifier, often called a *beacon*, at a regular interval. Users scan for base stations, often over multiple frequencies, then choose one to associate with based on specific criteria, such as signal strength and authorization. The user then contacts the base station, usually leveraging the same mechanism as random access, to request and coordinate access, e.g., authorization, encryption, and scheduling.

CSI Collection: To obtain CSI, the transmitter sends a pre-known sequence, called a *pilot*, which the receiver uses to compute this amplitude and phase shift for each subcarrier. However, this requires time-frequency synchronization, as without time synchronization the receiver would not reliably know where the pilot starts, and without frequency synchronization there would be inter-subcarrier interference that causes inaccurate channel estimation.

Traditional MU-MIMO systems employ explicit CSI estimation: the base station sends pilots from each of its antennas, the users estimate the CSI to each antenna, then send this CSI estimation back to the base-station. In CSMA systems, such as 802.11, this CSI collection is performed at the beginning of every frame, whereas in scheduled systems, such as LTE, this is performed continuously using reference symbols from each base-station antenna. These techniques do not scale well as the number of antennas and users increase, thus emerging many-antenna systems typically employ implicit CSI estimation: each user sends an uplink pilot which the base station receives on every antenna, which provides uplink CSI, then leverages reciprocal calibration to estimate the downlink CSI based on the uplink CSI [9, 16–19].

Paging and Random Access: Additionally, the control channel handles notifying users when they have incoming data, called *paging*, and coordinating users to randomly access the network when they have outgoing data. Both of these operations must take place before CSI is acquired, as the user has to be paged in order to know it needs to send pilots, or, for random access, it must be able to notify the base station that it has outgoing data so the base station knows to estimate the channel.

3. GAIN GAP EXPLAINED

Multi-antenna base stations operate in two modes: either with CSI or without CSI. With CSI the base station can achieve a gain of M^2 relative to the *peak-power of a single antenna*, whereas without CSI the base station only has a gain of 1 for some control channel operations, illustrated in Figure 1. Furthermore, while the channel is reciprocal for uplink and downlink transmissions, the transceiver hardware is not, which subsequently creates a second gain gap between uplink and downlink modes. In this section, we take a closer look at these gain gaps, taking in to account real-world constraints and hardware.

Mode	CSI	no-CSI	**Gap**
Uplink	$M \times P_U$	P_U	$\boldsymbol{M}$
Downlink	$M^2 \times P_{BS}/K$	P_{BS}	$\boldsymbol{M^2/K}$
Gap	$K \cdot P_U/M/P_{BS}$	P_U/P_{BS}	

Table 1: Gain gaps between no-CSI and CSI modes. M is the number of base-station antennas; K the number of concurrently served users; P_U the transmission power of a user antenna; P_{BS} that of a base-station antenna.

Table 1 summarizes the analytical results for all modes of operation: no-CSI vs. CSI and downlink vs. uplink assuming an M antenna base station serving K single-antenna users. Each base-station antenna has a transmit power of P_{BS} and each user antenna has a transmit power of P_U. While many theoretical analyses use a total transmit power budget, real systems are constrained by a peak transmit power *per antenna*. For simplicity we assume the average channel and antenna gains are normalized to 1, since they are constant across all modes, and include any non-reciprocal hardware effects, such as the gains from the low-noise amplifiers (LNAs) in the appropriate P, e.g., P_U includes the gain from the base station's LNAs.

We note the above M^2 gain gap is a point of contention, particularly among theoreticians, as they typically assume a total power budget, which reduces this gain to M. In a real system antennas are peak-power constrained so this would require a single antenna to be provisioned with a much higher total transmit power, which can be impractical. Regardless, there is still at least a gain gap of M, and we analyze both situations in our experiments in §7.

3.1 Without CSI

To the best of our knowledge there is no existing scheme which performs better than a single antenna for the no-CSI mode control channel operations of synchronization and channel estimation. Thus the no-CSI mode has a gain of 1, which becomes P_{BS} and P_U for downlink and uplink, respectively, as shown in Table 1. The gain of an M antenna base station in its no-CSI mode is dependent on what operation it is performing. For CSI collection, there is a fundamental gain limitation of 1 because CSI consists of only information about the link between one antenna and another antenna. Therefore, signals received at other antennas do not contain information about that link's CSI. On the other hand, this theoretical limitation doesn't exist for synchronization, as the desired signal can be sent from all the base-station antennas, which is exploited in our design.

While there are no-CSI mode techniques which achieve a theoretic gain of M, these methods are either impractical, or, in fact, reduce the performance of time-frequency synchronization. One naïve technique would be to use an RF combiner to merge the power output of the M base-station antennas to a single antenna. Not only is this difficult and expensive to implement in hardware, as it requires perfect phase matching to avoid feedback in to the antennas, and complex wiring, but it also loses the diversity gain of the M antennas; in essence this is just using a single high-power transmitter, i.e., it is no longer an $M \times K$ system. Despite these drawbacks, we include this scheme for comparison in our experimental analysis. Another method, which is currently used in multi-antenna systems, such as 802.11n and 802.11ac, is cyclic delay diversity (CDD), which cyclicly rotates the symbols by different amounts of time from each antenna [20]. CDD spreads the power output of all M antennas

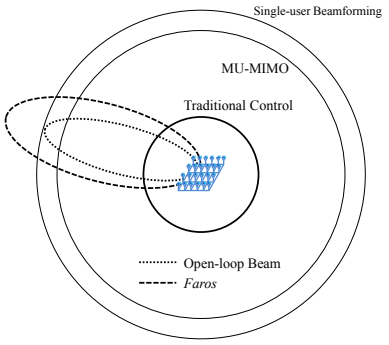


Figure 1: The downlink gain gap. Note that while the figure depicts omnidirectionality, the gap is equivalent for directional antennas.

spatially, and can be thought of as arbitrarily beamforming on different subcarriers. This causes time-domain distortion, which substantially degrades the performance of existing synchronization techniques, and, even worse, this performance degrades rapidly as more antennas are added [20]. Finally, both of these naïve schemes only help in the downlink, and do not provide any gain in the uplink.

3.2 With CSI

The potential power gain of an $M \times K$ MU-MIMO system with CSI, in both uplink and downlink, is well known to be $P \cdot M$, where P is the transmission power [21]. Leveraging CSI, the base station can direct radiation towards, or listen to radiation from, the intended K users using beams with an approximate width of $1/M$, which provides a spatial power gain of M. In the downlink, the base station transmits power from all M antennas, but has to split the power among K users, thus providing a per-link power of $P_{BS} \cdot M/K$, assuming equal power allocation among the users. In the uplink, the base station receives power from each user on all M antennas, thus providing a power of P_U. This renders a total gain of $M^2 \cdot P_{BS}/K$ and $M \cdot P_U$, respectively, as shown in Table 1. Note that a MU-MIMO base station capable of serving K users likely will not always serve K users simultaneously; with a single user the gap increases to a full M^2.

4. FAROS GAIN MATCHING

With the gain gaps above, we next present the design of *Faros* in two parts: (*i*) mechanisms to bridge the gain gaps (this section), and (*ii*) the control channel system design that overcomes the limitations of these mechanisms (§5).

To bridge the gain gap of the no-CSI and CSI modes in the downlink, *Faros* combines open-loop beamforming with a coding gain. It sweeps open-loop beams carrying orthogonal sequences, which enable the *synchronization* and *paging* operations. In the uplink, *Faros* exploits the natural per-antenna asymmetric transmit power and employs an additional coding gain to enable *CSI collection* and *random access* operations. By encoding a base-station ID in the downlink synchronization sequence and exploiting the random access operation, *Faros* facilitates the *association* operation.

4.1 Open-Loop Beamforming

Faros employs open-loop beamforming to exploit the power and diversity of all antennas on the base station. The combined power of the antennas provides a gain of M, and the beamforming provides another gain of M, for a total gain of $O(M^2)$. However, this beamforming gain does not come for free, as it focuses the radiated power on $1/M$ of the antennas' coverage area, thus *Faros* must sweep beams to provide complete coverage. Leveraging our key insight that *association* and *synchronization* are delay-tolerant, *Faros* employs open-loop beamforming for these operations *without* impacting user-perceived performance or creating significant channel overhead.

