ACHIEVING HIGH DATA RATES IN DISTRIBUTED MIMO SYSTEMS

by

Horia Vlad Balan

A Dissertation Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(COMPUTER ENGINEERING)

August 2013

to my parents

Table of Contents

List of Figures

List of Tables

Preface

Multiple Input Multiple Output (MIMO) transmission increases the spectral efficiency of wireless communication by multiplexing multiple streams of data, taking advantage of the spatial diversity characteristic to large wireless networks. The number of multiplexed streams and the achieved degree of spectral reuse have been so far limited by the constraint that all simultaneous transmissions had to occur from a single location, over a limited number of closely-placed antennas.

The current thesis presents a technique for realizing a distributed, coherent transmission system capable of achieving the full spatial multiplexing gain. The synchronization among different transmitters is realized over the air, replacing the commonly-used back-end network capable of clock signal distribution with a much more cost-effective data network. The reduction in deployment cost paves the way for distributed MIMO as a viable technology for enterprise wireless networks and cellular offloading, beyond its current application in single base-station distributed antenna systems. Past advancing the state of the art in wireless system construction, we hope that the techniques presented here will be followed by more advanced implementations, rendering high data rate wireless communication into a day-to-day commodity.

I would like to acknowledge the assistance of my thesis adviser, Professor Konstantinos Psounis and of my collaborators, Professor Giuseppe Caire, Ryan Rogalin and Antonios Michaloliakos during the development of the material presented in this thesis.

Los Angeles, CA
May, 2013
                                                                    Horia Vlad Balan

CHAPTER 1
Introduction

The enormous success of advanced wireless devices such as tablets and smartphones is pushing the demand for higher and higher wireless data rates and is causing significant stress to existing networks. While new standards (e.g., 802.11n/ac and 4G) are developed almost every couple of years, novel and more radical approaches to this problem are yet to be tested. The fundamental bottleneck is that wireless bandwidth is simply upper bounded by physical laws, in contrast to wired bandwidth, where putting new fiber on the ground has been the de-facto solution for decades. Advances in network protocols, modulation and coding schemes have managed steady but relatively modest spectral efficiency (bit/s/Hz) improvements. 4G-LTE, for instance, offers two to five times better spectral efficiency than 2.5G-EDGE. Denser spectrum reuse, i.e., placing more access points per square mile, has the potential to successfully meet the increasing demand for more wireless bandwidth [cis12]. On the other hand, in contrast to conventional cellular systems, a very dense infrastructure deployment cannot be carefully planned and managed for reasons pertaining to scale and cost. Therefore, the multiuser interference between different uncoordinated Access Points (APs) represents the main system bottleneck to achieve truly high spectral efficiency.

In theory, the ultimate answer to this problem is distributed multiuser MIMO (also known as "virtual MIMO"), where several (possibly multi-antenna) APs are connected to central server and operate as a large distributed multi-antenna base station. When

1
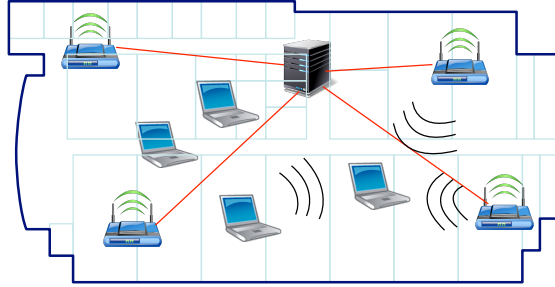
Figure 1.1: **Enterprise WiFi and Distributed MIMO.** Multiple access points connected to a central server through Ethernet (red lines) coordinate their transmissions to several clients by using distributed MIMO.

using joint decoding in the uplink and joint precoding in the downlink, all transmitted signal power is useful, as opposed to conventional random access scenarios (e.g., carrier-sense) which waste power through interference. This approach is particularly suited to the case of an enterprise network (e.g., a WLAN covering a conference center, an airport terminal or a university), or to the case of clusters of closely spaced home networks connected to the Internet infrastructure through the same cable bundle.

Distributed multiuser MIMO (MU-MIMO) is regarded today mostly as a theoretical solution because of some serious implementation hurdles, such as providing accurate timing and carrier phase reference to all jointly coordinated APs and the ability to perform efficient joint precoding at a central server connected to the APs through a wired backhaul of limited capacity.

We consider a typical enterprise network as illustrated in Figure 1.1. Since in such networks the wired backhaul is fast enough to allow for efficient joint processing at the server (see Section 3.1), the major obstacle to achieving the full distributed MU-MIMO multiplexing gain is represented by the lack of synchronization between the jointly processed APs. The perceived difficulty of this task has led some researchers to believe that it is practically impossible to achieve the full multiplexing gain in the context of distributed MU-MIMO.

In this thesis, we present two real-world testbed implementations which achieves the theoretical optimal gain by correcting, in real time, the instantaneous phase offsets

between geographically separated access points. The first one of these is a reference-based system, called *AirSync*, which synchronizes the access points surrounding a reference access point, enabling for the first time multiplexed transmissions from multiple access points. The second of these is a system that exploits channel reciprocity, called *DistSync*, and which implements a number of distributed synchronization techniques in order to enable spatial multiplexing on a much larger scale.

In a nutshell, AirSync tracks the instantaneous phase of a pilot signal broadcasted by a reference AP (the master AP), and predicts the phase correction across the duration of the MU-MIMO precoded slot in order to de-rotate the complex baseband signal samples. This enables APs to maintain phase coherence, which is necessary for MU-MIMO precoding. Notice that each AP *transmits* the precoded data signal and *receives* the master AP pilot signal simultaneously. This is accomplished by dedicating one antenna per AP to pilot reception, while the others are used for MU-MIMO transmission. We have implemented AirSync as a digital circuit in the FPGA of the WARP radio platform [Ric]. We have also implemented *Zero-Forcing Beamforming* [CS03] and *Tomlinson-Harashima Precoding* [WFVH04], two popular MU-MIMO precoding schemes widely investigated, from a theoretical viewpoint, in the literature. Using Airsync in a testbed consisting of eight WARP radios, four acting as access points connected to a central server and four acting as clients, we have shown that the theoretical optimal gain of multiuser MIMO is achievable in practice.

DistSync separates the problems of phase-locking a group of transmitters and of estimating and using the resulting fixed channel realization. It uses a distributed measurement scheme to determine and eliminate all carrier frequency offsets between APs, making time phase-locked for a downlink timeslot duration, to the accuracy of the estimates. Subsequently, it efficiently estimates the channel, precodes and transmits the client signals before the coherence time of the channel lapses. DistSync has been similarly implemented as a realtime digital circuit supported by a software stack and a network infrastructure capable of computing precoding information at a central server and returning it to the transmitters within the channel coherence time.

Multiuser downlink transmission involves taking into account channel state information and reacting to an unknown amount of interference arising from faulty measurements or imperfect synchronization. Consequently, we investigate protocol designs for the MAC layer in distributed MU-MIMO systems, and the possibility of advancing past the design of the 802.11ac medium access layer [80212],through the use of Incremental Redundancy rateless coding at the physical layer [Sho06] [PBS11].

Several extensions and improvements of this basic layout are possible, and are discussed in the thesis. For example, in the case of AirSync, the master pilot signal range can be extended by regenerating and repeating the pilot signal at different frequencies. Also, we discuss the possibility of estimating the downlink channel matrix from training signals in the uplink and exploiting the physical channel reciprocity of Time-Division Duplexing (TDD). In particular, this latter issue is discussed in detail in the recent work [SYA^{+}12] for a large centralized MU-MIMO system where all transmit antennas are clocked together (both timing and carrier frequency) and therefore are perfectly synchronous. Thanks to DistSync, the same approach for calibration can be used in a distributed MU-MIMO system, although in the present implementation we use a more conventional downlink training and feedback configuration. As opposed to AirSync, DistSync does not aim to eliminate random phases from the channel matrix, but to work with the composite channel. We explore in detail the overheads of our synchronization schemes and describe techniques of reducing the amount of information transferred, over the air and over the wire, to a minimum.

While recently there have been a number of very interesting and important works in which some of the gains of multiuser MIMO have been shown (see Section 2.3 for more details) none of these has managed to achieve both timing and carrier phase synchronization between remote transmitters precise enough to implement MU-MIMO with optimal multiplexing gain in the distributed transmitters scenario. It is also worthwhile to notice that while single-user MIMO and single AP (centralized) MU-MIMO can offer multiplexing gains for a given configuration of transmit and receive antennas, these can be further increased by extending the cooperation to the distributed case, provided that the transmitters (the APs) can be synchronized with sufficient accuracy. Therefore, our

approach can potentially provide *additional multiplexing gains* on top of an existing configuration.

A final part of the thesis deals with the issue of message passing in wireless networks. Message passing is the basic building block of many proposed schemes for apportioning the wireless capacity of a network among users and flows of data. We introduce a physical layer primitive that makes many of these techniques practical in real-world deployments and evaluate its impact on two example congestion control and medium access schemes targeted to wireless environments. Message passing complements the synchronization techniques presented in the rest of the thesis and makes distributed MIMO a viable alternative for adhoc deployments as well [OJTL10, OLT13].

■ 1.1 Contributions

This main contribution of this thesis is providing a clear implementation path for distributed multiuser MIMO in large, distributed wireless networks. It bridges the distance between the theoretical developments of the past years and their realization in practical wireless systems. By doing so, it enables distributed multiuser MIMO to realize its full potential as a practical, efficient way of increasing spectral reuse in enterprise wireless networks.

As a structure for building up practical implementations, we formalize the problem of distributed transmission and outline the challenges to be overcome, i.e. the physical effects present when using multiple transmission sources. A solid understanding of these effects, their scale and their influence on the received signals is a necessary prerequisite for designing synchronization schemes that address them.

The dissertation introduces two methods for realizing coherent downlink multiplexed transmissions: AirSync uses a single access point as a reference source for synchronizing a number of neighboring transmitters; DistSync used distributed synchronization in order to phase-lock a more distributed set of transmitters. Each one of these methods imposes its own limitation on the type of channel estimation that can be used and brings with it its own timing requirements.

We develop, in each case, complete protocols for the physical layer and investigate the design space of medium access layer protocols. Our system designs are capable of exploiting to the fullest the achieved multiplexing gains. In particular, we show that uplink channel estimation is a particular strong match for DistSync, allowing the sort of timeslot-by-timeslot operation that counters the many effects of imperfect synchronization and allows for considerably less centralization in the operation of the network. We study how different encoding schemes perform in a distributed MIMO environment, subject to the uncertainties of imperfect channel estimation and imperfect synchronization and make a strong case for incremental redundancy, rate-achieving schemes as a match to this physical layer.

Each of these systems has been implemented in actual hardware and their performance has been thoroughly evaluated. The experiments prove that the level of synchronization achieved allow them to approach the theoretical optimal gains without sacrificing a sensible amount of bandwidth due to protocol overheads.

■ 1.2 Organization

We conclude this introduction by providing a brief outline of the rest of the thesis.

Chapter 2 introduces distributed MIMO transmission, provides an overview of a number of common precoding schemes, some of which require transmitter-side channel state information and some of which operate without knowledge of the channel. It continues by establishing a theoretical model of distributed MIMO transmission that takes into account different oscillator-induced effects and it experimentally evaluates its accuracy. It closes by detailing the body of related work relevant to the problem, both theoretical and experimental.

Chapter 3 introduces AirSync, a single reference-based synchronization method for wireless transmitters. This method achieves phase coherence among several transmitters located within the range of the reference access point. It starts by presenting the construction of the synchronization scheme and the details of its implementation. It

continues by evaluating its performance in an actual deployment. Finally, it describes a medium access layer compatible with its operation.

Chapter 4 introduces DistSync, a scalable synchronization scheme aimed at large-scale deployments, beyond the effective range of a single access point. DistSync operates by running a distributed synchronization algorithm which attempts to phase-lock the APs in the network for the duration of a downlink transmission slot. The oscillator characteristics that influence the design of the synchronization scheme are presented in detail, along with experimental data that describes the changing oscillator behavior in time. We describe a synchronization scheme capable of compensating frequency drift with a high degree of precision while imposing a low wireless-transmission overhead on the network. The downlink transmission slot makes use of the short-term constancy of the channel and uses an efficient estimation algorithm followed by fast precoding in order to determine and use this ephemeral channel instantiation. The chapter continues by discussing the use of estimators in order to increase the precision of timing-specific parameter estimation and to decrease the amount of data sent over the wireless back-haul. The performance of the synchronization scheme and the achieved degree of multiplexing are quantified through an experimental implementation.

