Achieving High Data Rates in a Distributed MIMO System

Horia Vlad Balan Ryan Rogalin Antonios Michaloliakos Konstantinos Psounis
Giuseppe Caire
University of Southern California
{hbalan,rogalin,michalol,kpsounis,caire}@usc.edu

A distributed MIMO system consists of several access points connected to a central server and operating as a large distributed multi-antenna access point. In theory, such a system enjoys all the significant performance gains of a traditional MIMO system, and it may be deployed in an enterprise WiFi like setup. In this paper, we investigate the efficiency of such a system in practice. Specifically, we build upon our prior work on developing a distributed MIMO testbed, and study the performance of such a system when both full channel state information is available to the transmitters and when no channel state information is available. In the full channel state information scenario, we implement Zero-Forcing Beamforming (ZFBF) and Tomlinson-Harashima Precoding (THP) which is provably near-optimal in high SNR conditions. In the scenario where no channel information is available, we implement Blind Interference Alignment (BIA), which achieves a higher multiplexing gain (degrees of freedom) than conventional TDMA. Our experimental results show that the performance of our implementation is very close to the theoretically predicted performance and offers significant gains over optimal TDMA. We also discuss medium access layer issues in detail for both scenarios. To the best of our knowledge, this is the first time that the theoretical high data rates of multiuser MIMO systems have been showcased in a real world distributed MIMO testbed.

Categories and Subject Descriptors

C.2.2 [**Computer System Organization**]: Computer Communication Networks

General Terms

Design, Experimentation, Performance

Keywords

Software Radio, Synchronization, Virtual MIMO, Wireless

MobiCom'12, August 22–26, 2012, Istanbul, Turkey.

1. INTRODUCTION

One of the most fundamental challenges in wireless networking today is the increasing demand for higher data rates. In theory, denser deployments, which suppress interference and achieve high spectral efficiency via spatial multiplexing, may be realized in a cost effective manner via a distributed multiuser MIMO system. More specifically, in an enterprise environment, several access points (APs) can be connected to a central server and operate as a large distributed multi-antenna AP, ensuring that all transmitted signal power serves the purpose of data transmission, rather than creating "interference."

In this work, we investigate the efficiency of such a system in practice by focusing on the downlink, that is, the transmissions from the APs to the clients. The first challenge that needs to be addressed is how to achieve tight phase and timing synchronization of the different APs in order to allow for distributed multiuser MIMO precoding. We address this challenge by building upon our prior work on developing a distributed MIMO testbed (see Section 3). We then proceed to study the performance of such a system both when full channel state information is available to the transmitters (full CSIT) and when no channel state information is available (no CSIT).

In the full CSIT scenario we implement two schemes (see Section 4): Zero-Forcing Beamforming (ZFBF) and Tomlinson-Harashima Precoding (THP) [30]. ZFBF is known for its simplicity and high performance (see [6] and references therein), and has been implemented and studied recently in the case of a centralized MIMO system [2]. THP is provably near-optimal in high SNR conditions, the typical use case for WiFi. In the no CSIT scenario, we implement the so called Blind Interference Alignment (BIA) scheme [16], which achieves higher multiplexing gain than orthogonal access (TDMA) by exploiting antenna switching at the receivers in conjunction with a special form of linear precoding at the transmitters (see Section 5). It is worth noting that to the best of our knowledge, neither THP nor BIA have been implemented in the past even in the simpler case of a centralized MIMO setup.

We study the performance gains of the three schemes in our distributed MIMO testbed, comprised of two APs and two clients with one antenna each (see Section 6). Our results are particularly promising. Our implementation of both ZFBF and THP achieve close to 85% of the theoretically achievable gain, an 85% improvement in absolute terms in comparison to TDMA (this gain grows linearly with the number of clients when enough antennas and a sufficiently

rich environment is present). Also, our implementation of BIA achieves up to 60% of the theoretically achievable gain, i.e. a 22% improvement in absolute terms in comparison to optimal TDMA. We also study the performance of these schemes through simulations in a variety of scenarios, and we compare it to that of Dirty Paper Coding (DPC), an information theoretic coding scheme that is known to be capacity achieving in the full CSIT case, but is too complex to be implemented in practice. As a matter of fact, THP can be regarded as the practical approximation of the DPC idea. Finally, we discuss medium access layer issues for both the full CSIT and the no CSIT scenarios (see Section 7). To the best of our knowledge, this is the first time that the theoretical high data rates of a MIMO system have been showcased in a real world distributed MIMO testbed.

2. A MULTIUSER MIMO PRIMER

2.1 Precoding schemes

Since the introduction of multiuser MIMO [7] and the solution of its information theoretic capacity region [29], there has been extensive research on low-complexity practical precoding schemes able to achieve a large fraction of capacity with low complexity. Under the assumption of full CSIT, the capacity achieving scheme consists of Dirty-Paper Coding (DPC) [10] combined with optimal power allocation.

While DPC is essentially an information theoretic tool for realizing perfect interference pre-cancellation with no power penalty, its practical implementation is notoriously difficult [4,13]. THP is a well-known low-complexity scheme for the pre-cancellation of known interference which was proposed and widely studied in the context of precoding for Inter-Symbol Interference channels [14]. At high SNR, THP achieves the same rate as DPC minus a fixed penalty called the "shaping loss", which is an asymptotic power loss (see Section 4.2). For this reason, THP was proposed as a low-complexity alternative to DPC for multiuser MIMO in [30].

An even simpler alternative to non-linear precoding (DPC and THP) is represented by *linear* precoding, i.e., beamforming. Among linear beamforming schemes, the simplest and best known is Zero Forcing Beamforming (ZFBF), consisting of a column-normalized version of the right pseudo-inverse of the channel matrix. ZFBF may outperform THP at low SNR. However, when ZFBF is used in conjunction with practical QAM coded modulation [22] it is affected by the same shaping loss as THP at high SNR. In addition, it has been shown [31] that when the number of users is larger than the one of the transmit antennas, ZFBF may use user selection to approach the system sum capacity.

The conclusion of the above discussion is that, in general, there is no clear a priori ranking of the achievable performance between THP and ZFBF: this depends on the actual channel statistics (scattering environment), operating SNR, number of clients K and jointly coordinated AP antennas M. However, when ZFBF is used with practical codes based on QAM constellations, THP is generally superior to ZFBF unless the system works in low SNR (typically not the case for WLAN broadband access, targeting high spectral efficiency and operating in high SNR) or the number of clients is significantly larger than the number of AP antennas, and greedy selection of the clients is performed. The reason why ZFBF has been the preferred choice in testbed implementations (see Section 2.3) is that it is considered to be easier and lower complexity than THP. Our work, reported in this paper, shows that this is not the case.

The need for CSIT imposes certain constraints on the system hardware, and PHY/MAC protocols. For example, in a Time-Division Duplexing (TDD) scenario (common in wireless LANs), CSIT can be acquired at the APs from incoming uplink pilot symbols, provided that the RF hardware is designed in order to preserve the reciprocity of the uplink and downlink baseband equivalent channels. In order to be able to schedule the downlink clients, it is necessary that these clients send their uplink pilots right before the downlink slot, such that their channel state information is not outdated. This requires the design of special uplink pilots and downlink data frames, which are different from legacy 802.11 WLANs (see the discussion in Section 7). When the hardware is not specifically designed to achieve uplink-downlink reciprocity and/or when legacy MAC protocols must be used, it is impossible to have fresh CSIT of the scheduled clients for each downlink slot and precoding schemes based on CSIT cannot be used. In this case, it is known that if the clients are *statistically equivalent* (i.e., they have identical marginal channel statistics of their channel vectors), then TDMA is optimal. In order to achieve a performance gain over conventional TDMA it is therefore necessary to induce artificially *different* channel statistics for the clients.

Following this idea, a scheme known as Blind Interference Alignment (BIA) was proposed in [16]. The scheme assumes that each client terminal has M antennas, but only one RF chain. At each point in time, only the output of one antenna is demodulated and sampled. Therefore, the complexity and power consumption of the client terminals is the same as in a conventional single antenna terminal, with the addition of a switch that connects the different antennas to the RF demodulation front-end. The client channel statistics are differentiated by assigning to each client a specific antenna switching pattern. In this way, each client "sees" a different time-varying sequence of channel vectors. In [16] it is shown that BIA can send MK independent information streams over blocks of $M + K - 1$ channel uses, and, in idealized conditions, each client is able to perfectly remove the interference from streams destined to the other clients, even though the APs have no CSIT at all.