While there are many MIMO and diversity schemes that exploit the gains from multiple antennas, only open-loop beamforming is effective for time-frequency synchronization, as it provides the full potential combined power and directivity gain from all of the available antennas without causing time-domain distortion. Furthermore, open-loop beamforming has four practical benefits in a real-world MU-MIMO system: (*i*) the increased received power allows the user to employ cheaper RF components, e.g., the LNA, (*ii*) the increased directivity and lower total power reduce the interference to adjacent cells, (*iii*) it does not require any additional hardware or computation, as the beamforming precoders are already required on the base station for MU-MIMO, and (*iv*) it allows the coverage area to be finely tuned.

4.1.1 Beamsweeping

To overcome the spatial selectivity of open-loop beamforming, *Faros* employs *beamsweeping* that transmits a signal, s, in different spatial directions using beamforming. Fundamentally, beamsweeping trades off increased spatial coverage with additional time overhead. Guided by our second key insight that some control channel operations are delay-tolerant, we leverage beamsweeping for synchronization, and to help facilitate association.

Each beam is defined by a $M \times 1$ vector, $\mathbf{b_n}$, thus an N length sweep pattern can be defined by a $M \times N$ matrix, $\mathbf{B}$, composed of $\mathbf{b_1}, \mathbf{b_2}, ..., \mathbf{b_n}$. The M-antenna base station transmits an entire sweep pattern in N time-slots, as the transmission in a given time-slot n and given base station antenna m is simply: $s \cdot B_{m,n}$. Thus, if each beam is sent contiguously, then beamsweeping takes N times longer than a single omnidirectional transmission of the same sequence. Because *Faros* sends a beam at the beginning of each frame, an entire beamsweep takes $N \cdot F$, where F is the frame duration, as further described in §5.6, and shown in Table 2.

Complete Spatial Coverage: If $\mathbf{B}$ forms an orthogonal basis, i.e., it consists of $N = M$ orthogonal or pseudo-orthogonal beams, then it provides complete spatial coverage. Any complete M-dimensional basis used for beamsweeping will provide complete coverage of the CSI space, since, by definition, the CSI of any user can be represented by a linear combination of the basis. This ensures that for any given point in the coverage area at least one beam in B will not have a perfect null.

It is important to note that as M increases, the probability that a user detects a given beam is reduced, since the energy is more spatially selective. However, the probability that a user will detect at least one beam in the sweep pattern increases, as, given a complete orthogonal basis, at least one beam is pointed towards the user, and that beam has a higher EIRP since it is narrower.

Techniques and Range: *Faros* can leverage many beamforming techniques with compelling tradeoffs for specific implementations. Without detailed information about the environment and precise calibration, any orthogonal basis with low peak to average power ratio (PAPR) works well for open-

loop beamforming. While a complete basis guarantees spatial coverage, it does not guarantee a strong signal. Since it is statistically impossible that every user will have an open-loop beam pointed directly at them, the gain of beamsweeping is reduced by an inaccuracy factor of a, to M^2/a. As such, an overcomplete $\mathbf{B}$, i.e. $N > M$, can provide extended coverage by statistically reducing a. Otherwise, given careful consideration of the propagation environment and antenna placement, as well as hardware calibration, techniques such as DFT open-loop beamforming can be tuned to provide the desired coverage area. For our implementation we choose Hadamard beamforming weights, as further described in §6.

4.2 Coding Gain

The use of open-loop beamsweeping will reduce the gain gap between no-CSI and CSI modes. To close the remaining gap, *Faros* additionally employs a variable coding gain in both the downlink and uplink. In theory, a coding gain is achieved by sending a signal over a longer period of time, thus, the total received power, integrated over time, increases linearly as the duration increases. However, this gain comes at a cost of increasing the channel usage overhead linearly as well. Coding gains are ideal for tuning the gains to match between modes because they are easily adjustable and thus can be used to dynamically fine-tune the gain vs. overhead tradeoff.

While Table 1 analyzes the gain gap in terms of SINR, not all parts of the frame have the same SINR requirements. For example, data transfer can benefit from a higher SINR by altering the modulation and coding scheme. Higher-ordered modulation requires a higher SINR to be successfully decoded, thus it can be thought of as a negative coding gain in the CSI mode. For instance, in 802.11 OFDM BPSK modulation requires 15 dB SINR, whereas 64-QAM requires 31 dB [22]. In contrast, the detection threshold for a length 128 Kasami sequence is roughly -5 dB [23]. This effectively further reduces the gain gap between the CSI mode, which is used for tranmitting data, and no-CSI mode, but how much is dependent on actual data modulation rate. By leveraging a dynamic coding gain, the range and overhead of *Faros* can be tuned to the specific needs of each deployment.

Downlink Coding Method: In the downlink, *Faros* transmits variable length orthogonal synchronization sequences to encode the base-station ID and paging information, while simultaneously providing synchronization and achieving a gain, C_{down}, proportional to the length of the sequence. Orthogonal sequences are extensively used in wireless systems; an overview of them can be found in [24]. Since these downlink sequences need to be detected *prior* to synchronization, they must have low streaming auto-correlations, both with themselves and the other sequences in the orthogonal set. That is, since the sequences must be detectable without knowledge of when they start, the receiver must perform a full correlation at every sample, thus a time-shift of the sequences must produce a low correlation; otherwise it could cause an erroneous detection.

Uplink Coding Method: In the uplink, the *Faros* base station assigns orthogonal pilot slots to active users, and reserves dedicated slots for association and random access, as shown in Figure 2(b). These pilot slots are variable length to enable a coding gain based on users' channel quality, e.g., users on the cell edges will use longer pilots to increase the accuracy of their channel estimate.

By orthogonalizing pilots in frequency *Faros* is able to increase the accuracy of the channel estimates, and provide an uplink gain of at least K. Frequency orthogonalization (OFDMA) enables all the users to transmit simultaneously, which increases the instantaneous power received at the base station by a factor of K. To collect complete CSI for every frequency, users are further time orthogonalized, as shown in Figure 2b. As such, the total power received for a given user, integrated over time, also increases by a factor of K. Theoretically, to obtain accurate CSI each user must send a pilot for at least a duration of the inverse of the frequency coherence every coherence time interval. However, by scheduling users with poor channel quality to send even longer than required by the frequency coherence interval, *Faros* increases the coding gain, C_{up}; this ensures high-quality channel measurements across the entire cell and fully closes the gain gap.

For association and random access, users send orthogonal synchronization sequences on dedicated time-frequency blocks during the training phase. This allows the users to still achieve a coding gain, while simultaneously enabling collision avoidance and timing-advance estimation, as further discussed in §5.4.

4.3 Combined Gain

Faros employs a combination of open-loop beamforming and coding gain to close the gain gap, as depicted in Figure 1. Beamsweeping provides the majority of downlink gain by focusing the full power of the base station on a small portion of the coverage area; it achieves a gain of M^2/a, where a is the beamforming inaccuracy. In the downlink *Faros* reduces the gap between no-CSI and CSI gains from M^2/K to $M^2/K/(C_{down} \cdot M^2/a) = a/(C_{down} \cdot K)$, thus the coding gain should be tuned so that $C_{down} \approx a/K$.

In the uplink *Faros* leverages OFDMA and coding to achieve a gain of $C_{up} \cdot K$ in the no-CSI mode. This reduces the no-CSI to CSI gap from M to $M/(K \cdot C_{up})$, which suggests C_{up} should be roughly M/K to close the gap.

However, once a proper downlink coding gain, C_{down}, is applied, combined with beamsweeping, the *Faros* no-CSI downlink gain is M^2/K. In contrast, the no-CSI uplink gain is only $(C_{up} \cdot K \cdot P_U)$, which leads to a new gain gap. To mitigate this gap, in *Faros* the *total* transmission power of the base station and user need to be roughly the same, e.g. $O(P_U) \approx O(M \cdot P_{BS})$; this is typical of existing bidirectional communication systems, though macro cells can have as high as a 10 to 18 dB difference. This reduces the gap from $(C_{up} \cdot K \cdot P_U)/(M^2/K \cdot P_{BS})$ to $(C_{up} \cdot K^2)/M$, and suggests that the uplink coding gain should be tuned to approximately M/K^2, along with any residual discrepancy between P_U and P_{BS}, to finish closing the gap.