Chapter 5 introduces tags, wireless primitives aimed towards message passing. Tags enable multiple distributed control schemes in wireless networks. It discusses the construction of tags, characterizes their performance theoretically and experimentally and presents sample applications in scheduling and congestion control.

CHAPTER 2
Theoretical Considerations

The current chapter introduces multiuser MIMO communication and describes a number of commonly discussed multiuser precoding schemes. The focus is, on one side, on schemes that require full channel state information at the transmitter (CSIT) and, on the other side, on schemes that forgo channel state information at the expense of realizing a smaller number of degrees of freedom. We establish a theoretical model of a distributed MIMO transmitter that includes all oscillator-induced effects and test its accuracy through a series of microbenchmarks. This exposition constitutes a complete statement of the distributed MIMO transmission problem for which the synchronization algorithms presented in the following chapters will try to offer compelling solutions. Finally, we present a review of related work, theoretical and experimental, that paved the way to the current development.

■ 2.1 A Multiuser MIMO Primer

We consider the OFDM signaling format, as in the last generation of WLANs and cellular systems (e.g., IEEE 802.11a/g/n and 4G-LTE [Mol05]). OFDM is a block precoding scheme. One OFDM symbol corresponds to N frequency-domain information-bearing symbols. By inverse FFT (IFFT), an OFDM symbol is converted into a block of N time-domain samples. This block is augmented by the cyclic prefix (CP), i.e., by repeating the

$L \leq N$ last samples at the beginning of the block. The OFDM symbol length N and the CP length L are design parameters. With CP length L, any frequency selective channel with impulse response of length $\ell \leq L + 1$ samples is turned into a cyclic convolution channel, such that, by applying an FFT at the receiver, it is exactly decomposed into a set of N parallel frequency-flat discrete-time channels in the frequency domain. Typical CP length is between 16 to 64 time-domain samples. For example, for a 20 MHz signal, as in IEEE 802.11g, the time-domain sampling interval is 50 ns, so that a typical CP length ranges between 0.8 and 3.2 μs.

In a multiuser environment OFDM has also a significant side advantage: as long as the different users' signals align in time with an offset not larger than $L - \ell$, where L denotes the CP and ℓ is the maximum length of any channel impulse response in the system, their symbols after OFDM demodulation remain perfectly aligned in time and frequency. In other words, the timing misalignment problem between user signals, which in single-carrier systems creates significant complications for joint processing of overlapping signals (e.g., multiuser detection [Ver98], successive interference cancellation [TV05], Zig-Zag decoding [GK08]), completely disappears in the case of OFDM, provided that all users achieve timing alignment within the CP.

In a point-to-point MIMO link with N_r receive and N_t transmit antennas, the time-domain channel is represented by an $N_r \times N_t$ matrix of channel impulse responses. Thanks to OFDM, the channel in the frequency domain is described by a set of channel matrices of dimension $N_r \times N_t$, one for each of the N OFDM subcarriers. An intuitive explanation of the MIMO multiplexing gain can be given as follows: in the high-SNR regime, the receiver observes N_r (noisy) equations with N_t unknown coded modulation symbols on each time-frequency dimension, each of which carries $\sim \log(\mathsf{snr}) + O(1)$ bits, where $O(1)$ indicates constants that depend on the channel matrix coefficients but are independent of SNR. For sufficiently rich scattering, the rank of the channel matrix is equal to $\min\{N_r, N_t\}$ with probability 1. Therefore, using appropriate coding in order to eliminate the effect of the noise, up to $\min\{N_r, N_t\}$ symbols per channel time-frequency dimension can be recovered with arbitrarily high probability, thus yielding the high-SNR capacity scaling $C(\mathsf{snr}) = \min\{N_r, N_t\} \log \mathsf{snr} + O(1)$ bits/s/Hz.

Zero-Forcing Beamforming. In contrast to point-to-point MIMO, in a MU-MIMO system with one M-antennas sender and K single antenna receivers, [1] it is not generally possible to jointly decode all the receivers' observations, since the receivers are spatially separated and are not generally able to communicate with each other. In this case, joint precoding from the transmit antennas must be arranged in order to invert, in some sense, the channel matrix and control the multiuser interference. One of the techniques to achieve this is linear Zero-Forcing Beamforming (ZFBF).

In ZFBF, the transmitter multiplies the outgoing symbols by beamforming vectors such that the receivers see only their intended signals. For instance, let the received signal on a given OFDM subcarrier at user k be given by

$$y_k = h_{k,1}x_1 + h_{k,2}x_2 + \cdots + h_{k,N_t}x_{N_t} + z_k \tag{2.1}$$

where $h_{k,j}$ is the channel coefficient from transmit antenna j to user k and z_k is additive white Gaussian noise. Then, the vector of all received signals can be written in matrix form as

$$\mathbf{y} = \mathbf{H}^\mathsf{H}\mathbf{x} + \mathbf{z} \tag{2.2}$$

where $\mathbf{H}$ has dimension $M \times K$. Assuming $K \leq M$, we wish to find a matrix $\mathbf{V}$ such that $\mathbf{H}^\mathsf{H}\mathbf{V}$ is zero for all elements except the main diagonal, that is $\mathbf{H}^\mathsf{H}\mathbf{V} = \mathbf{\Lambda}^{1/2} = \mathrm{diag}(\sqrt{\lambda_1},\ldots,\sqrt{\lambda_K})$. Letting $\mathbf{x} = \mathbf{V}\mathbf{u}$, where $\mathbf{u}$ is the vector of coded-modulation symbols to be transmitted to the clients, we have

$$\mathbf{y} = \mathbf{H}^\mathsf{H}\mathbf{V}\mathbf{u} + \mathbf{z} = \mathbf{\Lambda}^{1/2}\mathbf{u} + \mathbf{z}, \tag{2.3}$$

so that each receiver k sees the interference-free Gaussian channel $y_k = \sqrt{\lambda_k}u_k + z_k$.

When $\mathbf{H}$ has rank K (which is true with probability 1 for sufficiently rich propagation scattering environments typical of WLANs and for $K \leq M$) a column-normalized

---

[1]We assume single-antenna receivers for simplicity of exposition. The extension to $1 \leq N_r \leq M$ antenna receivers is immediate.

version of the Moore-Penrose pseudo-inverse generally yields the ZFBF matrix. This takes on the form

$$\mathbf{V} = \mathbf{H}(\mathbf{H}^{\mathsf{H}}\mathbf{H})^{-1}\mathbf{\Lambda}^{1/2},$$

where $\mathbf{\Lambda}$ is chosen in order to ensures that the norm of each column of $\mathbf{V}$ is equal to 1, thus setting the total transmit power equal to $\mathrm{tr}(\mathrm{Cov}(\mathbf{Vu})) = \mathbb{E}[\|\mathbf{u}\|^2]$, i.e., equal to the power of the data vector $\mathbf{u}$.

As far as the achievable rate is concerned, since ZFBF converts the MU-MIMO channel into a set of independent Gaussian channels for each user, subject to the sum-power constraint $\mathbb{E}[\|\mathbf{u}\|^2] \leq \mathsf{snr}$, we have immediately that the maximum sum rate of ZFBF is given by

$$R_{\text{sum}}^{\text{zfbf}}(\mathsf{snr}) = \sum_{k=1}^{K} \log\left(1 + \lambda_k q_k\right), \tag{2.4}$$

where q_k denotes the power of the k-th data symbol in $\mathbf{u}$. The above expression can be maximized over the power allocation $\{q_k\}$, subject to the constraint $\sum_{k=1}^{K} q_k \leq \mathsf{snr}$, resulting in the classical water filling power allocation of parallel Gaussian channels [CT91].

Tomlinson-Harashima Precoding. In Tomlinson-Harashima Precoding (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 (2.2). 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. 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 $M \times K$ tall unitary, 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 $\hat{\mathbf{u}} = \mathrm{THP}(\mathbf{u})$ 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

$$\mathbf{x} = \mathbf{Q}\hat{\mathbf{u}}, \tag{2.5}$$

where $\mathrm{Cov}(\hat{\mathbf{u}}) = \boldsymbol{\Sigma} = \mathrm{diag}(q_1, \ldots, q_K)$ and, as before, q_k denotes the power allocated to the k-th data symbol. It follows that the channel reduces to

$$
\begin{aligned}
\mathbf{y} &= \mathbf{H}^{\mathsf{H}}\mathbf{x} + \mathbf{z} \\
&= \mathbf{R}^{\mathsf{H}}\mathbf{Q}^{\mathsf{H}}\mathbf{Q}\hat{\mathbf{u}} + \mathbf{z} \\
&= \mathbf{L}\hat{\mathbf{u}} + \mathbf{z},
\end{aligned}
\tag{2.6}
$$

where $\mathbf{L} = \mathbf{R}^{\mathsf{H}}$ is lower triangular. The signal seen at client k receiver is given by

$$
y_k = [\mathbf{L}]_{k,k}\hat{u}_k + \underbrace{\sum_{j<k}[\mathbf{L}]_{k,j}\hat{u}_j}_{\text{interference}} + z_k.
\tag{2.7}
$$

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 (2.7). 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 and would result in a suboptimal version of the linear ZFBF treated before.

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 $[s]_{\mathrm{mod}\,\tau} = s - Q_\tau(s)$, 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 $[s]_{\mathrm{mod}\,\tau}$ is the quantization error. We let

$$
\hat{u}_k = \sqrt{q_k}\left[u_k - \frac{\sum_{j<k}[\mathbf{L}]_{k,j}\hat{u}_j}{[\mathbf{L}]_{k,k}\sqrt{q_k}}\right]_{\mathrm{mod}\,\tau}.
\tag{2.8}
$$

In this way, the symbol $\hat{u}_k$ is necessarily bounded into the squared region of side $\tau\sqrt{q_k}$, 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 $\mathbb{E}[|\hat{u}_k|^2] = \tau^2/6q_k$. Letting $\tau = \sqrt{6}$ we have that the precoded symbols have the desired power 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

$$\widehat{y}_k = \left[u_k + \frac{z_k}{[\mathbf{L}]_{k,k}\sqrt{q_k}}\right]_{\text{mod }\tau}. \tag{2.9}$$

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 [FE91].

As far as the achievable rate is concerned, it is possible to show (see [BTC06, ESZ05]) that this is given by

$$R_{\text{sum}}^{\text{thp}}(\text{snr}) = \sum_{k=1}^{K}\left[\log(1 + |[\mathbf{L}]_{k,k}|^2 q_k) - \log(\pi e/6)\right]_{+}, \tag{2.10}$$

where $[\cdot]_+$ indicates the positive part. Again, this sum rate can be optimized with respect to the power allocation $\{q_k\}$, 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). [2]

Blind Interference Alignment. 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.

---

[2]It should be noticed that the same shaping loss at high SNR is incurred by any other scheme, including plain CSMA, when practical QAM constellations are used instead of the theoretically optimal Gaussian coding ensemble.
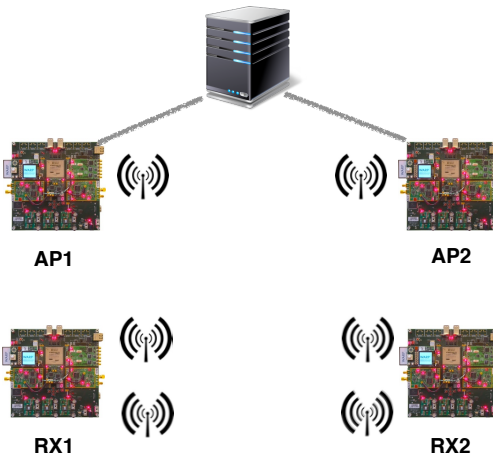
Figure 2.1: **BIA Testbed.** When using Blind Interference Alignment each receiver switches between two antenna modes.