2.2 Performance

Different schemes can be evaluated in terms of their achievable rates, implicitly assuming ideal coding. Studying achievable rates is much easier than implementing a whole coding scheme and evaluating throughput and packet error rate, since the achievable rates can be calculated on the basis of closed-form information theoretic formulas (see Sections 4 and 5). In general, a coarse performance measure is provided by the so-called "system degrees of freedom" (DoF), also known as "spatial multiplexing gain". This is the number of non-interfering data streams that we can simultaneously send to the clients, per time-frequency slot. Mathematically, the systems DoFs are given by the limit:

$$d_{\text{sum}} = \lim_{\text{SNR}\to\infty} \frac{\text{sum\_rate(SNR)}}{\log \text{SNR}}, \qquad (1)$$

where SNR denotes the ratio between total transmit energy per channel use and noise power spectral density. A system achieving a total DoFs count equal to d_{sum} has high-SNR sum rate of $\text{sum\_rate(SNR)} = d_{\text{sum}} \log(\text{SNR}) + o(\log(\text{SNR}))$.

It turns out that for all schemes considered in this work, the $o(\log(\text{SNR}))$ terms are effectively a constant, i.e. a $O(1)$ term. It is well-known that THP and ZFBF achieve under full CSIT $d_{\text{sum}} = \min\{M, K\}$ DoFs (recall that K is the number of clients and M the number of antennas), which is the maximum achievable. BIA achieves $d_{\text{sum}} = \frac{MK}{M+K-1}$, which is close to $\min\{M, K\}$ when either $M \gg K$ or $K \gg M$. Finally, TDMA (or, equivalently, any orthogonal multiple access scheme) achieves $d_{\text{sum}} = 1$.

The notion of DoFs becomes particularly relevant when fixed QAM constellations are used (e.g., determined by some standard coded modulation scheme), such that the maximum number of information bits per symbol per stream is capped by $\log_2$(constellation size). With fixed size QAM constellations, the achieved spectral efficiency at high SNR approaches $d_{\text{sum}} \times \log_2$(constellation size). We shall see in Section 6 that our testbed, comprising two coordinated APs $M = 2$ and two clients $K = 2$, and transmitting data streams formed by 16QAM symbols (carrying a maximum of 4 information bits per symbol per stream), achieves sum rates close to 8 bps/Hz for the CSIT schemes (THP and ZFBF), sum rates close to $\frac{16}{3}$ bps/Hz for the BIA scheme, and sum rate close to 4 bps/Hz for TDMA, reflecting the DoFs that we expect from the theory.

2.3 Related work with software radio testbeds

The performance of multiuser precoding schemes in real world scenarios has been investigated recently using software radio testbeds. A related study [2] focuses on ZFBF performance when using a single access point with multiple antennas. Achieving multiplexing gains with *distributed* transmit antennas (i.e., using different APs jointly coordinated by a central server through a wired backhaul network) is discussed in [15] for systems that do not provide phase synchronization between the different APs. This precludes the use of the known full-multiplexing achieving precoding schemes. Therefore, in [15] an alternative based on interference alignment and cancellation is proposed. Other streams of work related to spatial multiplexing in wireless systems have considered the use of directional antennas to limit interference [19], the use of additional antennas to transmit concurrently with other nodes without harming the ongoing transmissions [18], and applying multiuser detection techniques to the WiFi uplink [28].

Other related implementations are SourceSync [23], achieving diversity gains through the use of distributed space time coding, and Fine Grained Channel Access [27], which implements practical uplink bandwidth sharing. These systems rely on frame alignment but do not achieve tight phase synchronization. Our solution (see Section 3) achieves sufficiently accurate phase synchronization across the different jointly coordinated APs such that full distributed precoding is possible. This condition is *necessary* to implement full CSIT precoding schemes (in particular THP and ZFBF) for distributed multiuser MIMO. It should be remarked, however, that BIA only requires frame synchronization and sufficient carrier phase stability, and therefore BIA might be supported by system solutions such as SourceSync and Fine Grained Channel Access.

3. DISTRIBUTED MIMO

In this section we briefly describe the distributed MIMO testbed over which we implement THP, ZFBF, and BIA. We pay special attention to the requirement to synchronize the clocks of the APs which transmit simultaneously.

3.1 The Synchronization Requirement

Most precoding schemes designed for multiuser MIMO transmission assume that multiple transmit antennas are hosted on the same AP and share the baseband circuitry and the clocking circuitry which produces the passband signals. In fact, for schemes exploiting full CSIT it is essential that the signal amplitudes and the relative phase and timing offsets of the signals received from different antennas remain unchanged between channel estimation and the actual transmission.

In contrast, schemes such as Blind Interference Alignment [16] and distributed space-time coding through Alamouti encoding [1] and its generalizations do not require any CSIT and can be implemented under relaxed synchronization requirements, limited to frame synchronization and sufficient phase stability over the number of packets spanning a precoding block. For example, for the BIA scheme that we have implemented we need phase stability over three consecutive packets, forming a BIA precoding block, as detailed in Section 5.

Since full CSIT schemes achieve the full sum-DoFs of the distributed MIMO system, and in high SNR the advantage in spectral efficiency can be very significant, it is interesting to explore the possibility of achieving this level of synchronization in practice. We have implemented a testbed that achieves phase synchronization and enables full CSIT precoding for distributed multiuser MIMO. It is also evident that the schemes requiring less strict synchronization can be implemented in our testbed as well.

3.2 Platform Description

We have implemented our synchronization method in the FPGA of the WARP [24] software radio platform. The radio platform is comprised of a central board containing the programmable logic circuitry and up to four daughterboards serving as radio frequency front-ends. The FPGA and the digital-to-analog and analog-to-digital circuitry present on the daughterboards are clocked from a single oscillator called the sampling clock, while the transceivers on the RF front-ends derive their carrier frequency signals used in modulation and demodulation from a second oscillator, called the carrier clock.

All transmitters are connected to a central server through individual gigabit Ethernet links. The server is responsible for the joint encoding of the transmitted signals and for passing the resulting waveforms, in the form of frequency domain soft symbols, to the transmitter radios.

3.3 The Synchronization Method

Phase drifts and oscillators. Any discussion on phase synchronization of distributed wireless transmitters must necessarily start with the mechanisms through which phase errors occur. Digital wireless transmission systems are constructed using a number of clock sources, among which the two most important ones are the sampling clock and the carrier clock. In a typical system, signals are created in a digital form in baseband at a sampling rate on the order of megahertz, then passed through a digital-to-analog converter (DAC). Through the use of interpolators and filters, the DAC creates a smooth analog waveform signal which is

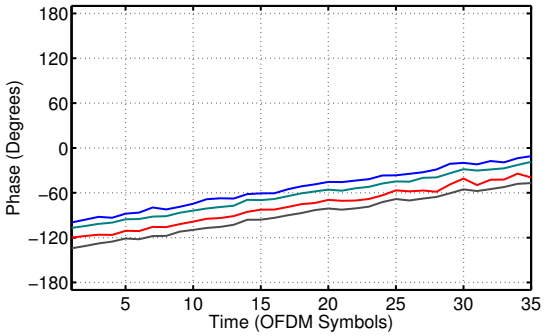


Figure 1: Pilot phases

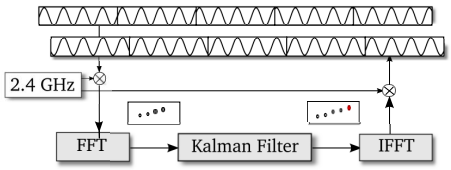


Figure 2: AirSync Schematic. The baseband signals are processed through an FFT which feeds phase estimates into a Kalman Filter. The IFFT produces a phase-adjusted data signal, with the same phase drift as the main transmitter. The modulation and demodulation use the same carrier clock.

then multiplied by a sinusoidal carrier produced by the carrier clock. The result is a passband signal which is then sent over the antenna.

Wireless receivers, in turn, use a chain of signal multiplications and filters to create a baseband version of the passband signal received over the antenna. Some designs, such as the common superheterodyne receivers, use multiple high frequency clocks and convert a signal first to an intermediate frequency before bringing it back to baseband. Other designs simply use a carrier clock operating at the same nominal frequency as the carrier clock of the transmitter and perform the passage from passband to baseband in a single step. We will be focusing on such designs in the ensuing discussion. After baseband conversion, the signal is sampled and the resulting digital waveform is decoded.