Comparing the C_{up} needed for the no-CSI vs. CSI, M/K, and uplink vs. downlink, M/K^2, we see there is a residual gap of K. Since the range of the base station is limited by the downlink mode, C_{up} should be selected to match the uplink-downlink gap, then the residual gain of K in the CSI uplink can be used to reduce transmission power or increase modulation rate. Notably, this full coding gain is only required at cell edges, where *Faros* uses extra-long pilots.

It is important to realize that when compared to existing systems, for a given coverage area *Faros reduces* the required per-antenna transmission power of the base station by M^2 and of the user by K.

5. FAROS CONTROL CHANNEL DESIGN

We next describe the design of *Faros* control channel design and how it realizes synchronization, association, CSI collection, random access, and paging.

5.1 Synchronization

Faros achieves both time and frequency synchronization by beamsweeping carefully designed, extended-length, sequences from the base station to the user.

5.1.1 Time Synchronization

With *Faros*, users perform a streaming cross-correlation on received samples to detect the synchronization sequence sent from the base station. That is, it computes the correlation of the received signal R with the sequence S, $\sum_{i=1}^{n}(R_{t-i} \cdot S_i^*)$, at every sample. This produces a peak at the single sample when R and S are aligned in time.

While this is the same concept employed by existing systems, *Faros* faces two new challenges: (i) *Faros* needs to detect multiple synchronization sequences simultaneously since it uses both beacon and paging sequences for synchronization, which are sent simultaneously on separate beams. (ii) *Faros* needs to perform time synchronization *without* coarse timing information or automatic gain control (AGC). As discussed in §2.2, existing solutions leverage coarse frame detection and AGC to achieve fine-grain time synchronization; however, these techniques are inefficient or even impossible for *Faros* to employ in the no-CSI mode. This is because *Faros*' beamsweeps and MU-MIMO downlink are highly spatially selective and, as a result, users receive every synchronization sequence with highly varying power. While *Faros* could precede every synchronization sequence with a training sequence to facilitate coarse frame detection and AGC, similar to the STS in 802.11, this training sequence would have to have significantly increased length to overcome the gain gap. Moreover, the gains set by this sequence would only be valid for a single beam, making it highly inefficient.

Faros addresses these two challenges with three techniques. First, it employs two full-precision correlators. Existing implementations, such as [25, 26], perform only 1-bit and 3-bit correlations, respectively, and only detect a single pre-set sequence. While this approach is computationally efficient, it does not work well without gain control, and performs poorly when trying to distinguish different sequences. By performing two parallel full-precision correlations, e.g. 12-bit for WARPv3, *Faros* is able to reliably detect synchronization sequences with highly varying signal strengths, as well as reliably distinguish paging and beacon synchronization sequences that are sent simultaneously.

Second, since performing AGC on every sequence is inefficient, *Faros* employs *transmit* gain control. That is, since *Faros* beamsweeps the sequence, a user receives every sequence with a substantially different signal strength. Therefore, the users can simply wait for a sequence in the sweep that is within their dynamic range. If they don't detect any sequences, e.g. before discovering any base stations, the users slowly vary their receive gain settings until they detect sequences. After synchronization is established the users listen to all of the subsequent synchronization sequences and adjust their gain accordingly. Notably, *Faros* performs uplink gain control identically to LTE [27], using feedback, and fine-grain downlink gain control is performed at the beginning of each downlink phase, as depicted by Figure 2.

Finally, *Faros* dynamically sets detection threshold by combining the running average of the correlator output and a spike detector. This is because without traditional AGC, the single-sample correlation peak varies drastically in magnitude. The average correlator output provides the average input power, but is additionally scaled by the power of the correlation sequence so that different sequences can be detected without adjusting the threshold. The spike detector simply raises the threshold exponentially when there is a short burst of power, thus avoiding erroneous false-positives. Existing techniques, such as [25, 26], employ a static threshold for peak detection, as they leverage AGC to consistently set the magnitude of the digital samples, and thus peak. Other reported correlator designs, e.g., [23], use the input power to set the detection threshold, however we found this approach by itself to be inadequate for *Faros*. This approach is susceptible to false-positives from power spikes without retrospective processing, and does not automatically adapt the threshold to sequences with different PAPRs.

5.1.2 Frequency Synchronization

To determine the carrier frequency offset (CFO), the user calculates the phase drift in the downlink synchronization sequence. This sequence consists two repetitions of the same sub-sequence; since the drift from CFO is constant, corresponding received samples in each repetition have the same phase offset. That is, for an n length sequence repeated twice to give the synchronization sequence S, $\theta(S_i, S_{i+n}) = \theta(S_j, S_{j+n})$, where θ is the phase difference between the complex samples. This is because S_i and S_{i+n} are the same symbol, thus in the absence of CFO $\theta(S_i, S_{i+n}) = 0$; with CFO there is a phase drift that is proportional to time n, which is thus constant across all i: $\theta(S_i, S_{i+n}) = drift(n)$. Thus, we can use the following equation to compute CFO:

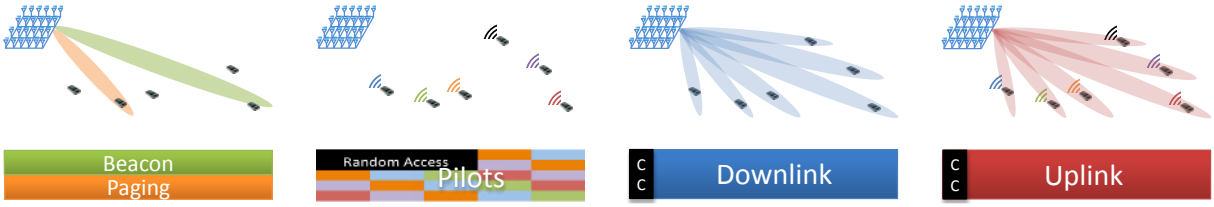
$$CFO = \frac{1}{2\pi \cdot n} \sum_{i=1}^{n} \theta(S_i, S_{i+n}) \quad (1)$$

Notably, in hardware the division by 2π is not actually performed, since the CFO is multiplied by 2π when generating the correcting complex sinusoid. Thus by selecting n to be a power of 2, the division becomes a trivial bitshift. In the presence of noise, longer sequences become more reliable, as the noise is filtered out by the averaging operation. While there are other techniques to compute CFO, such as the conjugate method adopted by LTE [27], *Faros* employs this technique since it enables two synchronization sequences to be simultaneously without affecting CFO recovery. Since both sequences have sub-sequences that repeat twice, the combined signal also repeats twice and can still be used to accurately calculate CFO.

To avoid frequency distortion in multipath environments, typically a cyclic prefix is prepended to the synchronization sequence. However, this cyclic prefix makes time synchronization less robust, as it can cause false positives in the correlator, since it aligns with a subset of the sequence. To avoid this, we use a cyclic *postfix*, then delay the CFO calculation accordingly, i.e., the sum in equation 1 starts at the length of the cyclic postfix. Note that this does not affect the correlator performance, as it operates in the time-domain.

5.2 Association Procedure

Faros enables association by: (i) encoding a unique base-station identifier in the beamswept synchronization sequence,



(a) Beamsweep (b) CSI collection (c) Downlink data (d) Uplink data

Figure 2: An example *Faros* frame structure. First, in (a), the base station beamsweeps a beacon that provides the users with time-frequency synchronization and the base-station ID. If a user needs to be paged, the base station will simultaneously beam a paging sequence towards that user. Next, in (b), users send orthogonal uplink pilots in scheduled slots. Users that require random access or association send an uplink pilot in the one of the reserved slots. Finally, in (c) and (d), the base station leverages the acquired CSI to provide downlink and uplink data connectivity, as well as any remaining control channel information, over the efficient MU-MIMO link.

i.e. the *beacon*, (*ii*) having users scan for these beacons to *select a base station*, and (*iii*) providing a *'soft' association* mechanism that allows users to quickly obtain more information about the base station over a MIMO link. We next elaborate on each of these steps.