The scheme that we have implemented sends 2 independent streams per client to two clients, over 3 time slots. Figure 2.1 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 2.1. 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 [GWJ10]:

$$R_{sum} = \sum_{k=1}^{K} \frac{\mathbb{E}\left[\log \det \left(\mathbf{I} + \frac{(K+M-1)P}{M^2K}\mathbf{H}_k^H\mathbf{H}_k\right)\right]}{M + K - 1} \tag{2.11}$$

where for the 2×2 case:

$$\mathbf{H}_k = \left[\frac{1}{\sqrt{2}}\mathbf{h}_{kA}, \mathbf{h}_{kB}\right] \tag{2.12}$$

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	$$\left. \begin{aligned} \widetilde{y_1}(1) &= y_1(t_1) - y_1(t_3) \\ &= \mathbf{h}_{1A}^{\mathsf{H}}\mathbf{x}^{[1]} + z_1(t_1) - z_1(t_3) \\ \widetilde{y_1}(2) &= y_1(t_2) \\ &= \mathbf{h}_{1B}^{\mathsf{H}}\mathbf{x}^{[1]} + z_1(2) \end{aligned} \right\} \Rightarrow \hat{\mathbf{x}}^{[1]} = \begin{bmatrix} \mathbf{h}_{1A}^{\mathsf{H}} \\ \mathbf{h}_{1B}^{\mathsf{H}} \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{y_1}(1) \\ \widetilde{y_1}(2) \end{bmatrix}$$		
User 2 Decodes	$$\left. \begin{aligned} \widetilde{y_2}(1) &= y_2(t_1) - y_2(t_2) \\ &= \mathbf{h}_{2A}^{\mathsf{H}}\mathbf{x}^{[2]} + z_2(t1) - z_2(t2) \\ \widetilde{y_2}(2) &= y_2(t_3) \\ &= \mathbf{h}_{2B}^{\mathsf{H}}\mathbf{x}^{[2]} + z_2(3) \end{aligned} \right\} \Rightarrow \hat{\mathbf{x}}^{[2]} = \begin{bmatrix} \mathbf{h}_{2A}^{\mathsf{H}} \\ \mathbf{h}_{2B}^{\mathsf{H}} \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{y_2}(1) \\ \widetilde{y_2}(2) \end{bmatrix}$$		

Table 2.1: Blind Interference Alignment for the 2×2 scenario

■ 2.2 Synchronization in distributed MIMO systems

In a distributed MU-MIMO setting, timing and carrier phase synchronization across the jointly precoded APs are needed in order for ZFBF and THP precoding to work. As discussed above, timing synchronization requires only that all nodes align their slots within the length of the OFDM CP. This is relatively easy to achieve, and it has already been implemented in software radio testbeds as in [RHK10, TFZ+10]. Carrier phase synchronization, however, is much more challenging. While a centralized MU-MIMO transmitter has a common clock source for all its RF chains [SYA+12], in a distributed setting each AP has an individual clock. The relative time-varying instantaneous phase

offset between the different transmitters may cause a phase rotation of the transmitter signals across a downlink slot such that, even though at the beginning of the slot we have ideal precoding (e.g., ZFBF or THP), the interference nulling effect is completely destroyed towards the end of the slot.

It is important to remark here that, while synchronizing a receiver with a transmitter for the purpose of coherent detection is a well-known problem for which robust and efficient solutions exist and are currently implemented in any coherent digital receiver [PS07], here we are faced with a different and significantly harder problem, which consists of synchronizing the instantaneous carrier phase of different *transmitters*. This requires that APs must track an RF carrier reference and compensate for the relative (time-varying) phase rotation while they are transmitting the downlink slot. Simultaneous transmission of the data signal and reception of the carrier reference signal cannot be implemented by standard off-the-shelf terminals. Instead, we have devised a system architecture to accomplish this goal.

Why is distributed MU-MIMO challenging? For simplicity of exposition, consider a distributed MU-MIMO system with two clients and two access points, each one with a single antenna and using ZFBF (the following considerations apply immediately to more general scenarios). For nomadic users, as in typical WLAN setting, the physical propagation channel changes quite slowly with time, so that we may assume that the channel impulse response is locally time-invariant. In order to use ZFBF, the channel matrix coefficients at each OFDM subcarrier must be estimated and known to the transmitter central server. Various methods for learning the downlink channel matrix at the transmitter

$$
\widetilde{\mathbf{H}}(n;t) = \underbrace{\begin{bmatrix} e^{j\left(\frac{2\pi}{NT_s}\tau_1 n + \phi_1(t)\right)} & 0 \\ 0 & e^{j\left(\frac{2\pi}{NT_s}\tau_2 n + \phi_2(t)]\right)} \end{bmatrix}}_{\boldsymbol{\Phi}(n;t)} \begin{bmatrix} H_{11}(n) & H_{12}(n) \\ H_{21}(n) & H_{22}(n) \end{bmatrix}
$$

$$
\underbrace{\begin{bmatrix} e^{-j\left(\frac{2\pi}{NT_s}\delta_1 n + \theta_1(t)\right)} & 0 \\ 0 & e^{-j\left(\frac{2\pi}{NT_s}\delta_2 n + \theta_2(t)\right)} \end{bmatrix}}_{\boldsymbol{\Theta}(n;t)}
$$

(2.13)

side have been proposed, including closed-loop feedback schemes (see [CJKR10] and the references therein) or open-loop schemes that exploit the uplink/downlink channel reciprocity of TDD systems [JAWV08]. For simplicity of exposition, we will assume here that the channel estimates correspond perfectly to the actual channel.

The central server computes the precoding matrix as seen in Section 2.1, for each subcarrier $n = 1, \ldots, N$. Let

$$\mathbf{H}(n) = \begin{bmatrix} H_{11}(n) & H_{12}(n) \\ H_{21}(n) & H_{22}(n) \end{bmatrix} \tag{2.13}$$

denote the 2×2 downlink channel matrix between the two clients and the two access point antennas on subcarrier n, and let $\mathbf{V}(n)$ denote the corresponding precoding matrix such that $\mathbf{H}^{\mathsf{H}}(n)\mathbf{V}(n) = \mathbf{\Lambda}^{1/2}(n)$ is diagonal. If the timing and carrier phase reference remain unchanged from the slot over which the channel is estimated and the slot over which the precoded signal is transmitted, we obtain perfect zero-forcing of the multiuser interference.

Suppose now that the timing reference and carrier phase reference between the estimation and transmission slots of the two APs is not ideal. With perfect timing, the downlink channel from AP i to client j would have impulse response $h_{ij}(\tau)$. Instead, due to lack of synchronization, the impulse response is $h_{ij}(\tau - (\tau_i - \delta_j))e^{j(\phi_i(t) - \theta_j(t))}$ where τ_i, δ_j denote the timing misalignment of AP i and client j, respectively, and $\phi_i(t), \theta_j(t)$ denote the instantaneous phase differences (with respect to the nominal RF carrier reference) of AP i and client j, respectively. For simplicity, we assume here that the sampling clock at all nodes is precise enough such that we may assume that the sampling frequency is the same and does not change significantly in time over the duration of a slot. Hence, τ_i and δ_j can be considered as unknown constants. Furthermore, we assume that they are multiples of the sampling interval T_s (i.e., the duration of the time-domain samples) otherwise the derivation is more complicated, involving the folded spectrum of the channel

frequency response, but the end result is equivalent to what derived here. Instead, we model the instantaneous phases of the RF carrier oscillators as

$$
\begin{aligned}
\phi_i(t) &= \phi_i(0) + 2\pi\Delta_i t + w_i(t) \\
\theta_j(t) &= \theta_j(0) + 2\pi\Delta_j t + w_j(t)
\end{aligned}
$$
(2.14)

where $\phi_i(0), \theta_j(0)$ are unknown constants, $\Delta_i = (f_{c,i} - f_c)(N + L)T_s$ is the normalized frequency offset of node i with respect to the nominal carrier frequency f_c, and $w_i(t)$ is a zero-mean stationary phase noise process, whose statistics depends on the hardware implementation. In the above expression, the time index t ticks at the OFDM symbol rate, i.e., at intervals of duration $(N + L)T_s$.

From the well-known rules of linearity and time-shift of the discrete Fourier transform, we arrive at the expression for the effective channel matrix in (2.13). The diagonal matrix of phasors $\mathbf{\Phi}(n; t)$ and $\mathbf{\Theta}(n; t)$ depend, in general, on both the subcarrier and OFDM symbol indices n and t. The multiplication of the nominal channel matrix $\mathbf{H}(n)$ from the right (receiver side, according to the channel model (2.2)) poses no problems, since these phase shifts can be recovered individually by each client as in standard coherent communication [PS07]. In contrast, the diagonal matrix $\mathbf{\Phi}(n)$ multiplying from the left (the transmitter side) poses a significant problem: since the server computes the MIMO precoding matrix $\mathbf{V}(n)$ based on its estimate $\mathbf{H}(n)$, it follows that when applied to the effective channel $\widetilde{\mathbf{H}}(n)$ in (2.13) the matrix multiplication $\widetilde{\mathbf{H}}^{\mathsf{H}}(n)\mathbf{V}(n)$ is in general far from diagonal. To stress the importance of this aspect, we would like to make clear that the resulting signal mixing takes place over the actual transmission channel, making it impossible for the receivers to eliminate it.

Why Synchronization Is Possible. 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 tens of MHz, 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 then multiplied by a sinusoidal carrier produced by the carrier clock. The result is a passband signal at a frequency of a few GHz 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 RF carrier clock and the receiver's RF carrier clock and sampling clock. All four clocks manifest phase "drift" (i.e., a linear time-varying term) and "jitter" (i.e., a random fluctuation term). We have assumed that the sampling clocks have no significant drift and jitter, and the only effect of timing misalignment (within the length of a CP) is captured by the constants τ_i and δ_j in (2.13). In contrast, the carrier clocks are affected both by drift and jitter (see (2.14)). [3] Furthermore, the phase noise term $w_i(t)$ may have some slow dynamics that can be linearized locally, over the duration of a slot, and add up to the linear phase term, such that the slope of the phase drift is constant over a single slot, but it is not constant over longer time intervals, in general.

We have verified experimentally the validity of our model, by letting a transmitter send several tone signals, i.e., simple unmodulated sine waves, corresponding to different subcarriers of the OFDM modulation, and using a receiver to sample, demodulate and extract the instantaneous phase trajectory of the received tones. In the absence

---

[3]In fact sampling clocks and carrier clocks have similar drift and jitter dynamics. However, since the carrier clock is multiplied in order to achieve a 2.4 GHz carrier frequency while the sampling clock frequency is in the tens of MHz, the phase effect of the carrier clock drift significantly outweighs the one of the sampling clock drift and is the only one relevant over the duration of a packet transmission.
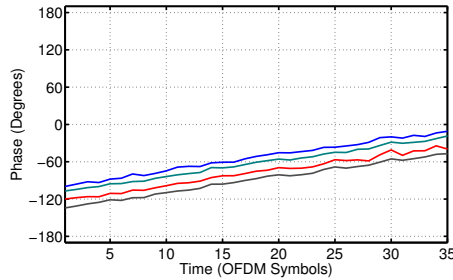
Figure 2.2: **Pilot phases.** The phases of different subcarriers drift at the same speed, suggesting that they are only subject to carrier frequency offset.

of phase offset these signals would exhibit a constant phase when measured over a sequence of several OFDM symbols. Instead, the measured instantaneous phase is time-varying and closely approximate parallel straight lines, as shown in Figure 2.2. The common slope of these straight lines is given by the carrier frequency offset Δ_i between transmitter and receiver. The spacing between the lines is given by constant phase terms $\frac{2\pi}{NT_s}\tau_i n$ for different subcarrier index n, and depends on the time misalignment τ_i between the AP and the nominal slot initial time. The small fluctuations around the linear behavior of the instantaneous phase is due to the phase noise, which is quite small for the WARP hardware used in our system, as it can be observed qualitatively from plots as in Figure 2.2.