There are four clocks in the signal path: the transmitter's sampling clock and carrier clock and the receiver's carrier clock and sampling clock. All four clocks manifest phase drift and jitter. The drift effect, when linear in time and happening at a relatively stable rate, can be assimilated to the presence of a carrier frequency offset.

Denote by $\omega_n = \frac{2\pi n}{NT_s}$ the subcarrier frequency and with ω_c the carrier frequency. Let the timing error of the sampling clock be Δt_s and the timing error of the carrier clock be Δt_c and assume that they are on the same order of magnitude. The phase error due to the sampling clock will be $\omega_n \Delta t_s$ while the phase error due to the carrier clock will be $\phi_i = \omega_c \Delta t_c$. Since ω_c is much greater than ω_n, the dominant phase rotation is due to the carrier clock and does not depend on the subcarrier frequency. Moreover, since time errors are additive, if the time error is approximately linear in time (linear clock drift) then the phase error will also be linear in time and almost equal for all subcarriers.

The assumptions behind the above statement are verified by the results presented in Figure 1. We have constructed an experiment in which a transmitter sends several tone signals, i.e. simple unmodulated sine waves, on several different subcarrier frequencies. In the absence of phase drift each of the tone signals would exhibit a constant phase when measured over several OFDM frames. In reality, the phase is not constant and the frame to frame phase shifts of the signals can be measured and recorded. In the figure, phase drifts have been plotted over the duration of a few tens of frames, a time length corresponding to that of a packet transmission according to the WiFi standards. Our experiment confirms what was anticipated above: the drift is indeed linear and does not depend on the subcarrier frequency. This allows us to design a scheme through which the drift can be tracked and predicted.

Design of a synchronization method. In a nutshell, our method consists of performing frame alignment similarly to prior work [23,27], and then broadcasting a reference signal from a master AP, on to which the rest of the APs transmitters will lock. Thanks to the high clock precision of the WARP platform, the carrier offset is sufficiently low to preclude the need for further frequency offset compensation at baseband frequency before decoding the signal through a Fourier transform. In general, when it is needed, such additional carrier estimation and compensation can be performed using standard methods [25]. No matter how small the residual carrier frequency offset is, it will lead to phase changes in the signal received from a transmitter, from one symbol to the next. We chose to track these phase shifts in real-time and compensate them separately for every frame. We base the tracking of the phase drift on pilot tone signals broadcasted by a master transmitter.

In order to reduce self-interference at the secondary transmitters, the tone signals are placed outside the data band, from which they are separated by a guard interval. The secondary transmitters place an analog baseband filter around their data band further limiting their interference with the pilots. Self-interference could have been avoided using a number of other techniques such as nulling by antenna placement [9], digital compensation [11], or simply relying on the OFDMA-like property of a frame aligned system [27] and preventing the secondary transmitters from using the pilot subcarriers.

In addition to tracking the common drift, the secondary APs measure the initial phase of each tone of the master AP in order to obtain absolute estimates, that is the intercepts of the lines describing the phase drifts in Figure 1. For this, the master AP transmits an initial synchronization header containing a set of known channel estimation symbols. The initial phase estimate, combined with the phase drift measured using the pilot signals, suffices to predict the absolute phase of any particular tone.

Figure 2 illustrates the process of creating a phase synchronous signal at the secondary transmitter. The secondary transmitter overhears a packet sent by the primary transmitter and uses the initial pseudo-noise sequence in order to determine the block boundary timing of this packet. Using a discrete Fourier transform the secondary transmitter decodes the successive frames of the incoming packet. It then employs the CORDIC algorithm on the complex-valued received soft symbols in order to obtain their phases in radi-

ans. The phases of the out-of-band pilot signals are tracked throughout the entire packet transmission in order to estimate the phase drift from the primary sender. The measurements from the four different pilots are averaged and passed through a simplified Kalman filter which maintains an accurate estimate and predicts, based on the current drift estimate, the phase drift after the passage of a few further frames.

The phase estimates are used in synthesizing a synchronized signal. The secondary transmitter uses an inverse discrete Fourier transform, whose output frames are timed such that they align with the frames of the main sender's signal. For every subcarrier the secondary transmitter rotates the soft symbol to be sent by an angle corresponding to the subcarrier's estimated phase offset. The result is a tone that, while not having the same phase as the corresponding tone from the main transmitter, follows that tone at a fixed, preknown phase difference.

Since the subcarrier tones of all transmitters are now synchronized, the transmitters act together as a commonly hosted set of MIMO radio-frequency front-ends.

3.4 Centralized joint encoding

By transmitting phase synchronous signals from multiple access points we have created the equivalent of a distributed MIMO transmitter, capable of employing multiuser MIMO precoding strategies in order to transmit to multiple users at the same time. However, the use of multiple access points complicates the design of the transmitter system. For most of the precoding schemes available, the encoding of the waveforms to be transmitted over the antennas must be done jointly, since reaching a single user usually involves transmitting over multiple antennas. While in theory the joint encoding process could be duplicated at each access point given the binary information destined to each user, we chose to do the encoding only once, at a central server and send the resulting waveforms to each access point for transmission[1].

Our central server has an individual gigabit Ethernet connection to each of the WARP radios serving as access points. We divide the downlink time into slots and in each slot schedule for transmission a number of packets destined to various users. For each of the access points, the server computes the waveform of the signal to be transmitted in the next downlink slot. However, it does not perform any phase correction at this point. The only information used in the precoding is the data to be transmitted and the channel state information between each access point antenna and each user antenna. The server assumes that all access points are phase synchronous, like in a normal MIMO system. The server transmits their corresponding waveforms to all secondary transmitters and finishes by sending the last waveform to the primary transmitter. The primary transmitter starts transmitting right away and the secondary transmitters follow.

The design of the phase alignment method ensures its scalability. There is no added overhead for synchronizing a larger number of secondary transmitters.

[1]This approach is practical in enterprise networks where a number of access points are already connected to a common server.

4. PRECODING USING TRANSMITTER CSI

Consider a system with M single-antenna jointly coordinated access points and K single-antenna clients. When using OFDM, the time and frequency selective fading channel is decomposed into subcarriers, where for each OFDM symbol and subcarrier the channel is characterized by a single frequency-domain complex coefficient for each transmit/receive antenna pair. The resulting baseband channel model for a single subcarrier over any given OFDM symbol is described by

$$\mathbf{y} = \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z}, \tag{2}$$

where the vectors $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{z}$ and the matrix matrix $\mathbf{H}$ have complex entries, and $^{\mathsf{H}}$ denotes Hermitian transpose. For the sake of notation simplicity, we neglect in Equation(2) both the OFDM symbol and the subcarrier (time and frequency indices), since they are irrelevant at this point. The received signal sample at client receiver k is the k-th entry of the vector $\mathbf{y}$, the symbol transmitted by access point j is the j-th entry of the vector $\mathbf{x}$, and the (j,k)-th element of the matrix $\mathbf{H}$ indicates the channel coefficient between the j-th access point transmitter and the k-th client receiver. Finally, $\mathbf{z}$ denotes a K-component vector of independent, identically distributed, additive white circularly symmetric Gaussian noise samples (AWGN).

Let $\mathbf{u}$ be a K-vector containing the symbols destined to each user. The transmitted vector $\mathbf{x}$ is obtained as a function of $\mathbf{u}$. The mapping $\mathbf{u} \mapsto \mathbf{x}$ is called *precoding*. When the mapping function depends on the matrix $\mathbf{H}$, assumed known to the transmitters, we say that the scheme makes use of full CSIT. If the mapping is independent of $\mathbf{H}$, then the scheme is "blind". Next, we review the two full CSIT schemes that we have implemented in our testbed: ZFBF (a linear mapping) and THP (a non-linear mapping).