Beacons: In *Faros*, every base station beamsweeps a synchronization sequence that encodes a locally unique identifier, called a beacon, as shown in Figure 2(a) and discussed in §4.2. This enables users to simultaneously synchronize with a base station, as well as identify it. For the sake of brevity, we assume that the base stations are coordinated so that they each have locally unique identifiers and can ensure that their beacons do not overlap in time, which prevents random access collisions and reduces pilot contamination. While there are straightforward techniques for achieving this coordination, e.g., through the backhaul or via a user, that discussion is outside the scope of this paper.

Base Station Selection: Before associating, a user listens for at least one entire sweep interval, perhaps on multiple frequencies, to determine the IDs of all nearby base stations, as well as the average power of the beacons from each base station. Since the beacon is beamformed, its received power does not indicate the actual channel quality between the user and the base station. Thus it is important for the user to listen to beacons for an entire sweep interval to obtain a rough estimate of the signal strength from each base station, but the true SINR and channel quality, cannot be accurately determined until after association due to the beamforming inaccuracy described in §4.1. Furthermore, the unique identifier contained in the beacon does not convey any additional information, such as authentication, encryption, and a human-readable identifier (e.g. an SSID). Therefore, the user may soft-associate to multiple base stations in order to search for the best match.

Soft-Association: Since *Faros* beacons only contain a unique identifier, we additionally provide a mechanism called *soft-association* which enables users to gather more information over the CSI mode. Traditional control channel designs broadcast information about the base station in their beacons. For example, 802.11 beacons include the BSSID, SSID, modulation rate, encryption information, and more. This information is essential for users to determine if they want to, or even can, connect to the base station. Moreover, the users need to be able to judge their channel quality to the base station, which can only be done in CSI mode.

Guided by our first key insight, that as much control information as possible should be sent over the more efficient MU-MIMO channel, *Faros* provides soft-association to enable users to quickly establish a MIMO link with the base station to efficiently exchange control channel information. To perform a soft-association, users must first synchronize with the base station by successfully decoding a beacon, then send a pilot in one of the slots reserved for random access, as discussed below. Once the base station successfully receives the pilot it has CSI for that user, which it leverages to open a MIMO link and convey the remaining control channel information. If the user proceeds with a full association, based on authorization, link quality, etc., the base station schedules the user dedicated pilot slots and a unique paging sequence to maintain the link. Otherwise, the user continues to scan for and soft-associate to other base stations.

5.3 Collecting CSI

After each beacon, all active users send uplink pilots in their scheduled slots which the base station leverages to collect CSI. It is best to think of the CSI collection phase as a number of time-frequency-code resource slots that can be arbitrarily assigned to users, with some resource slots dedicated to random access, including association requests and paging responses. Users which send reference signals in a given resource element gain spatial resource elements in the corresponding time and frequency coherence interval for both the uplink and downlink phases. That is, any given reference symbol provides an estimation that is valid both for the coherence time interval, as well as a wider frequency coherence interval. As noted in §4.2, *Faros* assigns longer pilot slots to users that have worse channels in order to improve CSI accuracy.

5.4 Random Access

Faros facilitates random access by reserving pilot slots at the beginning of each channel estimation phase, as shown by Figure 2(b). To initiate a connection users simply send an uplink pilot in one of these pilot slots. For the user to send in the correct pilot slots, without interfering with other users, it must have successfully received a beacon, and thus established synchronization. The base station uses this pilot to estimate the user's channel, as well as timing advance, and create a highly efficient MU-MIMO link to the user. As guided by our first key insight, this link is then used to convey all remaining control channel information, including modulation rates and pilot scheduling, as well as maintain/improve synchronization.

LTE already provides a compelling random access solution which fits well within the *Faros* design, with the exception that *Faros* allows for longer length sequences to be employed to finely tune the gain gap. Thus, due to space constraints, we defer to [28] to fully describe the LTE random access

Var	Description	Overhead	Description
L	Sequence Length	C	Channel Utilization
B	Bandwidth	D_A	Association Delay
F	Frame Duration	D_R	Random Access Delay
N	# of beams		

$$C = \frac{L/B}{F} \quad D_A = \frac{N \cdot F}{2} \quad D_R = \frac{F}{2}$$

L	B	F	N	C	D_A	D_R
128	$20MHz$	$15ms$	100	0.043%	$750\,ms$	$7.5\,ms$
128	$40MHz$	$1ms$	100	0.32%	$50\,ms$	$0.5\,ms$
256	$20MHz$	$10ms$	100	0.128%	$500\,ms$	$5\,ms$
256	$20MHz$	$5ms$	500	0.256%	$1250\,ms$	$2.5\,ms$
512	$40MHz$	$2ms$	1000	0.64%	$1000\,ms$	$1\,ms$
1024	$80MHz$	$1ms$	4000	1.28%	$2000\,ms$	$0.5\,ms$

Table 2: Analysis of *Faros*' beacon overhead. *Top:* Variable descriptions. *Middle:* Equations used for analysis. *Bottom:* Expected value of the worst-case overheads of the simplest version of *Faros* given various realistic system parameters.

scheme, including collision detection and avoidance, as well as timing advance, which we employ in *Faros*.

5.5 Paging

Faros enables many-antenna base stations to reliably and *quickly* page users across their entire coverage area. To accomplish this *Faros* applies the beamsweeping and coding gains described in §4; unfortunately, unlike synchronization and association, paging is not delay tolerant. Thus *Faros* leverages the users last known location to substantially reduce the delay from beamsweeping.

Upon association, the base station assigns each user a unique paging sequence. This paging sequence is constructed and transmitted almost identically to the beacon. That is, it is chosen from the same codebook as the beacon to ensure orthogonality, as well as repeated twice to facilitate time-frequency synchronization. To page a user the *Faros* base station beamsweeps their unique paging sequence with the beacon at the beginning of each frame, but on a separate beam, as shown in Figure 2(a). This additional spatial separation between the beacon and paging sequence helps improve the detection of either, as it reduces the inter-sequence interference. To detect the paging sequence, users perform the same synchronization correlation used for the beacon, described in §5.1. Successful detection similarly provides the user with synchronization, however in the case of a paging sequence the user immediately sends an uplink pilot in the dedicated random access pilot slot. This allows the base station to estimate CSI and begin MIMO communication.

One key challenge facing *Faros* is that while association and synchronization are not time-sensitive, the delay from beamsweeping is likely unacceptable for paging, e.g., up to 2 s in Table 2. To solve this challenge, *Faros leverages knowledge of the user's prior location to guide the beamsweep*, even in our naïve implementation this sped up paging by 400%, as demonstrated in §7.3. Note that leveraging the users last known location can only improve expected paging delay, as the sweep continues until the user is paged.

Link Maintenance: Additionally, or alternatively, users will periodically send a random access request to the base station. This serves the multi-purpose of maintaining the association, checking for missed page requests, and updating the users' last known location at the base station to assist with efficient paging and inter-base station handovers.

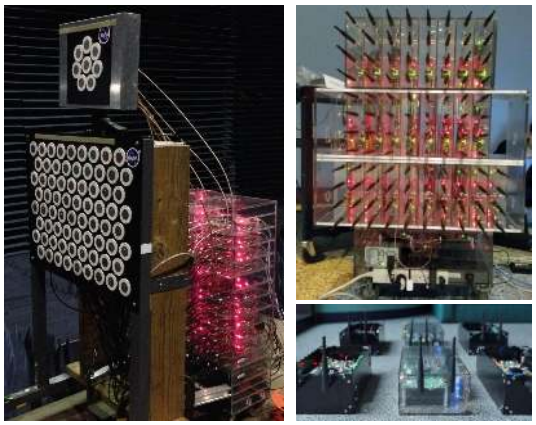


Figure 3: Our prototype, Argos. *Left:* 80-antenna array in an anechoic chamber. *Top Right:* 104-antenna array in an indoor environment. *Bottom Right:* ArgosMobile user devices.