It follows that by estimating the spacing between the phase trajectories (intercepts with the horizontal axis) and their common slope, we can track and predict across the slot the phase de-rotation coefficients to be applied at each AP in order to "undo" the effect of the matrix $\boldsymbol{\Phi}(n;t)$. Notice that the de-rotation factor must be predicted a few OFDM symbols ahead, in order to include the delay of the hardware implementation between when an OFDM symbol is produced by the baseband processor (FPGA) to when it is actually transmitted.

■   2.3 Related Work

Theoretical Foundations. The pioneering papers by Foschini [FG98] and Telatar [Tel99] have shown that adding multiple antennas both to the transmitter and to the receiver

increases the capacity of a point-to-point communication channel. At practical medium-to-high Signal to Noise Ratios (SNRs), this gain manifests as a multiplicative factor equal to the rank of the matrix representing the transfer function between the transmit and the receive antennas. For sufficiently rich propagation scattering, with probability 1 this factor is equal to $\min\{N_t, N_r\}$, where N_t and N_r denote the number of transmit and receive antennas, respectively. The MIMO capacity gain can be interpreted as the implicit ability to create $\min\{N_t, N_r\}$ "parallel" non-interfering channels corresponding to the channel matrix eigenmodes, and it is referred to in the literature as *multiplexing gain*, or as the *degrees of freedom* of the channel. Subsequently, Caire and Shamai [CS03] have shown that the MIMO broadcast channel, where the transmitter has N_t antennas and serves K clients with N_r antennas each, exhibits an analogous capacity factor increase of $\min\{N_t, KN_r\}$, suggesting that a transmitter with multiple antennas could transmit simultaneously on the same frequency to independent users. Such multiuser communication has two additional requirements. First, precoding of the transmitted data is needed to prevent the different spatial streams from mutually interfering. Second, the transmitter requires accurate knowledge of the channel matrix (channel state information) in order to realize this precoding.

The idea of precoding has spurred a large amount of research, well beyond the scope of this paper. Dirty Paper Coding (DPC) [Cos83] with a Gaussian coding ensemble achieves the capacity of the MIMO broadcast channel [WSS06], but is difficult to implement in practice. The well-known linear Zero-Forcing Beamforming (ZFBF) [CS03] achieves the same high-SNR capacity factor increase, with some fixed gap from optimal that can be reduced when the number of clients is large and the transmitter can dynamically select the clients to be served depending on their channel state information [YG06, KC06]. Tomlinson-Harashima Precoding is another well-studied, but infrequently implemented technique, which efficiently approximates DPC at high SNRs [WFVH04]. A number of other precoding strategies (e.g., lattice reduction, regularized vector perturbation) have been studied and the interested reader is referred to [SPSH04]

and references therein. For the purposes of this paper ZFBF and THP will be the primary methods of interest because of their conceptual simplicity and good complexity/performance tradeoff.

Practical Implementations. A number of recent system implementations have made forays into the topics of multiuser MIMO transmission and distributed, slot aligned OFDM transmission. MU-MIMO ZFBF as a precoding scheme in a centralized setting have been examined in [AASK10], for a system consisting of a single AP with multiple antennas hosted on the same radio board. A distributed system using a common clock source to drive a large number of radio boards is presented in [SYA+12]. This system uses conjugate beamforming [Mar10], a completely decentralized precoding scheme. The scheme requires a significantly larger number of antennas in order to provide rate gains comparable to the ones of centralized precoding schemes [SYA+12]. The use of interference alignment and cancellation as a precoding technique, which does not require slot synchronization or phase synchronization of the transmitters, has been illustrated in [GPK09]. While this solution achieves some spatial multiplexing, realizing the full spatial multiplexing gain using precoding schemes such as ZFBF requires tight phase synchronization between the jointly precoded transmitters [LSW05, VV09].

In order to be able to adopt the classical discrete-time symbol-synchronous complex baseband equivalent channel models used in communication and information theory, the fundamental underlying assumption is that transmissions from different nodes align within the cyclic prefix of OFDM (referred to as "slot alignment" in the following). If this is not verified, then inter-block interference arises and the channel does not decompose any longer into a set of discrete-time parallel channels. Slot alignment was used in SourceSync [RHK10] in conjunction with space-time block coding in order to provide a diversity gain in a distributed MIMO downlink system. In Fine-Grained Channel Access [TFZ+10], a similar technique allows for multiple independent clients to share the frequency band in fine increments, without a need for guard bands, resulting in a flexible OFDMA (OFDM with orthogonal multiple access) uplink implementation.

Distributed space-time coding and flexible orthogonal access do not increase the system degrees of freedom, since at most a single information symbol per time-frequency dimension can be transmitted. [4]

AirSync has been first published in [BRM+12] and then in [BRM+13]. A parallel work [RKK12] appearing at the same time used a similar set of ideas to develop a reference-based distributed MIMO system. To our knowledge, DistSync, first presented here, is the first system that extends distributed MIMO transmission to an arbitrary scale.

---

[4]A time-frequency dimension corresponds to one symbol in the frequency domain, spanning one OFDM subcarrier over one OFDM symbol duration, and spans (approximately) 1 s×Hz.

CHAPTER 3

Reference-based Synchronization

The current chapter introduces AirSync, a single reference-based synchronization method for wireless transmitters. This method achieves phase coherence among several transmitters located within the range of the reference access point. We start by presenting the construction of the synchronization scheme and the details of its implementation. We continue by evaluating its performance in an actual deployment. Finally, we describe a medium access layer compatible with its operation.

■ 3.1 AirSync

The fact that the common phase drift of all subcarriers can be predicted by observing only a few of them prompts the following approach to achieving phase synchronization between access points: a main access point (master) is chosen to transmit a reference signal consisting of several pilot tones [1] placed outside the data transmission band, in a reserved portion of the available bandwidth. An initial channel probing header, transmitted by the master access point, is used by the other transmitters in order to get an initial phase estimate for each carrier. After this initial estimate is obtained, the phase estimates will be updated using the phase drift measured by tracking the pilot signals.

---

[1]The use of multiple pilot tones ensures frequency diversity and spreads the pilot signal power over multiple frequency bins.

The estimate is used to calculate the difference between the carrier phase of each secondary transmitter and the phase of the master transmitter. This difference depends on the timing offset between the starting points of their frames and the frequency offset between the carrier frequency of the master AP (denoted by $f_{c,1}$) and the carrier frequency of each secondary AP (denoted by $f_{c,i}$, for $i > 1$). After obtaining the channel estimate, the secondary transmitters are able to undo effect of the instantaneous phase difference by derotating the transmitted frequency-domain symbols by the phase difference term along the whole transmission slot, thus eliminating the presence of the time-varying diagonal matrix $\Theta(n; t)$ in front of the estimated channel matrix and therefore achieving the desired MU-MIMO precoding along the whole transmission slot.

More specifically, at time $t = 0$, the n-th subcarrier signal generated by the master AP has the phase $\frac{2\pi}{NT_s}\tau_1 n + \phi_1(0)$, while the carrier generated by AP i has the phase $\frac{2\pi}{NT_s}\tau_i n + \phi_i(0)$. The phase of the instantaneous phase difference obtained from the master pilot tones is, ignoring the phase noise terms, $\frac{2\pi}{NT_s}(\tau_1 - \tau_i)n + \phi_1(0) - \phi_i(0) + \angle H_i(n)$, where $\angle H_i(n)$ is the phase of the channel coefficient between the master AP and AP i. If this phase estimate is added to the phase of the generated n-th subcarrier at AP i, the resulting phase becomes $\frac{2\pi}{NT_s}\tau_1 n + \phi_1(0) + \angle H_i(n)$, that is the phase of AP i is the phase of the master AP plus an offset $\angle H_i(n)$. To keep this offset constant over the duration of a transmission slot, the estimate must be adjusted by adding, for all t ranging over the transmission slot, the linear relative phase drift term $2\pi(\Delta_1 - \Delta_i)t$. In this way, after the phase compensation, all APs transmit at the *actual* frequency $f_{c,1}$ of the master AP.

The drift $2\pi(\Delta_1 - \Delta_i)t$ is estimated based on the out-of-band pilots using a sliding window smoothing filter over four samples to compute an updated value of the "slope" $\Delta_i - \Delta_1$. The secondary AP predicts, based on the current estimate, the instantaneous phase with a few OFDM symbols of look-ahead. The need for look-ahead prediction arises from the fact that the AP must align its phase to the phase of the reference at the moment of the actual transmission, not at the moment that the estimate has been recorded. Thus the look-ahead time of d OFDM symbols corresponds to the synchronization circuit delay. The prediction is obtained by simple linear extrapolation, by letting the correction term at time $t + d$ be given by $2\pi(\Delta_1 - \Delta_i)(t + d)$, where $\Delta_1 - \Delta_i$ is the

estimated slope at time t. The constant offset $\angle H_i(n)$ becomes a part of the downlink channel estimates and poses no further problems with regard to synchronization both when using downlink and uplink channel estimation schemes.

In our current implementation, for simplicity, we obtain an individual phase estimate of the form $\frac{2\pi}{NT_s}(\tau_1 - \tau_i)n + \phi_1(0) - \phi_i(0) + \angle H_i(n)$ for every subcarrier and use it independently of the estimates for other subcarriers in correcting the subcarrier phase. The form of the phase estimate suggests that it is possible to obtain a better estimate by breaking the estimation process into two distinct parts: obtaining an initial, high quality estimate of the constant $\angle H_i(n)$ during a system calibration step and then estimating just the two factors $\tau_1 - \tau_i$ and $\phi_1(0) - \phi_i(0)$ in subsequent packet transmissions. The constant estimate in this case is needed since undoing the angle $\angle H_i(n)$ amounts to equalizing the channel between the master AP and the i-th AP. After equalizing the channel, the resulting phases can be unwrapped along the carrier index n. It results that, after compensating for the angle $\angle H_i$, the phase of the estimate is $\frac{2\pi}{NT_s}(\tau_1 - \tau_i)n + \phi_1(0) - \phi_i(0)$, linear in the carrier index plus a constant term. A linear MMSE fitting can be applied in order to find the two factors mentioned, which are in fact the slope of the line (the carrier phase with regard to the subcarrier index) and its intercept.

Software Radio Implementation. We have implemented AirSync as a digital circuit in the FPGA of the WARP radio platform [Ric]. The WARP radio is a modular software radio platform composed of a central motherboard hosting an FPGA and several daughterboards containing radio frequency (RF) front-ends. The entire timing of each WARP is derived from two local reference oscillators, hosted on its clock board: a 20 MHz oscillator serving as a source for all sampling signals and a 40 MHz oscillator which feeds the carrier clock inputs of the transceivers present on the RF front-ends. The frequency accuracy of the carrier generating clock is on the order of 1.5 ppm, leading to an expected variance of the CFO of about 4 kHz, for a carrier frequency of 2.4 GHz. The sampling clock has a frequency accuracy on the order of 1ppm. These accuracy values are typical of oscillators used in 802.11 applications and do not preclude the need for phase drift compensation. Within each radio, sharing the local clocks among the RF front-ends assures that all signals sent and received using the different front-ends are phase

synchronous. Phase synchronicity for all sent signals or for all received signals is a common characteristic of MIMO systems. However, the fact that the design of the WARP ensures phase synchronicity among the sent and received signals, as opposed to using separate oscillators for modulation and demodulation, greatly simplifies the synchronization task. The system's data bandwidth is 5 MHz. We place the synchronization tones outside the data bandwidth, at about 7.5 MHz above and below the carrier frequency.

The slave APs have to track the out of band pilots (i.e., receive these signals) and transmit the data signal *at the same time*, in an FDD manner. We have dedicated one antenna of each secondary AP to receiving and tracking the reference signal, while the other antennas are used for transmitting phase-synchronous signals. The system design must mitigate self-interference between the transmit and receive paths.

In FDD transmission schemes in which the front-ends sample the entire system bandwidth, the dynamic ranges of the ADC and DAC circuitry plays an important limiting role. As opposed to a complete full-duplex system, in which self-interference cancellation is the main challenge to be solved, in bandwidth sharing systems the main challenge is accommodating both the incoming signal, i.e. the signal from the master AP, and the secondary AP's data signal within the limited dynamic range of the secondary's receiver front-end. A second challenge is shaping this data signal in order to prevent any significant power leakage outside the data band, mitigating the need for large guard bands between the data and the pilots.