4.1 Zero-Forcing Beamforming

The precoding mapping in ZFBF is given by

$$\mathbf{x} = \mathbf{V}\mathbf{u} \tag{3}$$

where $\mathbf{V}$ is a scaled version of the Moore-Penrose right pseudo-inverse of $\mathbf{H}^{\mathsf{H}}$, normalized in order to have unit-norm columns. Assuming $M \geq K$, [2] the matrix $\mathbf{H}$ has rank K with probability 1 for any practically relevant scattering environment. The pseudo-inverse takes on the form $\mathbf{H}^{\dagger} = \mathbf{H}(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}$ and the precoding matrix $\mathbf{V}$ is given by

$$\mathbf{V} = \mathbf{H}^{\dagger}\mathbf{\Lambda}^{1/2} \tag{4}$$

where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ and

$$\lambda_k = \frac{1}{[(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}]_{k,k}}$$

(inverse of the (k,k) diagonal element of $(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}$). It is immediate to check that $\mathbf{V}$ in (4) has unit norm columns.

With ZFBF precoding, the original channel (Equation 2) reduces to $\mathbf{y} = \mathbf{\Lambda}^{1/2}\mathbf{u} + \mathbf{z}$. Hence, each k-th client sees a separate (spatially decoupled) channel of the form $y_k = \sqrt{\lambda_k}u_k + z_k$. The total transmit power is given by $\mathbb{E}[\|\mathbf{x}\|^2] = \mathbb{E}[\|\mathbf{u}\|^2]\sum_{k=1}^{K} q_k$, where $q_k = \mathbb{E}[|u_k|^2]$ is the power allocated to the k-th data stream. The corresponding achievable sum

[2]If $K > M$, greedy user selection is used to serve a number of clients not larger than M.

rate assuming a Gaussian random coding ensemble is given by

$$R_{\text{sum}}^{\text{zfbf}}(\text{SNR}) = \sum_{k=1}^{K} \log(1 + \lambda_k q_k), \tag{5}$$

where, without loss of generality, we normalize the noise variance per component to 1, and let $\sum_k q_k = \text{SNR}$. The above rate can be maximized with respect to the power allocation to the data streams, i.e., with respect to $q_1, \ldots, q_K$, subject to the constraint that the total power must be not larger than SNR.

In the case of a distributed MIMO system, both inaccuracies in channel state information and phase synchronization errors between the transmitter antennas can lead to power leakage from each client's channel to the others, creating an error floor for downlink transmission. In previous work, during the development of our system, we have measured the amount of leakage arising from this cumulated effects. We have discovered that the residual self-interference power due to imperfect zero-forcing is, on average, about 24 dB below the useful power.

4.2 Tomlinson-Harashima Precoding

In THP the mapping from the data symbol vector $\mathbf{u}$ to the transmitted symbol vector $\mathbf{x}$ is *non-linear*. Consider again the channel model (2). THP imposes a given precoding ordering, and it pre-cancels sequentially the interference of already precoded signals. Without loss of generality, consider the natural precoding ordering to be from 1 to K, and let $\mathbf{h}_k$ denote the k-th client's $M \times 1$ channel vector. Let $\mathbf{H} = \mathbf{QR}$ be the QR factorization of $\mathbf{H}$, such that $\mathbf{R}$ is $K \times K$ upper triangular with real non-negative diagonal coefficients, and $\mathbf{Q}$ is such that $\mathbf{Q}^{\mathsf{H}}\mathbf{Q} = \mathbf{I}$. THP precoding is formed by the concatenation of a linear mapping, defined by the unitary matrix $\mathbf{Q}$, with a non-linear mapping that does the interference pre-cancellation. Let $\hat{\mathbf{u}} = \text{THP}(\mathbf{u})$ denote the non-linear mapping of the data vector $\mathbf{u}$ into an intermediate vector $\hat{\mathbf{u}}$, that will be defined later. The linear mapping component of THP is then given by

$$\mathbf{x} = \mathbf{Q}\mathbf{\Sigma}^{1/2}\hat{\mathbf{u}}, \tag{6}$$

where $\mathbf{\Sigma} = \text{diag}(q_1, \ldots, q_K)$, and q_k is, as before, the power allocated to the k-th stream. It follows that the channel reduces to $\mathbf{y} = \mathbf{L}\mathbf{\Sigma}^{1/2}\hat{\mathbf{u}} + \mathbf{z}$, where $\mathbf{L} = \mathbf{R}^{\mathsf{H}}$ is lower triangular. The signal seen at client k receiver is given by

$$y_k = [\mathbf{L}]_{k,k}\sqrt{q_k}\hat{u}_k + \underbrace{\sum_{j<k}[\mathbf{L}]_{k,j}\sqrt{q_j}\hat{u}_j}_{\text{interference}} + z_k \tag{7}$$

Next, we look at the non-linear mapping $\mathbf{u} \mapsto \hat{\mathbf{u}}$. The goal is to pre-cancel the term indicated by "interference" in Equation 7. Notice that this term depends only on symbols $\hat{u}_j$ with $j < k$. Therefore, the elements $\hat{u}_1, \ldots, \hat{u}_K$ can be calculated sequentially. A simple pre-subtraction of the interference term at each step would increase the effective transmit power.

The key idea of THP is to introduce a modulo operation that limits the transmit power of each precoded stream $\hat{u}_k$. This is defined as follows. Assume that the data symbols u_k are points from a QAM constellation uniformly spaced in the squared region of the complex plane bounded by the interval $[-\tau/2, \tau/2]$ on both the real axis and the imaginary axis. Then, for a complex number s, let s modulo τ be given by $[s]_{\text{mod }\tau} = s - Q_\tau(s)$, where $Q_\tau(s)$ is the point $(n + jm)\tau$ with integers n, m closest to s. In short, $Q_\tau(s)$ is the quantization of s with respect to a square grid with minimum distance τ on the complex plane, and $[s]_{\text{mod }\tau}$ is the quantization error. We let

$$\hat{u}_k = \left[u_k - \frac{\sum_{j<k}[\mathbf{L}]_{k,j}\sqrt{q_j}\hat{u}_j}{[\mathbf{L}]_{k,k}\sqrt{q_k}} \right]_{\text{mod }\tau}. \tag{8}$$

In this way, the symbol $\hat{u}_k$ is necessarily bounded into the squared region of side τ, and its variance (assuming a uniform distribution over the squared region, which is approximately true when we use a QAM constellation inscribed in the square) is given by $\mathbb{E}[|\hat{u_k}|^2] = \tau^2/6$. Letting $\tau = \sqrt{6}$ we have that the precoded symbols have unit energy and that the transmit power for stream k is exactly given by q_k.

Let's focus now on receiver k and see how the modulo precoding can be undone. The receiver scales the received symbol y_k by $[\mathbf{L}]_{k,k}\sqrt{q_k}$ and applies again the same the modulo τ non-linear mapping. Simple algebra then shows that

$$\widehat{y}_k = \left[u_k + \frac{z_k}{[\mathbf{L}]_{k,k}\sqrt{q_k}} \right]_{\text{mod }\tau}. \tag{9}$$

It follows that the interference term is perfectly removed, but we have introduced a distortion in the noise term. Namely, while u_k is unchanged by the modulo operation, since by construction it is a point inside the square, the noise term $\frac{z_k}{[\mathbf{L}]_{k,k}\sqrt{q_k}}$ is "folded" by the modulo operation, i.e., the tails of the Gaussian noise distribution are folded on the squared region. Noise folding is a well-known effect of THP [14].

As far as the achievable rate is concerned, it is possible to show (see [5, 12]) that this is given by

$$R_{\text{sum}}^{\text{thp}}(\text{SNR}) = \sum_{k=1}^{K} \left[\log(1 + |[\mathbf{L}]_{k,k}|^2 q_k) - \log(\pi e/6) \right]_+, \tag{10}$$

where $[\cdot]_+$ indicates the positive part. Again, this sum rate can be optimized with respect to the power allocation $q_1, \ldots, q_K$, subject to the sum power constraint $\sum_{k=1}^{K} q_k \leq \text{SNR}$. The rate penalty term $\log(\pi e/6)$ is the shaping loss, due to the fact that THP produces a signal which is uniformly distributed in the square region (therefore, a codeword of n signal components is uniformly distributed in an n-dimensional complex hypercube). [3]

5. PRECODING WITHOUT TRANSMITTER CSI

In the case when CSIT cannot be reliably acquired, when tight phase synchronization cannot be achieved, or when channel reciprocity does not hold because of hardware limitations, we need to resort to "blind" approaches. The classical choice, already implemented in software radio [23], consists of distributed space time coding (e.g., distributed Alamouti [1]). This, however, improves reliability through diversity, but achieves the single DoF of TDMA. Here, we implemented the BIA scheme proposed in [16], able to achieve $d_{\text{sum}} = \frac{MK}{M+K-1}$. We briefly outline the case of $M = K = 2$,

[3] It should be noticed that the same shaping loss is incurred by any other scheme such as ZFBF, BIA, as well as TDMA if practical QAM constellations are used instead of the theoretical Gaussian coding ensemble.