5.6 Overhead Analysis

By design, *Faros* has a small, if not negligible, overhead. This overhead can be measured by four metrics: (*i*) total channel overhead, (*ii*) association delay, (*iii*) random access delay, and (*iv*) paging delay. Table 2 provides the equations for determining these overheads, then provide example values for reasonable system configurations. For this analysis we assume that frames are sent continuously, with the beacon at the beginning of each frame similar to a scheduled MAC. Since the expected paging delay is dependent on the paging scheme, we discuss its real-world performance using a naïve scheme in §7.3, however it is upper-bounded by the association delay as that is how long it takes to perform a full beam-sweep.

It is important to note that active users *do not* need to receive valid beacons to maintain synchronization, as it is maintained in the CSI downlink control phase. Inactive, but associated users can also maintain synchronization by listening for beacons and paging signals. We note that the duration that time-frequency synchronization is valid depends on the accuracy of the oscillators, frame design, e.g., cyclic prefix, as well as fluctuations in temperature. Given the typical accuracy of oscillators in WiFi and LTE devices, and according to our measurements, the synchronization is usually valid for 100s of ms, but this can be determined on a per-system basis [29]. As such, beacons are only needed for association, and thus the sweep interval can be adjusted accordingly. We also find these overheads are very easy to tune by changing the system parameters. Note that per Table 2, *Faros* can support thousands of antennas with less than 2% overhead, at the cost of slightly increased association delay at the cell edges.

6. IMPLEMENTATION

We implement *Faros* on ArgosV2 [30], a prototype of many-antenna MU-MIMO base station that consists of an array of 27 WARP boards [26], driving 108 antennas, and 5 battery powered WARP-based ArgosMobiles that can be controlled wirelessly through a WiFi bridge, as shown in Figure 3. Both the implementations of *Faros* and Argos can support many times more antennas and users; the reported implementation is only limited by the number of WARP boards available to us. To the best of our knowledge, this

is the largest many-antenna MU-MIMO base station with publicly reported results.

Our implementation of *Faros* serves as the basis of Argos' realtime design, and involves development across all layers of the Argos architecture. To enable realtime operation we designed multiple custom Xilinx System Generator IP cores for both the base station and mobile nodes' Virtex 6 FPGA. The most computationally complex IP core we developed for the mobile nodes is the streaming correlator. The correlator enables realtime detection of beacon and paging codes simultaneously, can be dynamically reprogrammed with different sequences, and supports multiple rates and lengths. For the base station our most significant IP core is the MU-MIMO precoder, which we modified to support beamsweeping, as well as selecting and sending multiple paging sequences simultaneously on different beams. While System Generator IP cores are built with a graphical model and do not directly have lines of code, we use the Xilinx xBlock scripting language to dynamically build a significant portion of them, which constitutes over 4,000 lines of code. These IP cores are integrated with peripherals and other IP cores, including a Microblaze soft-core that is programmed with over 1,000 lines of embedded C.

We implement two versions of the central controller, one in Matlab, and one in Python, both of which are over 2,000 lines of code. The Matlab version facilitates flexible non-realtime experiments with rapid analysis, whereas the Python version supports realtime operation, including fully mobile channel estimation with a time resolution up to 200 μs. Our implementation of *Faros* is extremely versatile; it can be compiled to support detecting any code length, given adequate FPGA resources, and can support any beamforming technique by simply reloading the beamsweep buffers with the corresponding precomputed $\mathbf{B}$.

Open-loop beamsweeping: Our implementation uses Hadamard beamweights [31] for beamsweeping for the following reasons. First, they use a minimal number of weights to provide a complete, perfectly orthogonal, basis; this enables a full diversity gain and provides complete spatial coverage with the minimal amount of overhead. Second, they have a perfect peak-to-average power ratio (PAPR) of 1, which allows the antennas to use their full potential transmit power. Finally, calculating Hadamard beamweights does not require any knowledge of the antenna aperture or environment, enabling rapid deployment without calibration or environmental considerations.

Coding: The implementation uses Kasami sequences for the downlink coding. Kasami sequences [32] provide very good detection performance and have low, bounded, streaming correlation both with themselves and the other orthogonal sequences. This allows them to be reliably detected *without* time synchronization, as a streaming correlation on other sequences could produce peaks, and thus false positives, which is important since they are used for time synchronization. Moreover, they provide a large number of orthogonal sequences, e.g., 4096 for a length 256 Kasami sequence, which enables co-located users and base stations to be uniquely identified.

The implementation uses Zadoff-Chu sequences [33, 34] for the uplink channel estimation coding for the following reasons: First, they have a constant amplitude and thus have a perfect PAPR. Second, they can be used to detect multiple users' random access request simultaneously, along with each users' path delay to estimate timing advance, with small computational overhead. This is very similar to LTE's random access preamble [27]. However, in our design we allow variable length sequences in order to match gain requirements, as well as use the sequence for CSI estimation.

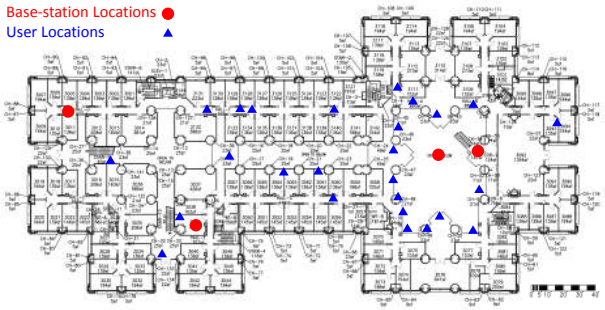


Figure 4: Floorplan depicting example locations of indoor measurements. Both the users and base station locations spanned three floors of elevation.

Thresholding and Variability: *Faros* leverages a realtime streaming time-domain correlator for the beacon, paging, and synchronization, which creates a very strong single-sample peak when the correct sequence is detected. As such, the performance range and accuracy is highly dependent on the detection *threshold*. This threshold is well understood theoretically with regard to false-positive and false-negative performance, and as such we defer to [23] for a more thorough analysis. Since we do not perform gain control for the beacon or paging code we must set this threshold dynamically based on the input power, as well as increase it during power surges to avoid false-positives. This dynamic threshold in *Faros* can be scaled by a constant via software; for the experiments we set the threshold somewhat aggressively so that it is close to impossible to receive a false positive, as we didn't across the 100,000s of synchronization sequences we sent during our experimentation. This threshold could be further optimized to increase range, particularly with mechanisms to deal with false positives.

7. REAL-WORLD PERFORMANCE

We evaluate the performance of *Faros* in bridging the gain gap in real-world topologies. We examine the fully functioning system, and evaluate its performance regarding synchronization, beacons, and paging in diverse environments. Our results demonstrate that *Faros* can extend the no-CSI mode range by over 40 dB when compared to traditional control channels. Furthermore, we find that leveraging knowledge of the users previous location can improve paging delay by 400%, and that *Faros* can reliably correct CFO of over 10 kHz. We first describe our experimental setup, then look at how each of the components performs individually.

7.1 Experimental Setup

We test the performance of the reported *Faros* implementation in 100 discrete user locations at varying distances from the base station in indoor environments and an anechoic chamber, using five ArgosMobiles simultaneously. Additionally, we perform an outdoor range and mobility test, presented in§7.2.1.

Antenna Configurations: Due to hardware availability, and to test the performance of different antennas, we employed *Faros* with three separate antenna configurations: (i) In the anechoic chamber with 80 directional 6 dBi patch an-

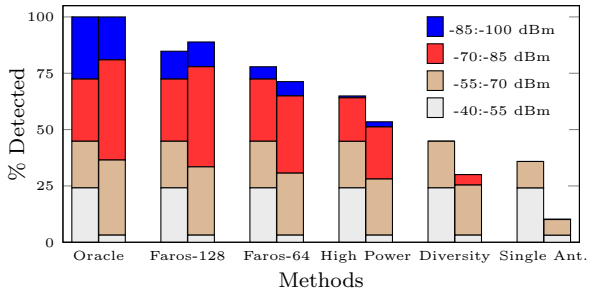


Figure 5: Beacon detection performance across all 32 anechoic chamber (left) and 68 indoor (right) experiment locations. *Oracle* denotes an oracle that detects every beacon sent.