The dynamic range needed can be computed as follows: assume that the secondary AP's signal and the master AP's pilots are broadcasted at the same power level. If the secondary's receiver antenna is α times closer to the secondary's transmitter antenna than to the one of the master's, assuming a free space propagation model in which the power decays as $\frac{1}{\alpha^2}$, it results that the data signal is received at $10*\log_{10}(\alpha^2)$ dB above the pilot signals. For α in the 32 to 128 range, this amounts to 30 dB to 42 dB. For comparison, the WARP's 14-bit ADCs offer a dynamic range of 84 dB. [2]

---

[2]This requirement could be further relaxed through the use of an analog rejection filter over the data band, before sampling, during the tracking period, thus decreasing the needed dynamic range through receive-side filtering.
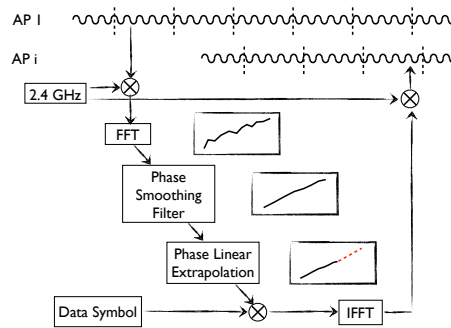
Figure 3.1: **AirSync operation.** A secondary AP (bottom) synchronizes its phase to the one of a reference signal (top) by adjusting the phase of its signal to match the phase of the reference.

For the second problem, the design of WiFi-NC [CRB+12], offers a clear indication of what can be achieved in a software radio using the same components as the WARP. To limit the size of required guard bands in a bandwidth-sharing system, in which different APs divide the data band into slices and can transmit in duplex over separate slices, the authors construct an OFDM transmitter with a sharp spectral footprint. By employing digital filters in the FPGA, they achieve a 60 dB power decay with guard bands that total 4% of the data bandwidth, as proved by spectrum analyzer plots. Their filter response time is well within the cyclic prefix. This approach allows for decreasing the over-the-channel power leakage into the pilot band through sender-side filtering. In our system, we achieve a similar effect by using the baseband sender filter present in the transmit signal path of the WARP's transceiver. In general, self-interference can be avoided using a number of other techniques such as antenna placement [CJS+10], digital compensation [DDS11], or simply relying on the OFDMA-like property of a symbol aligned system [TFZ+10] and preventing the secondary APs from using the pilot subcarriers.

We have implemented a complete system-on-chip design in the FPGA, taking advantage of the presence of hard-coded ASIC cores such as a PowerPC processor, a memory controller capable of supporting transfers through direct memory access over wide data buses and a gigabit Ethernet controller. Atop this system-on-chip architecture we have ported the NetBSD operating system and created drivers for all the hardware components hosted on the platform, capable of setting all system and radio board configuration parameters. The operating system runs locally but mounts a remote root filesystem

through NFS. In the same system-on-chip architecture we integrated a signal process-ing component created in Simulink which provides interfaces for fast direct memory access. This latter component is responsible for all the waveform processing and for the synthesis of a phase synchronous signal and interfaces directly with the digital ports of the radio front-ends. We interfaced the Ethernet controller and the signal processing component using an operating system kernel extension responsible for performing zero-copy, direct memory access data transfers between the two, with the purpose of passing back and forth waveform data at high rates between a host machine and the WARP plat-form. The large data rates needed (160 Mbps for a 5MHz wireless signal sampled at the 16 bit precision of the WARP DACs for both the real (I) and imaginary (Q) parts of the corresponding baseband signal) required optimizing the packet transfers into and out of the WARP. For example, consider the direct memory access ring associated with the receive end of the Ethernet controller on the board, which is shared between packets destined to the signal processing component and packets destined to the upper layers of the operating system stack. We do not release and reallocate the memory buffers occu-pied by packets destined to the signal processing component. Instead, we use a lazy garbage collection algorithm in order to reclaim these buffers when they are consumed in a timely manner or reallocate them at a later point if they are not consumed before the memory ring runs low on available memory buffers. The rationale for this particular optimization is that the overhead of managing the virtual-memory based reallocation of memory buffers of tens of thousands of packets every second would bring the processor of the software radio platform to a halt.

All transmitting WARP radios are connected to a central processing server through individual Ethernet connections operating at gigabit speeds. Most of the signal synthe-sis for the packet transmission is done offline, using Matlab code. We produce precoded packets in the form of frequency domain soft symbols. However, the synchronization step and the subsequent signal generation is left to the FPGA. The server, a fast machine with 32 processor cores and 64GB of RAM, encodes the transmitted packets and streams the resulting waveforms to the radios. Figure 3.1 illustrates the process of creating a phase synchronous signal at the secondary AP.

Centralized joint encoding. By transmitting phase synchronous signals from multiple APs, we have created a virtual single MU-MIMO transmitter, for which standard MU-MIMO precoding strategies can be used. However, the use of distributed APs complicates the design of the transmitter system. In order to eliminate multiuser interference, the data streams to different clients must be jointly precoded, as we have seen in Section 2.1. For systems with a very large number of jointly processed antennas and targeting mobile cellular communications (e.g., see [SYA$^+$12]), the centralized computation of the precoding matrix, of the precoded based band signals, and distribution of these signals to all the antennas would require a large delay, which is incompatible with the short channel coherence time due to user mobility. In contrast, in our enterprise network or residential network scenario, the channel coherence time is much longer (typical users are nomadic, and move at most at walking speed). Therefore, computing the precoding matrix does not represent a significant problem, and it is in fact better to perform centralized precoding and distribution of the baseband precoded signals. For example, using the conjugate beamforming scheme of [SYA$^+$12], it is possible to compute the precoded signals in a decentralized way, since each AP i needs just to combine the clients' data streams with the complex conjugates of its own estimated channel coefficients, i.e., with the elements of the i-th row of the channel matrix. In the notation of Section 2.1, this corresponds to letting $\mathbf{x} = c\mathbf{H}\mathbf{u}$, for some power normalizing constant c, such that the precoded channel becomes $\mathbf{y} = c\mathbf{H}^{\mathsf{H}}\mathbf{H}\mathbf{u} + \mathbf{z}$. Unless $M \gg K$, the resulting matrix $\mathbf{H}^{\mathsf{H}}\mathbf{H}$ is far from diagonal, and the system is interference limited, i.e., by increasing the transmit power, the system sum rate saturates to some constant value (the system multiplexing gain in this case is 1, corresponding to serving only one client on each time-frequency dimension, as in standard FDMA/TDMA). Hence, while conjugate beamforming is an attractive scheme for very large M, relatively high client mobility and limited power (as in a cellular system), it turns out that in the WLAN setting with not so large M, low client mobility and large operating SNR (due to communication range of at most a few tens of meters) this is not a competitive choice.

As a matter of fact, centralized ZFBF or THP precoding is much better in our setting. It should also be noticed that by centralized precoding we need only to send the I and Q

components of the frequency-domain OFDM baseband (precoded) symbols to the APs. This requires roughly $2b \times W$ bit/s, for signal bandwidth W Hz and b quantization bits per real sample. Instead, decentralized processing requires to send *all* client data streams to all APs. Assume for example that K clients are receiving at 4 bit/s/Hz (corresponding to 20 Mbps over a $W = 5$ MHz bandwidth). This requires $20 \times K$ Mbps to be sent to all APs, while in the case of centralized processing, with $b = 16$ bits of quantization, we need only 32×5 Mbps. Here, for $K > 5$, centralized processing is convenient also in terms of the backhaul data rate. For sufficiently large K, centralized processing is eventually less demanding than decentralized processing in terms of the backhaul data rates.

Our central server has an individual gigabit Ethernet connection to each of the WARP radios serving as APs. We divide the downlink time into slots and in each slot schedule for transmission a number of packets destined to various clients, according to an algorithm that will be presented in Section 3.3. For each AP, the server computes the I and Q components of the precoded baseband frequency domain waveform 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 APs and clients. The server transmits their corresponding waveforms to all secondary APs, and finishes by feeding the master AP, so that the master AP starts transmitting right away and the secondary AP can immediately synchronize and follow.

At the moment we obtain CSI using a downlink estimation procedure, similar to the one presented in 802.11ac. In a future refinement of our system, we would like to reduce the overhead of obtaining CSI by using an uplink estimation scheme that takes advantage of channel reciprocity, thereby reducing considerably the length of the channel estimation procedure.

The current implementation relies on a single master broadcasting a reference pilot for all secondary transmitters. A straightforward extension of our system would have
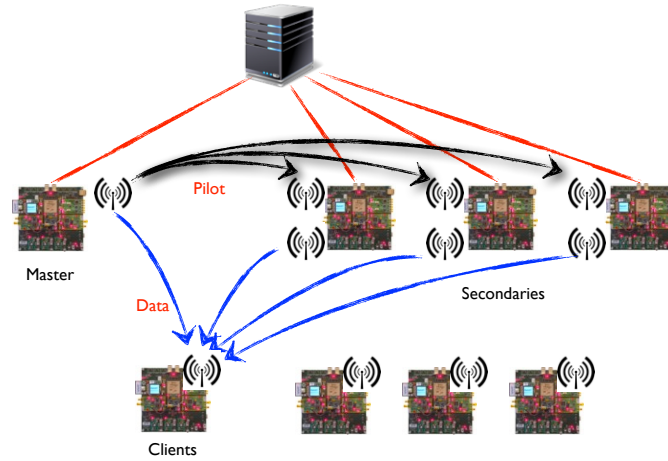
Figure 3.2: **Testbed diagram.** The central server is connected to four transmitters, the main transmitter on the left and the three secondary transmitters on the right. Four receivers act as clients.

some of the secondary access points relaying the pilot signals on different carrier frequencies. By using a set of alternating frequencies, in a fashion similar to cellular networks, the synchronization scheme can cover larger topologies.

◼ 3.2 Performance Evaluation

Our system setup is presented in Figure 3.2. It consists of a primary transmitter, three secondary transmitters and four receivers. The main sender uses a single RF front-end configured in transmit mode, placing an 18 MHz shaping filter around the transmitted signal. The secondary senders use an RF front-end in receive mode and a second RF front-end in transmit mode, with a 12 MHz shaping filter. As mentioned previously, the pilots used in phase tracking are outside the secondary's transmission band, therefore the secondary transmitter will not interfere with the pilot signals from the main transmitter. The series of experiments is intended to test the accuracy of the synchronization, the efficiency of channel separation and the extent to which we achieve the theoretical gains that multiuser MIMO promises in our setup.

The SNR values were measured using the received waveforms, not the RSSI indicator of the WARP transceiver. We initially measured the noise figure of the receiver and then,
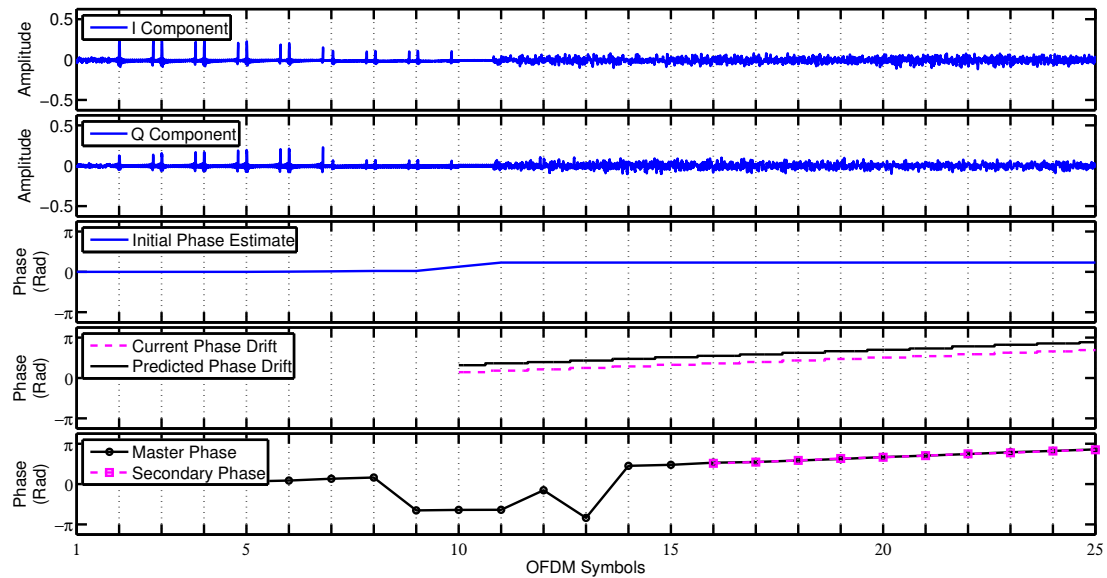
Figure 3.3: **Phase Synchronization Acquisition.** The secondary transmitter receives in-phase and quadrature components (real and imaginary components) of the master signal (top two figures). It then obtains an initial phase estimate (middle figure) from these samples. The secondary tracks the phase drift of the subcarriers beginning at the 10th symbol (second from bottom figure) and uses a filter to predict its value a few symbols later (bottom figure).

in subsequent measurements, integrated the received power to obtain the signal plus noise figure.