Slot	$t = t_1$	$t = t_2$	$t = t_3$
$\begin{bmatrix}\text{Tx1 Sends}\\ \text{Tx2 Sends}\end{bmatrix}$	$\mathbf{x}^{[1]} + \mathbf{x}^{[2]} = \begin{bmatrix} u_1^{[1]} + u_1^{[2]} \\ u_2^{[1]} + u_2^{[2]} \end{bmatrix}$	$\mathbf{x}^{[1]} = \begin{bmatrix} u_1^{[1]} \\ u_2^{[1]} \end{bmatrix}$	$\mathbf{x}^{[2]} = \begin{bmatrix} u_1^{[2]} \\ u_2^{[2]} \end{bmatrix}$
User 1 Antenna	A	B	A
User 2 Antenna	A	A	B
User 1 Receives	$y_1(t_1) = \mathbf{h}_{1A}^{\mathsf{H}}(\mathbf{x}^{[1]} + \mathbf{x}^{[2]}) + z_1(t_1)$	$y_1(t_2) = \mathbf{h}_{1B}^{\mathsf{H}}\mathbf{x}^{[1]} + z_1(t_2)$	$y_1(t_3) = \mathbf{h}_{1A}^{\mathsf{H}}\mathbf{x}^{[2]} + z_1(t_3)$
User 2 Receives	$y_2(t_1) = \mathbf{h}_{2A}^{\mathsf{H}}(\mathbf{x}^{[1]} + \mathbf{x}^{[2]}) + z_2(t_1)$	$y_2(t_2) = \mathbf{h}_{2A}^{\mathsf{H}}\mathbf{x}^{[1]} + z_2(t_2)$	$y_2(t_3) = \mathbf{h}_{2B}^{\mathsf{H}}\mathbf{x}^{[2]} + z_2(t_3)$
User 1 Decodes	$\left.\begin{array}{rcl}\widetilde{y_1}(1) & = & y_1(t_1) - y_1(t_3) \\ & = & \mathbf{h}_{1A}^{\mathsf{H}}\mathbf{x}^{[1]} + z_1(t_1) - z_1(t_3) \\ \widetilde{y_1}(2) & = & y_1(t_2) \\ & = & \mathbf{h}_{1B}^{\mathsf{H}}\mathbf{x}^{[1]} + z_1(2)\end{array}\right\} \Rightarrow \hat{\mathbf{x}}^{[1]} = \begin{bmatrix}\mathbf{h}_{1A}^{\mathsf{H}} \\ \mathbf{h}_{1B}^{\mathsf{H}}\end{bmatrix}^{-1}\begin{bmatrix}\widetilde{y_1}(1) \\ \widetilde{y_1}(2)\end{bmatrix}$		
User 2 Decodes	$\left.\begin{array}{rcl}\widetilde{y_2}(1) & = & y_2(t_1) - y_2(t_2) \\ & = & \mathbf{h}_{2A}^{\mathsf{H}}\mathbf{x}^{[2]} + z_2(t1) - z_2(t2) \\ \widetilde{y_2}(2) & = & y_2(t_3) \\ & = & \mathbf{h}_{2B}^{\mathsf{H}}\mathbf{x}^{[2]} + z_2(3)\end{array}\right\} \Rightarrow \hat{\mathbf{x}}^{[2]} = \begin{bmatrix}\mathbf{h}_{2A}^{\mathsf{H}} \\ \mathbf{h}_{2B}^{\mathsf{H}}\end{bmatrix}^{-1}\begin{bmatrix}\widetilde{y_2}(1) \\ \widetilde{y_2}(2)\end{bmatrix}$		

Table 1: Blind Interference Alignment for the 2×2 scenario

when each client has two antennas connected through a switch to a single RF front-end.

The fundamental idea of BIA is to differentiate the users by inducing special signatures in their channel temporal variations. This is obtained by allocating to each user an antenna switching sequence, according to which they demodulate the signal from one of their antennas. Only one antenna in every given slot is used, so that a single RF front-end and demodulation chain are needed.

The scheme that we have implemented sends 2 independent streams per client to two clients, over 3 time slots. Figure 4 contains a sketch of the testbed. Receiver 1 uses the switching sequence A,B,A, indicating that it uses antenna A in slots 1 and 3 and antenna B in slot 2 of a precoding frame formed by 3 slots. Receiver 2 uses the switching sequence A,A,B, with analogous meaning. Denoting by $u_i^{[k]}$ the i-th data symbol of user k, with $i = 1, 2$ and $k = 1, 2$, the BIA scheme transmits $\mathbf{x}^{[1]} + \mathbf{x}^{[2]}$ in the first slot, $\mathbf{x}^{[1]}$ in the second slot, and $\mathbf{x}^{[2]}$ in the third slot, where $\mathbf{x}^{[1]}$ and $\mathbf{x}^{[2]}$ are formed out of the symbol streams as illustrated in Table 4.2. Letting $\mathbf{h}_{kA}$ and $\mathbf{h}_{kB}$ denote the 2×1 channel vectors seen at antennas A and B of user k, we observe that the 2×2 matrix with columns $[\mathbf{h}_{kA}, \mathbf{h}_{kB}]$ has rank 2, and that the channels remain constant over the precoding block spanning 3 slots.

After linear interference cancellation at each client, the achievable sum rate with Gaussian random coding ensembles is given by [16]:

$$R_{sum} = \sum_{k=1}^{K} \frac{\mathbb{E}\left[\log\det\left(\mathbf{I} + \frac{(K+M-1)P}{M^2K}\mathbf{H}_k^H\mathbf{H}_k\right)\right]}{M+K-1} \qquad (11)$$

where for the 2×2 case:

$$\mathbf{H}_k = \left[\frac{1}{\sqrt{2}}\mathbf{h}_{kA}, \mathbf{h}_{kB}\right] \qquad (12)$$

6. PERFORMANCE EVALUATION

6.1 Experimental results

In this section we characterize the performance of the precoding schemes. The first series of experiments studies how the achieved rates vary in respect to variations in the transmitted power, for a single channel realization. We then study the performance of BIA when the channel experiences phase variations. We have focused on the THP and BIA precoding schemes. While we have implemented and tested ZFBF as well, we do not plot the results since the resulting performance values in our setup are quite close to those of THP.

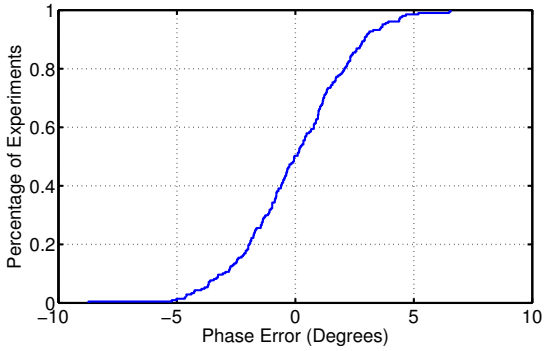


Figure 3: The Precision of the Phase Synchronization. AirSync achieves phase synchronization within a few degrees of the source signal.

Synchronization Accuracy. In this particular experiment we have placed the two transmitters and the two receivers at random locations. We placed a third RF front-end on the secondary sender and configured it in receive mode.

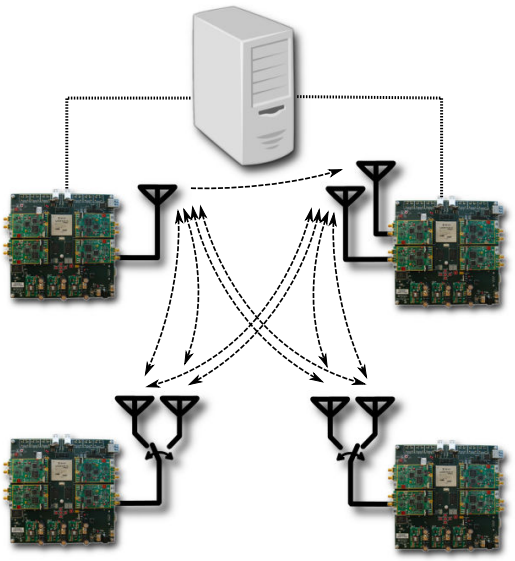


Figure 4: BIA Testbed. When usng Blind Interference Alignment each receiver switches between two antenna modes.