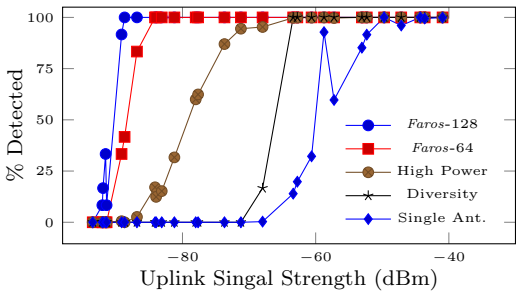


Figure 6: Beacon detection performance vs. uplink RSSI (range) for *Faros* in an anechoic chamber. *Faros* outperforms traditional by over 40 dB. Number indicates beacon length.

tennas, (*ii*) indoors and outdoors with 104 omnidirectional 3 dBi monopole antennas, and (*iii*) indoors with 108 of the same omnidirectional antennas. In all configurations the users also leveraged the 3 dBi omnidirectional antennas.

Power Settings: In the downlink we use a power of approximately -12 dBm per antenna and in the uplink we use 10 dBm. This downlink power is chosen since it results in a ~10 mW total power and an EIRP of up to 1 W, which is the FCC limit. Our prototype is capable of over 10 W total power, and EIRPs exceeding 1 kW, which is only appropriate in licensed bands. For high-power single antenna mode we use the approximate combined power of all of the base station antennas, which is 8 dBm.

Environments and Range: As Figures 5 and 6 depict, we selected locations at increasing distances until the beacons couldn't be detected, providing a fairly uniform selection of signal strengths. In indoor locations this required spacing the users at up to 80 m away, across 3 floors of elevation, as illustrated by the sample of locations presented in Figure 4; in outdoor line-of-sight locations it was over 250 m away. For the anechoic chamber experiments users were spaced up to 15 m away from the base station, and we used variable attenuators with up to 60 dB of suppression to simulate increased distance.

Measurement: At each location we test the *Faros* control channel system over a 20 MHz bandwidth at 2.4 GHz and analyze the performance with regard to the accurate detection of the beacon, paging signal, and uplink pilot, which demonstrate *Faros*' performance in no-CSI mode. As a control, we additionally send an unbeamformed beacon and paging signal from each base-station antenna, i.e. a "beamsweep" using the identity matrix, in both low and high-power modes using a 64 length code to compare the performance with traditional single antenna systems and the naïve high-power solution. While *Faros* is capable of running in realtime, we briefly pause after every beam in order to collect performance statistics from the nodes, such as successful detections, false positives, and received signal strength indicators (RSSIs). Because of this measurement delay, these experiments were conducted without mobility, in relatively stationary channels. We use these results to analyze the performance of *Faros*' beacon, paging, and CSI collection vs. traditional methods, which we present below. Additionally we setup a controlled experiment to test the performance of our CFO estimator, presented in §7.4.

7.2 Beacon Performance

Figures 5 and 6 show the probability of successfully receiving the base station's beacon, i.e., the synchronization sequence encoded with the base-station ID, with various configuration parameters. We compare single-antenna transmission, both high power (*High Power*) and low power (*Single Ant.*), diversity, and *Faros* using code lengths of 64 and 128. In the single antenna diversity mode (*Diversity*) the base station rotates which antenna is transmitting, thus exploiting the full diversity of the array; this is equivalent to *Faros* using the identity matrix for beamsweeping.

The figures sort the results based on the average uplink CSI signal strength across all base-station antennas for the given location, which is an approximation of distance and a fair metric for coverage area. We note that downlink RSSI is not a good metric, since it varies per-beam. Distance is not a good metric since scatterers can significantly alter signal strength. Clearly, changing uplink transmission power will simply shift the same plot either left or right, which indicates how code length and both uplink and downlink transmission powers should be balanced in a real system.

The results across all locations are shown in Figure 5, with separate bars for the 36 anechoic chamber locations and 64 indoor, including 104- and 108-antenna, locations. We see that indoor locations *Faros* is able to reliably serve over 8.8 times more locations than the traditional control channel, and 1.6 times more than a single high power antenna. Even when users have over a -70 dBm average RSSI to the base station, they miss almost 25% of the beacons sent with the high-power single-antenna scheme. This is due to multipath; in some locations, even fairly close, two paths will destructively interfere and create a null, which is not easily overcome with additional signal strength. While the diversity scheme performs better than the single antenna, it is still unable to reliably receive many beacons where users have lower than -70 dBm uplink RSSI. This illustrates the necessity of *Faros*, which leverages both the power and diversity of the entire array, in many-antenna MU-MIMO systems.

The results from the anechoic chamber are shown in Figure 6. Since there is no multipath in the anechoic chamber, the detection rate of each technique is very closely related to RSSI, thus these results accurately demonstrate the relative performance of each technique. We find that *Faros* is able to outperform a single-antenna scheme by over 40 dB, and the high-power scheme by 20 dB.

7.2.1 Range and Mobility Performance

To demonstrate the realtime capability of *Faros*, as well as test its range and mobility performance, we performed an outdoor experiment where we ran *Faros* at full speed. Unfortunately, the previous tests required us to pause the experiments after every beacon or paging signal was trans-

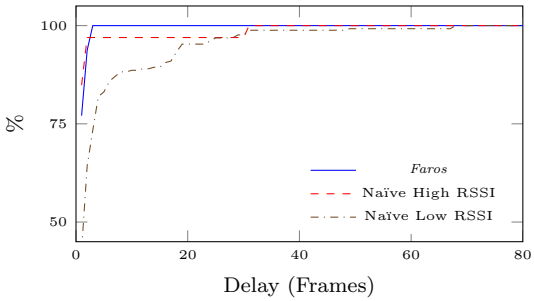


Figure 7: Cumulative distribution functions of paging delay. The naïve method does not use location data to sweep. *Faros* improves mean paging delay by 400% at low RSSIs.

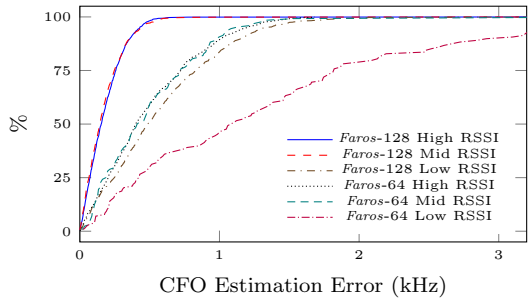


Figure 8: Cumulative distribution functions of CFO estimation error with various sequence lengths and RSSIs. *Faros* provides frequency synchronization within 800 Hz at up to -75 dBm.

mitted and collect measurements, which prevented realtime operation. For this experiment we had the base station continuously beamsweep the beacon at a frame rate of one beam per 10 ms, then had users move away from the base station at a walking pace. In line-of-sight the users performed reliably, and concurrently, at over 250 meters at multiple angles from the base station, and only began to lose reliability the users had to move behind buildings due to space constraints.

7.3 Paging Performance

To demonstrate *Faros*' ability to leverage location information to accelerate paging, we tested a simple scheme which guided the paging sweep based on the intended user's last location. These experiments were performed on the 108-antenna base station configuration in the last 44 locations. In the prior locations we had employed RSSI to guide the sweep, but realized that due to multipath distortion this was not the best detection performance metric, since the time-domain correlation essentially filters individual paths. Instead, we paged mobiles based on each beam's detectability, which is determined by the correlation magnitude to threshold ratio.

We find that *Faros* was able to successfully page 94% of users by the second frame, compared to only 70% without leveraging the user location, as shown in Figure 7. When users are near the base station they receive the majority of the beams in a sweep, and thus optimizing based on their location does not provide much benefit, as shown by the low RSSI plot. However, we still see the paging delay reduced from an average of 4.8 frames to 1.2 frames, an improvement of 4 fold, and a worst-case improvement of 68 frames to 3 frames. This system is very naïve, and is intended to demonstrate *Faros*' ability to leverage spatial information to drastically improve the performance of the control channel without additional overhead.