Synchronization Accuracy

In this particular experiment we have placed two transmitters and two receivers at random locations. We placed a third RF front-end on the secondary sender and configured it in receive mode. 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 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. The result was a synthesized signal that closely follows the phase of the signal broadcast by the master transmitter, as illustrated in Figure 3.3. The figure illustrates the initial phase acquisition process, the initial phase estimation, the tracking and estimation of the phase drift, as well as the synthesis of the new signal. The phase discontinuities appearing in the main transmitter's signal are due to the presence of the PN sequence along with a temporary disturbance needed in order to tune the feedback circuit.
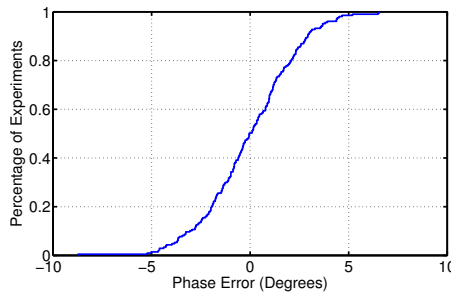
Figure 3.4: **The Precision of the Phase Synchronization.** AirSync achieves phase synchronization within a few degrees of the source signal.

Figure 3.4 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.

The radios were placed in a typical office environment. We have measured the SNR value of the synchronization pilots in the signal received by the secondary transmitter to be around 28.5 dB above the noise floor. This is easily achievable between typically placed access points.

Beamforming gain

Our second experiment was done using two transmitters and a receiver. Using Airsync synchronization, the transmitters broadcast their signals over the air in a phase coherent fashion. We measured the channel coefficients between the two transmitters and the receiver using standard downlink channel estimation techniques. Based on these measurements, we arranged the amplitudes and the phases of the transmitted signals such that at the receiver the amplitudes of the two signals would be equal, while their phases would align in a beamforming fashion. The maximal theoretic power gain over transmitting the two signals independently is 3.01dB. We compared the average power of the individual transmissions from the two senders to the average power of the beamformed joint transmission. Our measurements show an average gain of 2.98 dB, which is consistent with the precision of the synchronization determined in the previous experiment. This result shows that for all practical purposes we are able to achieve the full beamforming gain in our testbed.

Zero-Forcing Accuracy
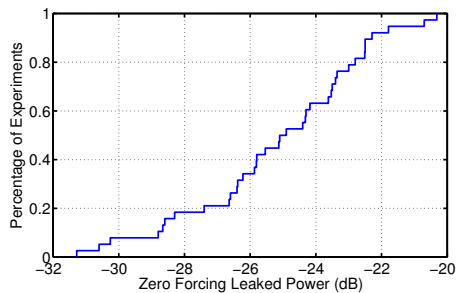
Figure 3.5: **The Power Leakage of Zero-Forcing.** The leaked power is significantly smaller than the total transmitted power, transforming each receiver's channel into a high SINR channel.

The following experiment measures the amount of power which is inadvertently leaked when using Zero-Forcing to non-targeted receivers due to synchronization errors. Again we have placed two transmitters and a receiver at random locations in our testbed. We have estimated the channel coefficients and arranged for two equal amplitude tones from the two transmitters to sum as closely as possible to zero. The

residual power is the leaked power due to angle mismatching. Figure 3.5 illustrates the CDF of this residual power for different measurements. The average power leaked is -24.46 dB of the total transmitted power. This establishes that Zero-Forcing is capable of almost completely eliminating interference at non-targeted receiver locations.

Zero-Forcing Beamforming Data Transmission
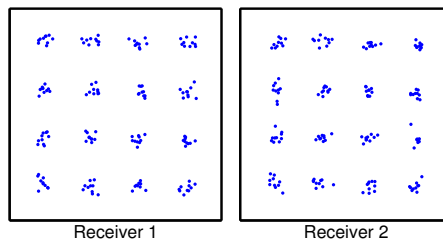
Figure 3.6: **Zero-Forcing Scattering Diagram.** The scattering diagram for two independent data streams transmitted concurrently using ZFBF demonstrates that AirSync achieves complete separation of the user channels.

This experiment transmits data from two transmitters to two receivers using ZFBF. We have used symbols chosen independently from a QAM-16 constellation at similar power levels. The scattering plots in Figure 3.6 illustrate the received signals at the two receivers. From the figure it is clear that we have created two separate channels. The actual rates achieved will depend on the quality of the two channels.

We would like to compare the performance of the multiuser MIMO system to a current standard. In current enterprise WiFi networks transmissions within a small area occur from single access points to single clients and are separated in time using TDMA. We use the best achievable point-to-point rate as an upper limit for the rates that the TDMA approach can achieve and compare the rates achieved by our system.

The SINR values at the two receivers are 29 dB and 26 dB respectively. In the same experiment, we measured the best point-to-point link to have a 32 dB SNR value. Using Shannon's formula, these values translate to maximally achievable rates of 9.96 bits/second/Hz (bps/Hz)for the point-to-point channel and 18.27 bps/Hz for the compound MIMO channel. Thus, when using ideal codes, we achieve a multiplexing rate gain of 1.83, which is close to the theoretical value of 2.
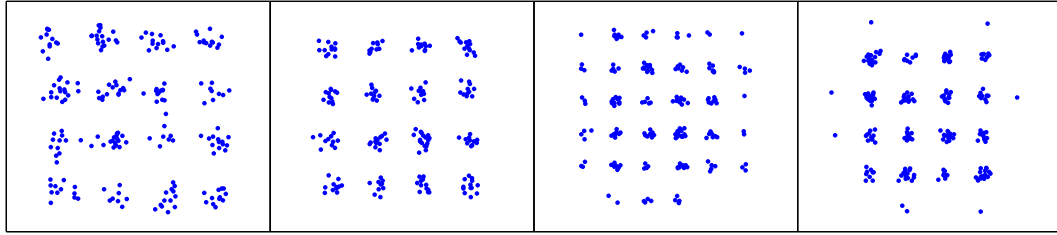
Figure 3.7: **Tomlinson-Harashima precoding.** Tomlinson-Harashima precoding based on QAM-16 constellations. The achieved spectral efficiency is 16 bits/second/Hz

At all the mentioned SNR levels 802.11g (a point-to-point standard) uses the same 64-QAM modulation, resulting in a rate of 6 bps/Hz (ignoring the error correcting code overhead, which is identical for all three SNR levels). Thus, we can say that both of the channels obtained through zero-forcing support WiFi operation at the highest commonly used rates and therefore equivalate to independent WiFi channels. We conclude that, using practical modulations, the experimental multiplexing gain equals the theoretical value of 2.

Tomlinson-Harashima precoding

The final experiment uses four transmitters and four receivers. We employ Tomlinson-Harashima precoding. The results are illustrated in Figure 3.7, which presents the four distinct wireless channels created for the four users. Thus, we have achieved a multiplexing factor of 4. As before, the actual rate gains will depend on the quality of the channels.

We measured the SINR values of the four channels to be 16.8 dB, 19.2 dB, 21.4 dB and 20.8 dB. The lower SINR values are caused by increased levels of power leakage due to the presence of more transmissions to other receivers (see Figure 3.5 for the distribution of leaked power from a single interfering transmission). Again, the Shannon rate formula predicts achievable channel rates of 5.6 bps/Hz, 6.4 bps/Hz, 7.11 bps/Hz and 6.91 bps/Hz. The sum rate is 26 bps/Hz. As mentioned before, the best point-to-point channel in our setup has a quality level of 32 dB, allowing for 9.96 bps/Hz. Therefore the rate gain is about 2.6 when using four degrees of freedom and ideal codes.
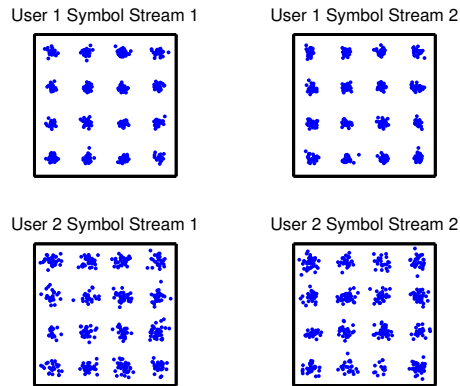
Figure 3.8: **BIA Scattering Diagram.** Blind Interference Alignment enables the multiplexing of four user streams over three timeslots.

More practically, we can compare the performance of our system when employing an extended 16-QAM constellation on every channel with the performance of 802.11g using a typical modulation. At 32 dB SNR, 802.11g would use a 64-QAM constellation and achieve (ignoring the error correcting code overhead) a spectral efficiency of 6 bps/Hz. In the MIMO case, we can achieve a sum rate of 16 bps/Hz using four 16-QAM constellations, leading to a multiplexing gain of 2.66 under practical modulations, while the theoretical value is 4. In a commercial implementation, we expect the leakage to be further reduced and we expect to be able to come closer of a rate gain of 4. In general, nearing the theoretical rate gains through spatial multiplexing requires precise channel state information and tight synchronization, as evidenced by our experiments.

Blind Interference Alignment

Achievable rates. We have used the testbed topology illustrated in Figure 2.1 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 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

(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)
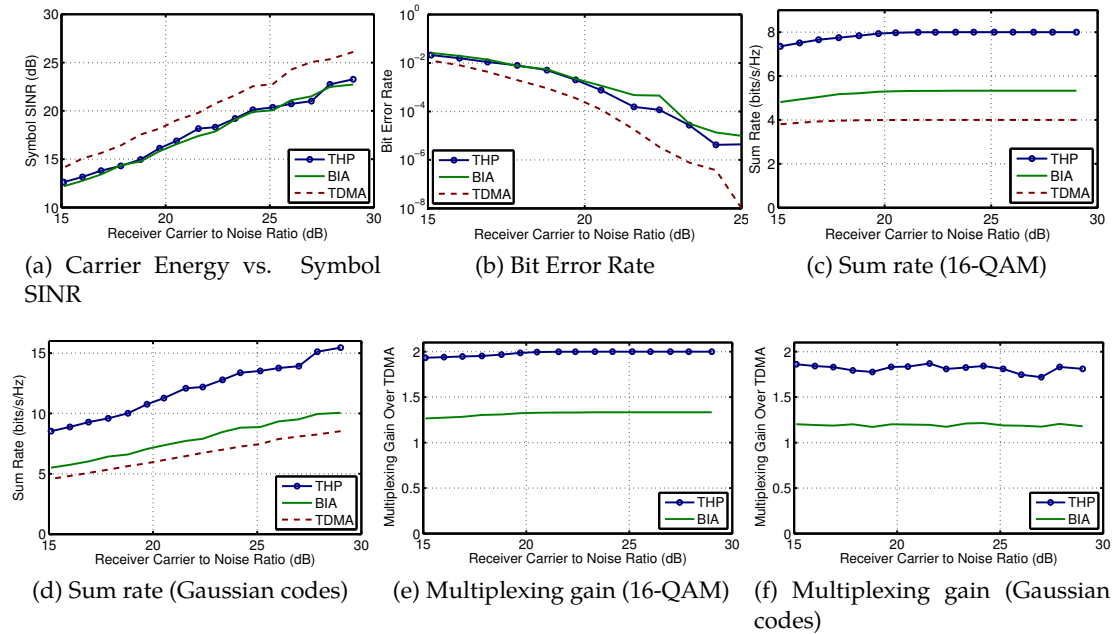
Figure 3.9: **Experimental Results.** The absolute and relative rates of BIA and THP at different SNR values, under different modulations.