The secondary transmitter samples its own synthesized signal over a wired feedback loop and compares it with the main transmitter's signal. The synchronization circuit measures and records the phase differences between these two signals. Since we use the primary transmission as a reference, in this experiment we do not broadcast the signal synthesized by the secondary transmitter in order to protect the primary transmission from unintended interference. We note that the use of a third RF front-end is not needed in the general case.

We have modified the synchronization circuit to produce a signal that is not only phase synchronous with that of the primary transmitter but has the exact same phase when observed from the secondary transmitter. To achieve this, the circuit estimates the phase rotation that is induced between the DAC of the secondary transmitter and the ADC through which the synthesized signal is resampled. It then compensates for this rotation by subtracting this value from the initial phase estimate. It is worth noting that this rotation corresponds to the propagation delay through the feedback circuit and is constant for different packet transmissions, as determined through measurements.

Figure 3 illustrates the CDF of the synchronization error between the secondary transmitter and the primary transmitter. The error is measured on a frame-to-frame basis using the feedback circuit. In decimal degree values, the standard deviation is 2.37 degrees. The 95th percentile of the synchronization error is at most 4.5 degrees.

Achievable rates. We have used the testbed topology illustrated in Figure 4 throughout our experiments, placing the receivers in arbitrary locations in a closed environment. In order to compare the performance of THP and BIA to the one of a typical TDMA system, we introduced a third transmission scheme, in which instead of multiuser precoding we transmit to one user at a time from the closest access point. In this scheme, transmissions to different users happen in a time-shared manner, just like in 802.11. As opposed to

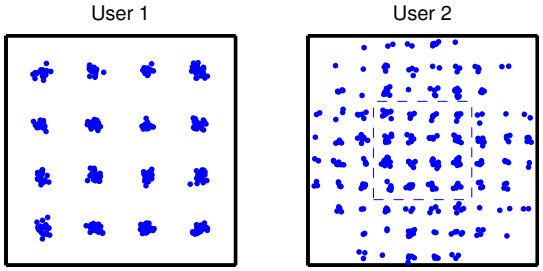


Figure 5: The scattering diagram for Tomlinson-Harashima Precoding.

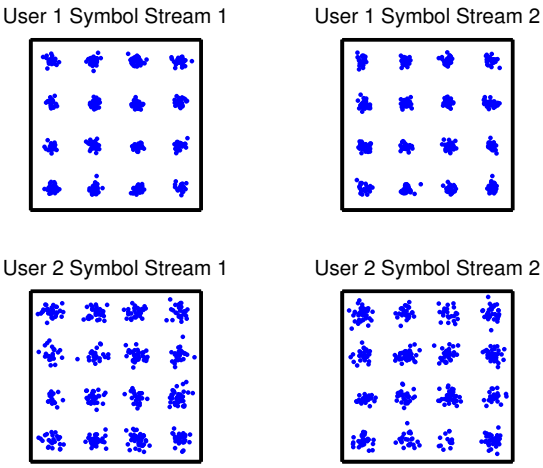
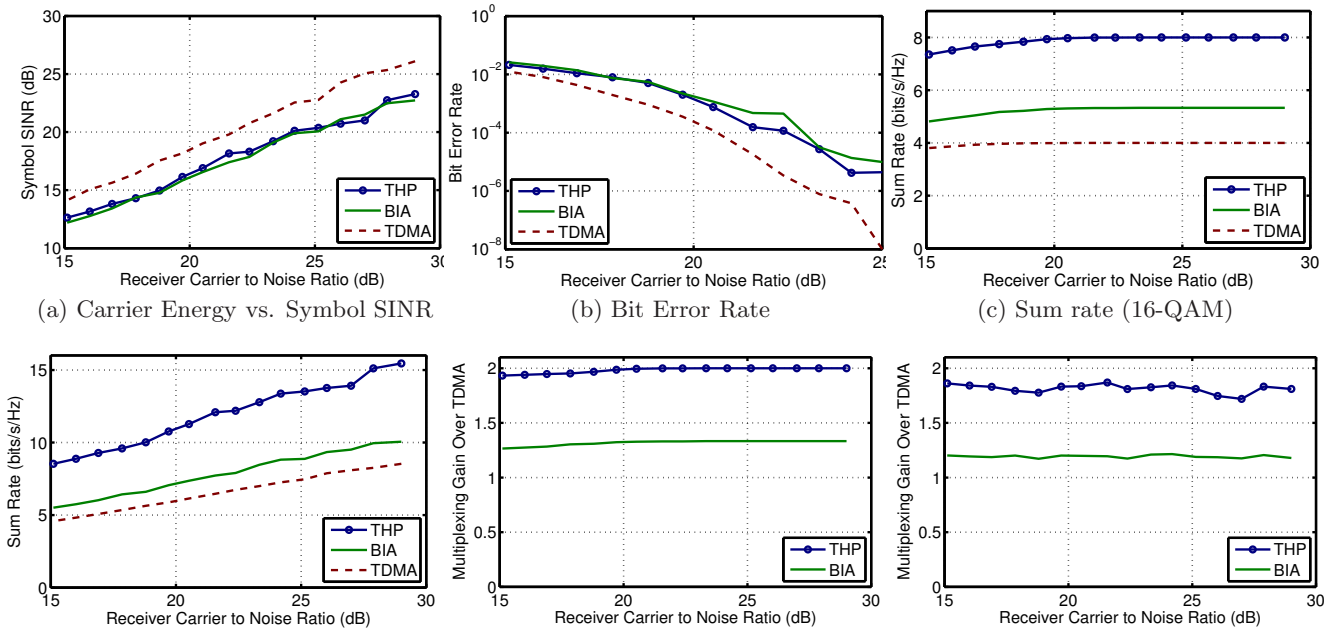


Figure 6: The scattering diagram for Blind Interference Alignment.

802.11, we assume that different access points do not collide when doing channel access, i.e. they perform perfect downlink scheduling. We investigate the sum rates achievable during downlink transmission. The unit of measure is the number of bits per second per Hertz (bps/Hz) transferred by each scheme, where the comparison was done looking only at the portion of the bandwidth used for data transmission (i.e. we considered only the data carriers and ignored the overhead of null carriers, pilots and cyclic prefix). Since the OFDM framing for all three schemes is identical and similar to the one of 802.11, we obtain a fair comparison of their throughputs.

We have varied the transmitters' signal powers in a proportional way, trying to obtain a typical range of SNRs at the receivers. The receive-side SNR values span the typical high range encountered in WiFi signal transmission, from 15 dB to 30 dB. The received SNR values (or carrier to noise ratios) in our figures were estimated using non-precoded and non-synchronized isotropic broadcasts, measuring the raw received power and comparing it to the receiver noise. The same levels of total transmit power were used in the precoded synchronous transmissions.

We evaluate the SINR (Signal to Noise plus Interference Power Ratio) values of the different symbols streams decoded by the receivers. Determining the symbol SINR values requires more effort in our scenario than in classic point-to-point transmission. Since our system is susceptible to power leakage from one stream to another, we would like to con-



(a) Carrier Energy vs. Symbol SINR (b) Bit Error Rate (c) Sum rate (16-QAM)

(d) Sum rate (Gaussian codes) (e) Multiplexing gain (16-QAM) (f) Multiplexing gain (Gaussian codes)

Figure 7: Experimental Results

tinuously transmit over all channels in order to assess the impact of interference.

To this end we sampled each symbol stream using symbols chosen from a relatively sparse QAM-16 constellation. We measured the variance of the constellation points on the receiver side in order to determine the sum of the noise and interference powers. The amplitude of the constellation reflects the received signal power. At the high SNR values present in our system, the clusters of constellation points are spaced sufficiently to allow for an accurate mapping of the received symbols to constellation points. In order to assess the effects of interference produced by streams that follow other encodings, we have, in some experiments, fixed a QAM-16 constellation on one symbol stream while employing symbols chosen according to a Gaussian or uniform distribution on the other stream. Our results have shown that at the low interference levels measured, none of the statistics collected shows considerable variance depending on the type of interference.