7.4 CFO Correction Performance

While successful detection of a beacon or paging sequence inherently provides time-frequency synchronization, to more accurately test the accuracy of our realtime CFO correction we setup a more controlled experiment. We shared a reference clock between the base station and user, effectively removing CFO, and placed the user at 0.5 m from the the base station. Then we *induced* a controlled CFO in our beacon sequence by multiplying it with a complex sinusoid ranging from -10 kHz to 10 kHz. To measure the performance vs. coding gain and SNR, we sent beacons of length 64 and 128, as well as used attenuators on the base station to reduce the transmission power from -12 dBm to -42 dBm. These attenuations resulted in the user receiving roughly -60 dBm (*High*), -75 dBm (*Mid*), and -90 dBm (*Low*) RSSIs. We present the cumulative distribution of the error magnitude of our CFO estimates in Figure 8. For clarity, these results are derived from a single estimation, however multiple estimates can be employed to reduce the error by an order of magnitude, as shown in [29].

We find that with mid and high RSSI *Faros* is always able to correct CFO within 0.8 kHz using a 128-length beacon, and within 1.3 kHz using a 64-length beacon. This estimation error is sufficient to not restrict the capacity of an LTE system, [35]. In the low RSSI regime we see that the 64-length beacon begins to perform poorly, and is only able to correct 80% of the beacons to within 2 kHz error. In contrast, the 128-length beacon with low RSSI is performs similarly to the high RSSI 64-length, which indicates extending the beacon length could further reduce CFO estimation error. The amount of induced CFO did not affect accuracy, and thus is not shown separately.

8. DISCUSSION

Broader use of Faros: Our original goal for *Faros* is to provide a very efficient control channel for many-antenna base stations. More fundamentally however, it represents an interesting paradigm that provides fine-grained control over time, code, and spatial resources, enabling previously impossible optimizations both within a single base station, and across the network. *Faros* allows base stations to leverage existing information, such as users' last known location, traffic patterns, and environmental properties to intelligently optimize timing, coding gains, and spatial coverage. Moreover these same properties can be used to further extend the range of the cell in sparse networks, restrict coverage area, carefully tune interference, or dynamically incorporate more antennas to increase the capacity of a given base station.

MAC and standards: So far we have intentionally avoided discussing the MAC, as the conceptual *Faros* design is MAC-agnostic. The primary requirement of *Faros* is a short regularly scheduled downlink phase in order to perform the beamsweep and paging; and both scheduled and CSMA MACs have this. However, we do not intend *Faros* to be a plug-n-play solution for either LTE or 802.11: both standards must be revised to integrate *Faros*. Applying *Faros* to a scheduled MAC is more intuitive, and likely more efficient: the phases depicted in Figure 2 simply have to continuously repeat, though not necessarily in that order. Schemes to adapt *Faros* to CSMA are also fairly straightforward and we provide one example below.

Design sketch of Faros in 802.11: In an 802.11-like CSMA MAC, the beacon would need to be replaced by the beamformed *Faros* beacon immediately followed by the CSI collection phase, including the dedicated random access and association slots. Since 802.11 typically supports only a small number of users with relatively low mobility, each user could have a dedicated CSI slot, which they use at every beacon interval when they are active. The AP can page inactive users during the beacon phase, making them become active, or, if the channel is idle, the AP could page users asynchronously, prompting them to send a pilot immediately. Since these mechanisms allow the AP to maintain accurate CSI for the users, the downlink phase is straightforward: when the channel is idle the AP simply sends a MU-MIMO transmission to the intended users. Uplink MU-MIMO is difficult to efficiently coordinate in CSMA, which, combined with typical asymmetric data requirements, is why current 802.11 standards do not support uplink MU-MIMO. However, one naïve solution would be to allow the users to indicate an uplink request during the CSI collection phase. The AP could then respond with a "clear-to-send" to selected users over the MU-MIMO channel. Of course, to reduce latency users would not have to wait for the beacon to send a single-user uplink packet.

9. RELATED WORK

To the best of our knowledge, *Faros* is the first reported control channel design that effectively addresses the gain gap between the CSI and no-CSI modes for many-antenna MU-MIMO systems. Nevertheless, various previous works are related to *Faros* in terms of both problem and solution. The challenge of control channel design for many-antenna MU-MIMO is well-known. The authors of [5] suggest utilizing space-time block coding for the control channel, but do not address the gain gap or suggest a design. The authors of [36] discuss control channel operation from a purely theoretical and feasibility perspective, which is complementary to our work. However, its assumption that "the only reasonable transmit strategy is to spread the power omnidirectionally" is questionable: *Faros* beamforms the control and provides a working counter-example. It also assumes that the total base-station power can be sent omnidirectionally in the first place, e.g., there is no peak-power per antenna constraint, which is incorrect for real systems, as discussed in §3. Another recent work from Samsung mentions the control channel briefly, but suggests the solution is to carefully create a wide open-loop beam using all of the antenna elements [11]. This approach requires careful calibration of the antenna elements, is environment and deployment specific, and, more importantly, does not completely serve the full potential coverage area of the base station.

802.11ad suffers from a related gain gap and employs a beamsweeping mechanism to initiate communication. Because 802.11ad does not employ MU-MIMO but phased arrays, its gain gap is fundamentally different and scales with less than M. Moreover, the contiguous Sector Level Sweep (SLS) that 802.11ad performs for synchronization and discovery is naïve, unscalable, and highly inefficient. An 802.11ad SLS with 128 elements can take over 1.5 ms [37], whereas a comparable *Faros* beacon would take less than 150 ns. This indicates that 802.11ad and other mm-wave technology could benefit substantially by incorporating design principles from *Faros*, particularly as they adopt MU-MIMO.

Faros' use of Kasami sequences to send a small portion of the control channel information is inspired by 802.11ec [23], which uses time-domain BPSK modulated Kasami sequences to encode control information in the preamble of 802.11 packets. However, *Faros* addresses an entirely different problem, the gain gap in many-antenna MU-MIMO, and as such, it employs different techniques and contributes an entire from-scratch control channel design. Other recent works, such as [38], have also used similar sequences for other purposes including control messages and power reduction.

Like most modern digital wireless systems, *Faros*' synchronization is based on the seminal works in [39, 40]. More recently, some research has focused on over-the-air time-frequency synchronization in distributed antenna systems, including [41, 42]. However, these works deal with the distributed antennas, not between the distributed system and users. As such, they do not address the synchronization range gap that emerges with multiple antennas on a single base station, or the challenge of paging in such a system. Since these distributed systems require backhaul, this synchronization can also similarly be solved with CPRI, [43], or PTP and SyncE, as employed by CERN's WhiteRabbit [44].

Open-loop beamforming techniques have been thoroughly researched. While we do not advocate for a specific technique in this work, our experiments leveraged Hadamard matrices for the beamweights; recent work in [31] covers the performance of Hadamard beamforming more thoroughly. Fourier transform based beamforming is a classic technique, however we note that it requires precise antenna calibration in order to be effective, as discussed in [9, 16–19].

10. CONCLUDING REMARKS

In this work we present the design, implementation, and experimental validation of *Faros*, a fundamental re-design of the wireless control channel in many-antenna MU-MIMO systems. By holistically considering the practical design constraints of many-antenna base stations, we are able to achieve a flexible design which improves the range, or transmission efficiency, by over 40 dB on a 108 antenna base station with negligible overhead. On a more fundamental level, *Faros* provides flexible optimzation of space, time, code, and frequency resources, enabling it to scale from a few antennas up to 1000s of antennas. Not only does *Faros* drastically improve the performance of basic control channel operations by leveraging MU-MIMO as much as possible, but it also utilizes spatial information to make paging operations as quick and efficient as possible. *Faros* unlocks the full potential of real-world many-antenna MU-MIMO, and brings it one significant step closer to real-world adoption.

Acknowledgements

This work was funded in part by NSF grants MRI 1126478, NeTS 1218700, EARS 1444056, and CRI 1405937. Clayton Shepard was supported by an NDSEG fellowship. We thank Hang Yu, Eugenio Magistretti, Ashutosh Sabharwal, and Nathan Zuege for their input, support, and help. We appreciate the support of the Xilinx University Program, and NASA, JSC for the use of their Antenna Test Facility. We thank the reviewers and shepherd for their constructive input; we especially thank Reviewer B for correcting a mistake in our original uplink gain gap analysis.

REFERENCES

[1] Samsung takes first 5G steps. http://www.computerworld.com/article/2497385/data-center/samsung-takes-first-5g-steps-with-advanced-antenna.html.