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 continuously 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 3.9a presents the SINR values for symbols received when using each of the three precoding schemes. Figure 3.9b 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 3.9c 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 3.9e 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 3.9d presents the resulting sum rates and Figure 3.9f 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 trans-mitter. 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 2.10.

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

In the case of a distributed MIMO system, we would expect that phase synchro-nization error could lead to random rotations of the received soft symbols. We investi-gated this effect by comparing the variance of soft symbols corresponding to constella-tion 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 CDFs of the resulting SINR distributions are shown in Figure 3.10. The high variance of the distribution has profound implications on the design of a coding and medium access scheme for BIA. 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.
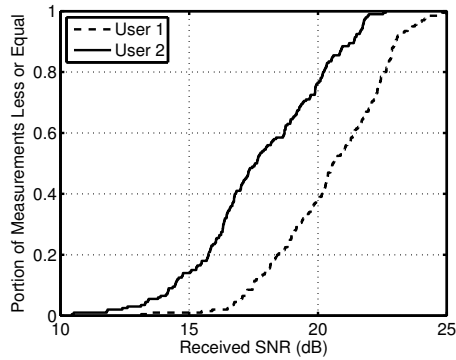
Figure 3.10: **BIA Channel Quality.** The cumulative distribution function of received SNRs under the Blind Interference Alignment Scheme.

∎   3.3   Medium Access Control

Given that we have achieved the necessary synchronization accuracy between access points and realized the full multiplexing gain, we turn to the large body of work on optimal scheduling for centralized multiuser MIMO systems (see for example [KC06,DS05]). Inspired by this work, we propose a MAC layer that significantly departs from the classic networking layered architectural model and adopts a cross-layer "PHY/MAC" design strategy.

High level description

Time Division Duplexing. First, we consider the issue of allocating air time and frequency spectrum between the uplink and the downlink. We can choose between two natural strategies for separating the uplink from the downlink: time division duplex (TDD) and frequency division duplex (FDD). TDD has the following two advantages. First, with TDD one can exploit channel reciprocity and measure the uplink channel, using pilots from the users to infer the downlink channel. In the case of FDD, an explicit closed-loop channel estimation (from the downlink pilots sent by the access points) and feedback (from the clients to the server) needs to be implemented, with a protocol overhead that increases linearly with the number of jointly precoded access point antennas [JAWV11]. Second, TDD is ideally suited for the transport of asymmetric traffic, as
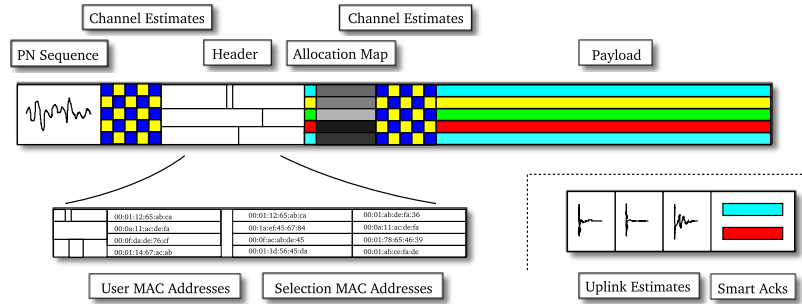
Figure 3.11: **Packet Design.** Downlink data packet (left) and uplink acknowledgment (bottom right).

is typical in an enterprise WiFi scenario, whereas an FDD system provides less flexibility for managing different traffic patterns. Specifically, with TDD, the downlink channel estimation procedure and the downlink time reservation proposed in the 802.11ac standard [80212] can be applied to our distributed MIMO system as well. We shall consider the scheduling of users in the uplink and downlink periods separately. In the uplink, clients compete for bandwidth using regular CSMA/CA. Thus, in the rest of this section we focus on the downlink. We note here that in order for our system to be backward compatible with legacy 802.11 clients and access points, protection mechanisms and modes of operation have to be implemented. Such mechanisms are described in the 802.11n/ac standards [80212] where, using RTS/CTS, CTS-to-self frames and legacy format preambles, nearby devices can sense that the channel is in use and avoid collisions.

Protocol Design

Our protocol design focuses on the downlink channel. Figure 3.11 presents a simplified schematic of the downlink data packets and corresponding uplink acknowledgments. The MAC layer protocol is tuned for enabling multiuser MIMO broadcasts. The crucial design constraint is to provide the central server with timely estimates of the channel state information for all clients to which it is about to transmit or which are considered for the next round of transmissions. For this purpose, we can collect channel estimates either at the access points through uplink pilots (based on TDD reciprocity) or at the receivers using a standard downlink estimation procedure as described in 802.11ac. The

central server uses the estimates to select a set of clients for the following transmission slots, according to the scheduling algorithms introduced earlier.

The choice between uplink and downlink estimation has an important impact on the design of the synchronization system: Uplink channel estimation is a "closed loop" design introducing a very small delay between the time one receives the channel estimation, and the time one transmits. In such a scenario, it is reasonable to expect that once the CFO is compensated for, the residual CFO will not result in significant phase drift among the carriers of the different transmitters during a channel measurement/packet transmission cycle. This simplifies synchronization as only the CFO (linear time-varying phase rotation) needs to be taken care of. [3]

Our protocol design follows the lines of 802.11ac [80212]: before a downlink transmission period the access points broadcast a request for a number of clients to estimate their channels based on a channel probing message broadcasted shortly after. The access points then transmit requests for feedback in succession to each targeted client and wait for the corresponding feedback. Once all the information has been collected, the downlink period can begin. We note that the use of a STBC for control frames can improve their robustness, given that from a client perspective the phases of the access points are essentially random during this phase.

The downlink data packet starts with a transmission from the main sender containing a pseudo-noise sequence used to achieve frame alignment by the transmitters and for block boundary detection by the receivers. The master access point then transmits the first set of channel estimation pilots which are used by the other access points to determine the initial phases of the subcarrier tones, as described in Section 3.1. After this point, all access points take part in the downlink transmission.

---

[3]To be precise, looking at Equation (2.13) let us distinguish between $\theta(t)$, which is a time-varying frequency-independent phase rotation due to CFO, and $2\pi\delta_i n/(NT_s)$, which is a time-invariant, frequency-dependent phase rotation due to timing misalignment. If uplink channel estimation is used, and the channel estimation subslot (uplink) and the data transmission subslot (downlink) are closely spaced in time such that the time axis reference remains the same in both subslots, then the effect of the delay δ_i on the channel frequency response is automatically included in the uplink channel estimation, and therefore these phase terms are included in the frequency domain channel coefficients for which the precoder is calculated. In this case, there is no need for explicit phase de-rotation of these terms. In contrast, if the channel estimation subslot and the data transmission subslot are separated by too long of a time interval or by a random (possibly fractional) number of samples, then we have to explicitly compensate for these phase rotation terms. This is usually the case in downlink estimation like the one we have implemented.

The packet header that follows is broadcast to all clients, including the non-targeted ones, using the Alamouti encoding [Ala98]. , as in SourceSync. Due to phase alignment between transmitters, the clients do not need to track the secondary senders in order to decode this header. The MAC addresses of the hosts targeted in the current transmission and the MAC addresses of the clients that are required to provide the server with channel estimates during the next acknowledgment are the most important pieces of information contained in the header fields. The positions of the addresses in the header fields create an implicit ordering of the clients, which will be used in the uplink period. The following part of the header is an allocation map, similar to the one found in the LTE standard, which assigns carriers to small groups of different clients and specifies the constellations used in broadcasting to them. The header is followed by a second set of channel estimation pilots, transmitted this time around by all access points using ZFBF, which are used by all clients in order to obtain the channel estimates for their individual downlink channels. The clients use the downlink estimates together with the synchronization pilot tones in order to gain a lock on the subcarriers. The downlink transmission continues with payload transmission.

In current 802.11 MIMO implementations, the channel estimates are obtained using downlink pilots which are in turn quantized by the receivers and communicated back in numerical form to the transmitter. The quantization and communication steps incur a large overhead. Using the reciprocity property of wireless channels, we can reduce the complexity of the channel estimation process significantly. First, we prefer to perform uplink channel estimation since uplink estimates can be received simultaneously by all access points, reducing the number of pilot transmissions needed by a factor equal to the total number of access point antennas. Second, uplink estimates are sent using analog pilot signals in an unquantized form, leaving the quantization step to the access points. This reduces the overhead of the transmission significantly. Third, while the usual estimation pilots are full OFDM symbols, we choose to send pulse-like signals, measure the channel response, and fill the non-significant taps with zeros before taking a Fourier transform in order to determine the frequency domain response. This ensures

that our pilots need to be spaced only by an interval that can accommodate a long channel response, i.e. the length of a cyclic prefix.

After the downlink transmission has finished, the clients who have been requested to send their channel estimates start sending these short estimation pilots in quick succession. We note that there is a large degree of similarity between the functioning of the downlink channel estimation for receive decode purposes and the uplink channel estimation step. The timing of the system remains unchanged during the uplink slot and the roles of the transmitters and the receivers are switched. The uplink pilots are followed by smart acknowledgments for the data packets sent using the technique detailed in [DSGS09a].

We tested each component of the downlink and uplink protocol slots. However, since our radios do not switch from receive to transmit in a timely manner, we could not perform complete real-time MAC experiments.

Overhead. A note on the overhead of the above MAC is in order. the overhead of our MAC is not more than that of 802.11n. The additional signaling overhead comes from requiring a few frames to predict the initial phase, and a few frames to dictate the MAC addresses of the nodes from which we wish to request channel state information for the next time slot. Even with very conservative estimates this will be less than a 20% increase in header time duration over that of a traditional 802.11 system. Note, however, that we get a bandwidth increase that grows almost linearly in the number of clients. This means that our overhead, normalized such that we consider the total control bits over the total data bits transmitted during a fixed airtime slot, is much less than in a traditional 802.11 system.

Achieving Scalability and Efficiency

The current chapter introduces DistSync, a scalable synchronization scheme aimed at large-scale deployments exceeding the effective range of a single access point. Dist-Sync uses a distributed synchronization algorithm which attempts to phase-lock the network's APs for the duration of a downlink transmission slot. As opposed to AirSync, DistSync does not attempt to suppress the random phases that different transmitters introduce in the channel model, but tries to estimate and use the effective channel before its coherence time ends. The downlink transmission slot makes use of the short-term constancy of the channel and uses an efficient estimation algorithm followed by fast precoding over a short timeslot duration. The presentation in the current chapter complements the theoretical description given in [RBP$^+$] and focuses on system design problems. The characteristics of oscillator-produced signals that influence the design of the algorithm are presented in detail, along with experimental data that describes the drift of oscillator frequency in time. We describe a synchronization scheme capable of compensating frequency offset with a high degree of precision while imposing a low wireless-transmission overhead on the network. We introduce optimal timing and phase estimators that increase estimation precision and decrease the amount of data sent over the wireless backhaul. DistSync has been implemented as a hardware proto-type. The performance of the synchronization scheme and the achieved spatial reuse are quantified through an experimental evaluation.

■ 4.1   Distributed Synchronization

Consider again the Equation 2.13:

$$\widetilde{\mathbf{H}}(n;t) = \mathbf{\Phi}(n;t)\mathbf{H}(n;t)\mathbf{\Theta}(n;t)$$

where $\mathbf{\Phi}(n;t) = \text{diag}(e^{j\left(\frac{2\pi}{NT_s}\tau_1 n + \phi_1(t)\right)}, e^{j\left(\frac{2\pi}{NT_s}\tau_2 n + \phi_2(t)\right)}, \ldots, e^{j\left(\frac{2\pi}{NT_s}\tau_M n + \phi_M(t)\right)})$ and $\mathbf{\Theta}(n;t) = \text{diag}(e^{-j\left(\frac{2\pi}{NT_s}\delta_1 n + \theta_1(t)\right)}, e^{-j\left(\frac{2\pi}{NT_s}\delta_2 n + \theta_2(t)\right)}, \ldots, e^{-j\left(\frac{2\pi}{NT_s}\delta_K n + \theta_K(t)\right)})$.