Figure 7a presents the SINR values for symbols received when using each of the three precoding schemes. Figure 7b illustrates the inferred symbol error rates for the QAM-16 constellation transmitted. It can be easily seen that the THP and BIA curves closely follow the TDMA curve, with only a few dB difference.

Figure 7c presents the sum rate achievable by the three different schemes (THP, BIA and plain TDMA) for different levels of the total transmit power, when employing a capacity achieving code on top of the transmitted QAM-16 constellation. Figure 7e presents the relative gains of THP and BIA over TDMA. It can be easily seen that each scheme quickly saturates at the maximum rate of 4 bits/DoF. Since THP and BIA provide extra degrees of freedom, they achieve their theoretical multiplexing gain over TDMA.

We would like to know how the quality of the resulting symbol streams affects the achievable rates. To this end we have estimated the rates achievable when using capacity-achieving codes instead of the QAM-16 modulation. Figure 7d presents the resulting sum rates and Figure 7f presents the multiplexing gains. THP achieves an average increase in sum rate of 85%. While this may seem shy of the theoretical achievable multiplexing gain of 2, we must remember that THP allocates power among two degrees of freedom, while TDMA allocates its whole transmitted power to a single transmitter. The second reason for this discrepancy is the shaping loss present in the rate calculation in the case of THP, which was indicated in Equation 10.

The average gain for BIA is 22%. Again, the transmitted power is distributed between the two transmitters. Additionally, as mentioned in Section 5, BIA suffers from noise enhancement, which affects the received symbols.

In the case of a distributed MIMO system, we would expect that phase synchronization error could lead to random rotations of the received soft symbols. We investigated this effect by comparing the variance of soft symbols corresponding to constellation points of different amplitudes. We would expect that due to random rotations, the variance of the outer constellation points would be higher. However, our measurements could not identify such an effect for any of the transmission schemes.

Since BIA does not provide the transmitter with channel state information, to allow it to guess an appropriate transmission rate, it is interesting to find out by how much the received symbol quality is affected by small variations in the positioning of the antennas. Such an effect is analogous to fast fading, where small phase changes affect the channel amplitude at different frequencies. We have conducted an experiment in which we have varied the transmitter antenna positions within one wavelength of their initial position and measured the channel SINR for the two user symbols. The

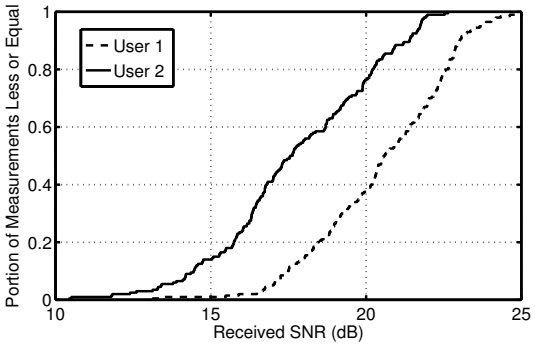


Figure 8: The cumulative distribution function of received SNRs under the Blind Interference Alignment Scheme.

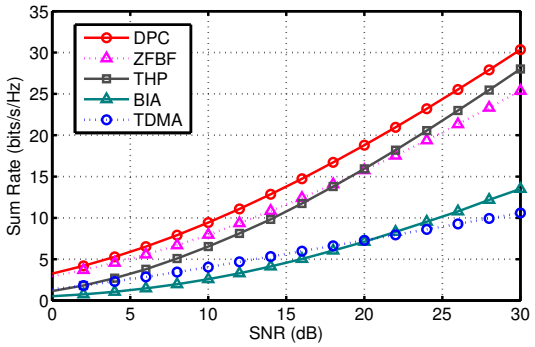


Figure 9: Simulation Results. Sum rate for four users and four access points.

CDFs of the resulting SINR distributions are shown in Figure 8. The high variance of the distribution has profound implications on the design of a coding and medium access scheme for BIA, as will be discussed in Section 7. The higher SNR present in one of the CDFs can be easily explained by the fact that the two symbols are transmitted by antennas placed on different transmitters. The placement of the users relative to the corresponding transmitter determines each symbol's average power.

6.2 Simulation results

To further evaluate the different schemes that were implemented on our testbed, we have also performed off-line simulations to compare them with respect to information theoretic benchmarks. We collected a large set of channel matrices measured at the receivers of our testbed, based on the downlink pilot symbols, and we calculated the achievable sum rate of the various precoding schemes assuming ideal Gaussian random coding, and averaging over the ensemble of measured channel matrices. Specifically, we compare the sum rates between the capacity achieving theoretic DPC, THP and ZFBF, BIA and finally TDMA. In Fig. 9 we can see the sum rates plotted against the SNR (in dBs) at the transmitter for $K = 4$ users and $M = 4$ antennas.

For the full CSIT schemes, we observe that both THP and ZFBF approach the optimal theoretic capacity and show a constant gap of $2.5 - 5$ dB with respect to the system sum capacity (achieved by DPC), at sufficiently large SNR. Notice that THP suffers from the shaping loss even in the case of ideal coding, while ZFBF with ideal coding does not. This makes THP outperform ZFBF only in very high SNR. However, in a real system implementation, coded modulation schemes based on QAM constellations would be used also with ZFBF, and these suffer from the same shaping loss of THP. Therefore, in an actual implementation, the performance of ZFBF would decrease also by roughly 0.5 bit/s/Hz per stream. Thus, the effective gain of THP over ZFBF in an actual implementation can be significant, even at moderate SNR. This is consistent with the testbed experiments shown in Section 6.1, where symbols from a 16QAM constellation were used.

Next, we compare the two schemes that require no CSIT. First, notice that in order for BIA to achieve the theoretic DoFs, we have to be in the high SNR region. It is therefore expected that BIA's sumrate will suffer from the noise enhancement until we reach a high SNR. However, we again remark that these conclusions driven by the use of ideal Gaussian coding may be misleading. For example, when the system is constrained to use constellations of fixed size, as observed before, the advantage of BIA over TDMA can be significant. For example, in the results of Section 6.1, we see that BIA achieves almost a rate 4/3 times larger than TDMA.

7. MEDIUM ACCESS

Our performance evaluation has shown that multiuser transmission at high data rates is possible and that we can approach the theoretical multiplexing gains in a real-world scenario. In the following we discuss issues pertaining to the MAC layer design for the full CSIT and no CSIT cases.

MAC design for full CSIT schemes. As opposed to standard 802.11, distributed multiuser MIMO transmission requires a high degree of coordination between the APs. Part of that coordination, necessary for obtaining frame alignment and phase synchronization, has been detailed in Section 3. The precoding schemes of Sections 4 and 5 require joint precoding of the packets to be transmitted to the different users. It results that the system design of a distributed MIMO system is necessarily more centralized than the current enterprise WLAN design.

We envision a system that, at the MAC layer, resembles more current cellular deployments than the standard CSMA/CA-based large wireless deployment. The downlink and uplink phases of the system should be completely separated in a Time Division Duplexing fashion. In the downlink phase, there will be almost no random behavior, since downlink transmissions are centrally planned by the server and then executed by the APs, see [3] for a very recent study of this matter. With full CSIT, the clients can be selected and jointly precoded on the basis of the knowledge of the system channel matrix $\mathbf{H}$ (see Equation (2)). In particular, the coding rate allocated to each data stream can be dynamically adapted depending on the effective SINR of the stream. This is in line with what is currently done in today's WLAN standards [20], that make use of adaptive coding and modulation based on a family of coded-modulation schemes.

MAC design for schemes with no CSIT. When CSIT is not available, the AP adapts its transmission rate (i.e., its coded-modulation scheme) on the basis of some average signal quality level, reported from the clients through a Channel Quality Indicator (CQI) feedback message, or sim-

ply obtained by measuring the quality of the uplink packets. The CQI depends on the *average* SINR, rather than on the instantaneous realization of the channel matrix **H**, and allows to adapt to coarse and slow signal strength variations (e.g., due to distance and walls attenuation). We assume that for practical reasons, obtaining a quality measure in a timely manner is not possible. For example, in the case of a moving receiver, the symbol SNR figures for BIA vary widely due to phase changes (see Section 6). It follows that the channel quality is, as far as the transmitter is concerned, a random variable with quite high variance and that the rates supported may be very different from one packet transmission to the next. We may choose one of two strategies in order to deal with the rate uncertainty: a conservative rate adaptation approach that chooses a small rate for every transmission or an incremental redundancy scheme.