[2] Steve Perlman and Antonio Forenza. pCell: Wireless reinvented. http://www.rearden.com/artemis/An-Introduction-to-pCell-White-Paper-150224.pdf, 2014.

[3] P. Rudnick. Digital Beamforming in the Frequency Domain. *Acoustical Society of America Journal*, 46:1089, 1969.

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson. Scaling up MIMO: Opportunities and challenges with very large arrays. *IEEE Signal Processing Magazine*, 30(1):40–60, 2013.

[5] E. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta. Massive MIMO for next generation wireless systems. *IEEE Communications Magazine*, 52(2):186–195, 2014.

[6] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang. An overview of massive MIMO: Benefits and challenges. *IEEE Journal on Selected Topics in Signal Processing*, 8(5):742–758, October 2014.

[7] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski. Five disruptive technology directions for 5G. *IEEE Communications Magazine*, 52(2):74–80, February 2014.

[8] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang. What will 5G be? *IEEE Journal on Selected Areas in Communications*, 32(6):1065–1082, June 2014.

[9] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong. Argos: Practical many-antenna base stations. In *Proc. ACM MobiCom*, 2012.

[10] Qing Yang, Xiaoxiao Li, Hongyi Yao, Ji Fang, Kun Tan, Wenjun Hu, Jiansong Zhang, and Yongguang Zhang. BigStation: Enabling scalable real-time signal processing in large MU-MIMO systems. In *Proc. ACM SIGCOMM*, 2013.

[11] Yang Li, Yan Xm, Mian Dong, Gary Xu, Jianzhong Charlie Zhang, Younsun Kim, and Juho Lee. Implementation of full-dimensional MIMO (FD-MIMO) in LTE. In *Proc. IEEE Asilomar Conference*, pages 998–1003, 2013.

[12] Joao Vieira, Steffen Malkowsky, Karl Nieman, Zachary Miers, Nikhil Kundargi, Liang Liu, Ian Wong, Viktor Owall, Ove Edfors, and Fredrik Tufvesson. A flexible 100-antenna testbed for massive MIMO. In *IEEE GLOBECOM Wrkshp. Massive MIMO: from theory to practice*, 2014.

[13] Nutaq. TitanMIMO. http://nutaq.com/en/products/titanmimo.

[14] Xilinx. Xilinx and BEEcube announce highly scalable prototyping platform for 5G massive MIMO antenna systems. http://press.xilinx.com/2015-02-25-Xilinx-and-BEEcube-Announce-Highly-Scalable-Prototyping-Platform-for-5G-Massive-MIMO-Antenna-Systems.

[15] 3GPP. Study on elevation beamforming/full-dimension (FD) MIMO for LTE. http://www.3gpp.org/ftp/tsg_ran/tsg_ran/TSGR_65/Docs/RP-141644.zip.

[16] Per Zetterberg. Experimental investigation of TDD reciprocity-based zero-forcing transmit precoding. *EURASIP Journal on Advances in Signal Processing*, 2011:5, 2011.

[17] Ryan Rogalin, Ozgun Y. Bursalioglu, Haralabos C. Papadopoulos, Giuseppe Caire, Andreas F. Molisch, Antonios Michaloliakos, Horia Vlad Balan, and Konstantinos Psounis. Scalable synchronization and reciprocity calibration for distributed multiuser MIMO. *IEEE Transactions on Wireless Communications*, pages 1815–1831, 2014.

[18] Kentaro Nishimori, Keizo Cho, Yasushi Takatori, and Toshikazu Hori. Automatic calibration method using transmitting signals of an adaptive array for TDD systems. *IEEE Transactions on Vehicular Technology*, 50(6):1636–1640, 2001.

[19] K. Nishimori, T. Hiraguri, T. Ogawa, and H. Yamada. Throughput performance on IEEE 802.11ac based massive MIMO considering calibration errors. In *Intl. Symp. Antennas and Propagation (ISAP)*, Dec 2014.

[20] Eldad Perahia and Robert Stacey. *Next Generation Wireless LANs: 802.11 n and 802.11 ac*. Cambridge university press, 2013.

[21] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[22] Cisco Systems, Inc. *Cisco Wireless Mesh Access Points, Design and Deployment Guide, Release 7.3*. 2012.

[23] Eugenio Magistretti, Omer Gurewitz, and Edward W Knightly. 802.11ec: collision avoidance without control messages. In *Proc. ACM MobiCom*, 2012.

[24] Farooq Khan. *LTE for 4G mobile broadband: air interface technologies and performance*. Cambridge University Press, 2009.

[25] C. Dick and F. Harris. FPGA implementation of an OFDM PHY. In *Proc. IEEE Asilomar Conference*, November 2003.

[26] Rice University Wireless Open Access Research Platform. warp.rice.edu.

[27] Farooq Khan. *LTE for 4G mobile broadband: air interface technologies and performance*. Cambridge University Press, 2009.

[28] Almamy Touray. LTE: Random access. http://www.lmk.lnt.de/fileadmin/Lehre/Seminar09/Ausarbeitungen/Ausarbeitung_Touray.pdf, 2009.

[29] P. Murphy and A. Sabharwal. Design, implementation, and characterization of a cooperative communications system. *IEEE Transactions on Vehicular Technology*, 60(6):2534–2544, July 2011.

[30] Clayton Shepard, Hang Yu, and Lin Zhong. ArgosV2: A flexible many-antenna research platform. In *Extended Demonstration Abstract in Proc. ACM MobiCom*, 2013.

[31] Young Gil Kim and N.C. Beaulieu. On MIMO beamforming systems using quantized feedback. *IEEE Transactions on Communications*, 58(3):820–827, March 2010.

[32] Dilip V Sarwate and Michael B Pursley. Crosscorrelation properties of pseudorandom and related sequences. *Proceedings of the IEEE*, 68(5):593–619, 1980.

[33] Robert L Frank. Polyphase codes with good nonperiodic correlation properties. *IEEE Transactions on Information Theory*, 9(1):43–45, 1963.

[34] David Chu. Polyphase codes with good periodic correlation properties (corresp.). *IEEE Transactions on information theory*, pages 531–532, 1972.

[35] Qi Wang, Christian Mehlfuhrer, and Markus Rupp. Carrier frequency synchronization in the downlink of 3gpp lte. In *IEEE Int. Synp. Personal Indoor and Mobile Radio Communications (PIMRC)*, pages 939–944, 2010.

[36] Marcus Karlsson and Erik G Larsson. On the operation of massive mimo with and without transmitter csi. In *IEEE Int. Wrkshp. Signal Processing Advances in Wireless Communications (SPAWC)*, pages 1–5, 2014.

[37] Thomas Nitsche, Adriana B Flores, Edward W Knightly, and Joerg Widmer. Steering with eyes closed: mm-wave beam steering without in-band measurement. In *Proc. IEEE INFOCOM*, 2015.

[38] Xinyu Zhang and Kang G Shin. E-Mili: energy-minimizing idle listening in wireless networks. *IEEE Transactions on Mobile Computing*, 11(9):1441–1454, 2012.

[39] Timothy M Schmidl and Donald C Cox. Robust frequency and timing synchronization for OFDM. *IEEE Transactions on Communications*, 45(12):1613–1621, 1997.

[40] Fredrik Tufvesson, Ove Edfors, and Mike Faulkner. Time and frequency synchronization for OFDM using PN-sequence preambles. In *Proc. IEEE VTC*, volume 4, pages 2203–2207, 1999.

[41] Hariharan Rahul, Swarun Kumar, and Dina Katabi. JMB: scaling wireless capacity with user demands. In *Proc. ACM SIGCOMM*, 2012.

[42] H.V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire. AirSync: Enabling distributed multiuser MIMO with full spatial multiplexing. *IEEE/ACM Transactions on Networking*, 21(6):1681–1695, Dec 2013.

[43] Common public radio interface. http://www.cpri.info/spec.html.

[44] Javier Serrano, M Lipinski, T Wlostowski, E Gousiou, Erik van der Bij, M Cattin, and G Daniluk. The white rabbit project. http://www.ohwr.org/projects/white-rabbit, 2013.