In incremental redundancy schemes, transmission occurs in blocks. An encoder produces several consecutive encoded versions of a block which are sent over the channel to the decoder. Each successive pass increases the amount of mutual information transfered. At some point the decoder accumulates sufficient information to decode the block and verifies its correctness through a checksum check. At this point the decoder informs the transmitter, through an acknowledgment, that it can proceed to the following information block.

Rateless codes have been proposed as efficient building blocks for incremental redundancy systems. Raptor codes [26], for example, utilize a belief propagation-based decoder and implement the process described above. However, Raptor codes, along with a large class of rateless codes, must be tuned depending on the channel SNR figure in order to effectively approach capacity. More recent approaches do not suffer from this shortcoming. Strider [17] is able to operate in the presence of unknown amounts of interference and when paired with a capacity achieving code comes very close to the actual channel rate. Spinal codes operate with a very dense constellation suitable to many channel conditions [21]. The spinal code decoder makes use of the channel SNR figure, however this quantity can be estimated efficiently [8] for each block transfer and used appropriately in symbol likelihood estimation.

8. DISCUSSION AND FUTURE WORK

In this work we have implemented for the first time ZFBF, THP, and BIA in a distributed MIMO setting. All schemes were shown to achieve almost all of the multiplexing gain predicted by theory in our testbed, consisting of two access points and two clients. In the rest of this section we discuss scalability and implementation overhead issues.

Scalability. For full CSIT schemes, the main issues affecting the SINR of distributed multiuser MIMO are the accuracy of channel estimation and the ability to maintain accurate phase synchronization between the APs. Both issues are not affected by the number of clients, but depend on the number and geometry (distances) of the jointly coordinated APs. In future work, we seek to understand how many APs can be jointly coordinated through the synchronization scheme that we have designed. In general, we may envisage that "clusters" of jointly processed APs can operate according to a distributed multiuser MIMO scheme.

Implementation Overhead. All schemes discussed in this paper, namely ZFBF, THP, BIA, as well as TDMA, require downlink pilots, similarly to 802.11. The main difference when it comes to overhead is the need for CSIT. Specifically, ZFBF and THP need to estimate the channel from the uplink in special slots, and rely on reciprocity to get an estimate for downlink. In contrast, BIA and TDMA do not have this need. An additional benefit of BIA is that it requires a single RF chain and very low power, making it ideal for small devices like smartphones. That said, the performance gain from CSI aware schemes is quite significant, as shown in Section 6. When it comes to implementing the MAC, all schemes require scheduling to materialize high rates. It is part of future work to implement in our testbed MAC protocols that utilize efficient schedulers, as well as to implement several incremental redundancy schemes.

9. ACKNOWLEDGMENTS

The authors would like to thank the reviewers and our shepherd, Xinbing Wang, for their assistance in improving the manuscript of this paper.

This work was supported by the Ming Hsieh Institute, the Army Research Laboratory (ARL) Collaborative Technology Alliance (CTA) number W911NF-09-2-0053, Marie Curie grants 256416 and 274523, Cisco Systems under a CRC grant and NSF grant CIF-0917343.

10. REFERENCES

[1] S. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Sel. Areas Commun.*, 16(8):1451–1458, 1998.

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

[3] H. Balan, A. Michaloliakos, R. Rogalin, G. Caire, and K. Psounis. Efficient MAC for distributed multiuser mimo systems. In *Information Theory and Applications Workshop*, La Jolla, CA, Feb. 2012.

[4] A. Bennatan, D. Burshtein, G. Caire, and S. Shamai. Superposition coding for side-information channels. *Information Theory, IEEE Transactions on*, 52(5):1872 – 1889, may 2006.

[5] F. Boccardi, F. Tosato, and G. Caire. Precoding schemes for the mimo-gbc. In *Communications, 2006 International Zurich Seminar on*, pages 10 –13, 0-0 2006.

[6] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Multiuser MIMO achievable rates with downlink training and channel state feedback. *IEEE Trans. Inf. Theory*, 56(6):2845–2866, 2010.

[7] G. Caire and S. Shamai. On the achievable throughput of a multiantenna gaussian broadcast channel. *IEEE Trans. Inf. Theory*, 49(7):1691 – 1706, Jul. 2003.

[8] B. Chen, Z. Zhou, Y. Zhao, and H. Yu. Efficient error estimating coding: feasibility and applications. In *ACM SIGCOMM*, pages 3–14, New Delhi, India, 2010.

[9] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti. Achieving single channel, full duplex wireless communication. In *IEEE MobiCom*, Chicago, IL, 2010.

[10] M. Costa. Writing on dirty paper (corresp.). *IEEE Trans. Inf. Theory*, 29(3):439–441, May 1983.

[11] M. Duarte, C. Dick, and A. Sabharwal. Experiment-driven characterization of full-duplex wireless systems. *CoRR*, abs/1107.1276, 2011.

[12] U. Erez, S. Shamai, and R. Zamir. Capacity and lattice strategies for canceling known interference. *Information Theory, IEEE Transactions on*, 51(11):3820 – 3833, nov. 2005.

[13] U. Erez and S. ten Brink. A close-to-capacity dirty paper coding scheme. *Information Theory, IEEE Transactions on*, 51(10):3417 –3432, oct. 2005.

[14] J. Forney, G.D. and M. Eyuboglu. Combined equalization and coding using precoding. *Communications Magazine, IEEE*, 29(12):25 –34, dec. 1991.

[15] S. Gollakota, S. D. Perli, and D. Katabi. Interference alignment and cancellation. In *ACM SIGCOMM*, Barcelona, Spain, 2009.

[16] T. Gou, C. Wang, and S. Jafar. Aiming perfectly in the dark - blind interference alignment through staggered antenna switching. In *IEEE GLOBECOM*, Dec. 2010.

[17] A. Gudipati and S. Katti. Strider: automatic rate adaptation and collision handling. In *ACM SIGCOMM*, Toronto, Ontario, Canada, 2011.

[18] K. C.-J. Lin, S. Gollakota, and D. Katabi. Random access heterogeneous mimo networks. In *ACM SIGCOMM*, pages 146–157, Toronto, Ontario, Canada, 2011. ACM.

[19] X. Liu, A. Sheth, M. Kaminsky, K. Papagiannaki, S. Seshan, and P. Steenkiste. Pushing the envelope of indoor wireless spatial reuse using directional access points and clients. In *IEEE MobiCom*, Chicago, Illinois, USA, 2010.

[20] A. Molisch. *Wireless communications*. Wiley-IEEE, 2005.

[21] J. Perry, H. Balakrishnan, and D. Shah. Rateless spinal codes. In *ACM HotNets*, Cambridge, Massachusetts, 2011.

[22] J. Proakis and M. Salehi. *Digital communications*. McGraw-Hill, New York, NY, 2007.

[23] H. Rahul, H. Hassanieh, and D. Katabi. SourceSync: a distributed wireless architecture for exploiting sender diversity. In *ACM SIGCOMM*, New Delhi, India, 2010.

[24] Rice University. Rice university WARP project.

[25] T. Schmidl and D. Cox. Robust frequency and timing synchronization for ofdm. *Communications, IEEE Transactions on*, 45(12):1613 –1621, dec 1997.

[26] A. Shokrollahi. Raptor codes. *IEEE Trans. Inf. Theory*, 52(6):2551 –2567, Jun. 2006.

[27] K. Tan, J. Fang, Y. Zhang, S. Chen, L. Shi, J. Zhang, and Y. Zhang. Fine-grained channel access in wireless LAN. In *ACM SIGCOMM*, New Delhi, India, 2010.

[28] 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 *IEEE MobiCom*, pages 49–60, Beijing, China, 2009.

[29] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of gaussian mimo broadcast channels. *Information Theory, IEEE Transactions on*, 49(10):2658 – 2668, oct. 2003.

[30] C. Windpassinger, R. Fischer, T. Vencel, and J. Huber. Precoding in multiantenna and multiuser communications. *Wireless Communications, IEEE Transactions on*, 3(4):1305 – 1316, Jul. 2004.

[31] T. Yoo and A. Goldsmith. On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming. *IEEE J. Sel. Areas Commun.*, 24(3):528 – 541, Mar. 2006.