The equation describes the effective channel matrix of a system with K clients and M access points. Keeping with the notation of the previous chapter, the impulse response of the downlink channel from AP i to client j is $h_{ij}(\tau - (\tau_i - \delta_j))e^{j(\phi_i(t) - \theta_j(t))}$ where τ_i, δ_j denote the timing misalignment of AP i and client j, respectively, and $\phi_i(t), \theta_j(t)$ denote their respective instantaneous phase differences (when compared to the nominal RF carrier reference).

The previous chapter described a system that diagonalizes the transmitter side matrix $\mathbf{\Phi}(n;t)$ by using a single access point as a phase and timing reference for the other access points. The secondary access points mimic the phase of the reference and the transmitter-side matrix becomes constant in time. While this system allows several access points to transmit coherently, the need for a reference impacts its scalability. The effective deployment range is the beacon transmission range of a single access point, which can be increased by adding relays and their corresponding signaling overhead, but remains nevertheless limited. This approach does not manage to solve the essential problem of single-reference distributed MIMO synchronization, namely that due to the limited range of wireless beacons, synchronizing from a single point over the air is not the most effective way to attain phase coherence and is subject to error accumulation as the range over which the reference is repeated is increased.

A simple observation simplifies significantly the synchronization problem: for any one client, the signals that it receives, i.e. the beamformed packet transmitted to this particular user, is broadcasted from a limited number of access point within its own wireless range. Similarly, the packets that might affect its reception, for example the

packets concurrently transmitted to *other* users simultaneously, which are broadcasted such that their signals null at this client, are sent from its wireless proximity. It results that the SINR figure at the user and the characteristics of the signal propagated to it are affected only be the precision of the synchronization, in frequency, time and phase, between the access points within its range. The global synchronization requirement is thus replaced by a local synchronization requirement, and the network has to fulfill a more relaxed precision constraint.

In the current chapter we replace the centralized architecture presented in the previous chapter with a decentralized synchronization method which measures the different frequency and time offsets between pairs of neighboring access points and then gathers, over the wired network connecting them, this information to a central server and computes for each of them the necessary adjustments. Due to the locality of the wireless transmissions, frequency and timing offsets between APs need only be measured when the APs are within the range of the same client. That is, the measurement problem can be decomposed into subproblems over multiple, possibly overlapping, neighborhoods. Ideally, we would like to decentralize the final computational step as well. However, as long as the wireless neighborhoods of the simultaneously active users create a connected cover of the access points, the synchronization problem cannot be partitioned easily and optimally into independent subproblems, due to the interdependency between the adjustments computed for APs belonging to multiple neighborhoods. Sometimes such a connected cover does not exist and each group of independent users can be approached as a separate problem and the access points can be clustered into independent components. In the general case, the system will have to deal with a synchronization problem involving all access points in the network. In case the size of the problem makes the computation intractable, a quick solution would be to assign clients to a number of sets of concurrently served clients where each set can be separated in small enough components.

In the case of synchronization, the total delay is given the overhead of measurements, the time necessary for collecting data throughout the wired network, the computational hardness of the adjustment computation and returning the results to the APs. In the

case of precoding, the delay consists of collecting channel state information, precoding and returning the precoder to the APs. The complexity of the problem must be limited, such that the afore mentioned delays are under the maximum tolerable time for solving the problem, much less than the channel coherence time for precoding and less than the time it takes for oscillators to significantly change their frequency offsets due to thermal effects in the case of time and frequency synchronization. The size of the instances to be solved influences all the above delay factors.

The central element of our new approach is a different way of dealing with the effective channel matrix. Instead of modifying the transmitter side matrix $\Phi(n;t)$, we seek, through the use of precise timing and frequency offset estimation and compensation, to maintain it constant throughout a slot comprised of a channel estimation time and several downlink packet transmissions. We precode using the effective channel matrix as an input to the precoding algorithm and accept the complication that the precoding algorithm must be run within a slot duration, whenever the effective channel matrix has significantly changed [1]. While this approach introduces a hard real-time constraint on our precoding calculation, it allows the system to continue functioning based on over-the-air synchronization while removing any barriers caused by the wireless coverage area of the reference access point. The possibility of using the effective channel matrix in precoding was discussed in [BRM+13], while the theoretical considerations behind this approach have been discussed in [RBP+].

■   4.2 System Description

The novel synchronization techniques presented here are affected mainly by the behavior of the frequency drift of AP oscillators. We expand here the discussion of the hardware characteristics of our system and insist on the details that impact frequency estimation. Following that, for completeness purposes, we describe the system architecture introduced in [RBP+]. While this presentation will focus on giving a high-level protocol

---

[1]The quantitative meaning of this statement will be discussed later in this chapter.

overview, the following sections will present in more detail the pilots and the estimation procedure used at different steps in the protocols' operation.

Hardware

Our system is based on the WARP hardware platform [Ric]. WARP radios provide an FPGA capable of hosting digital logic for real-time waveform processing, a number of RF front-ends capable of transmitting and receiving in the 2.4 GHz frequency band and a clock board hosting oscillators used for creating the sampling clock of the digital baseband circuitry as well as the carrier clock used in passing signals from passband to baseband and in reverse. The oscillators have frequency stability on the order of 2.5 ppm (parts-per-million), which create carrier frequency offsets of up to 4 kHz between the different transmitters. Moreover, the oscillator frequencies are not fixed but are influenced by thermal variations and drift in a random fashion. Figure 4.1 illustrates the evolution of the carrier frequency offset present between two transmitters over a one second interval. It can be readily observed from this figure that the thermal effects continuously affect the two oscillators and that the drift they induce can change their relative frequency offset by a few tens of Hertz every second. The stability of these oscillators is by no means uncommon in low-cost WiFi equipment. Most routers, access points and client devices are constructed using TCXOs (thermal-controlled crystal oscillators) whose frequency precision is typically in the ppm range. A different class of oscillators, OCXOs (oven-controlled crystal oscillators) offers frequency stability in the parts-per-billion range, leading to CFOs of only a few Hertz. Such accurate oscillators are heavy and expensive, with prices currently ranging in the hundreds of dollars. While such high-end components are an appropriate choice in the design of cellular base stations, their use in WiFi routers, access points or femtocells is quite prohibitive. However, allowing carrier frequency offsets on the order of KiloHertz between transmitters would completely cancel their ability to beamform coherently over a time duration comparable to a packet transmission time, i.e. their relative phases would drift significantly

over such a short period. We choose therefore to eliminate the carrier frequency off-sets present between transmitters by carefully estimating them and compensating them during transmission.
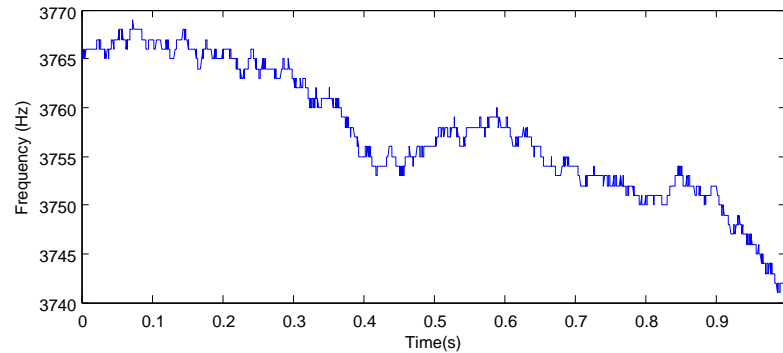
Figure 4.1: **CFO Drift.** The evolution of the carrier frequency offset between two WARP boards as measured over a one second interval.

In designing a wireless system, it is possible to use a single oscillator for creating both sampling and carrier clocks, which has the advantage of linking in an unambiguous way the carrier frequency offset (CFO) and the sampling frequency offset (SFO) of its radio transmitter. In fact, with this optimization, measurements of the former can be used to exactly compute the later. In general, CFO estimation is a simpler procedure than estimating the SFO, due to the fact that the carrier frequency is much higher than the sampling frequency. Phase changes induced by the CFO are much more rapid and can be estimated over shorter time intervals and with greater accuracy than the corresponding phase changes induced by SFO. It results that, under this setup, the presence of a separate sampling frequency offset estimator is not needed and the system's complexity can be greatly reduced.

The different radios fulfilling the role of access points are connected through a switch followed by a single gigabit Ethernet connection to a central server. The server has the role of collecting channel data sent by the access points, computing precoding coefficients and transmitting them back to the access points in a very short time interval, typically well under a millisecond. The software in charge of collecting the channel data and computing the precoding coefficients was implemented as a simple userspace UDP server which makes use of ATLAS [WPD01], LaPACK [ABD+90] and BLAS [WP05]

libraries for performing its linear algebra operations. The only optimization done on the server-side for reducing the response time was setting the network card's interrupt coalescing parameters to values that assured that a network interrupt was generated as soon as the expected number of coefficient packets arrived at the server and that the network card would transmit any outbound packets without any delay.

Architecture

The architecture of our system is hierarchical but decentralized. There are three main tasks that we aim to achieve through over-the-air synchronization: establishing a rough symbol alignment between transmitters, capable of supporting OFDMA (OFDM with Multiple Access) operation; eliminating, to the extent of achievable accuracy, carrier frequency errors between neighboring access points; lastly, using the resulting near-constant downlink channels for downlink data transmission. We aim to achieve all these goals in a completely distributed manner, without the need for a central wireless reference point as present in MegaMIMO [RKK12] or AirSync [BRM+13].
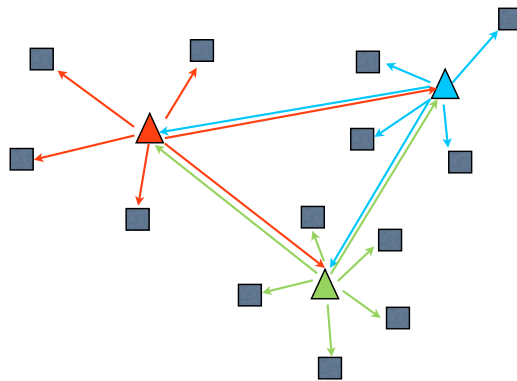
Figure 4.2: **Hierarchical structure.** The nodes are organized in clusters centered on a set of anchor nodes whose can communicate wirelessly with their anchor neighbors such that together they constitute a connected set of nodes. [RBP+]

Hosts are organized in a two-level hierarchy comprising a set of anchor nodes and a set of regular node. Anchor nodes are placed further apart but are capable of wirelessly communicating with each other (see Figure 4.2). Time alignment and frequency offset are measured and compensated first between the anchor nodes and then between all

nodes present in the system. The rate at which the synchronization algorithm must be run is about ten times per second. This number was obtained by quantifying the drift rate of the frequency offsets induced by the oscillators and limiting the maximum drift allowed before computing a fresh estimate to a value of at most 10 Hz. When the access points are symbol aligned and their relative frequency offsets are within a few Hertz of each other, their downlink channels can be assumed to be constant for millisecond-scale durations, since any residual phase drifts are unlikely to cause phase changes larger than a few decimal degrees and cause significant rate impairments.

As previously mentioned, we take advantage of the static nature of the channel at this short time scale and try to estimate it and use it for precoding and transmitting data before any significant change occurs. This approach is different from the one followed by MegaMIMO [RKK12] and AirSync [BRM+13], which were distributing a reference wireless signal in order to eliminate any transmitter-induced wireless channel changes and keep the transmitter-side phase matrix constant for indefinite time durations.

Our system uses centralized precoding. All channel coefficients collected at the different access points are sent over a wired backbone to a single server. The server computes the precoding coefficients using the Zero-Forcing technique [CS03].

■ 4.3 Efficient Estimation

As previously mentioned, hosts must perform periodic estimation of their relative timing offsets and carrier frequency offsets, since these quantities vary in time according to the thermal drift of their oscillators. We choose to interleave estimation periods with data transmission periods in a TDD manner, as illustrated in Figure 4.3. For the level of precision required by coherent beamformed transmission, the estimation procedure must be repeated several times per second. Moreover, every pair of neighboring access points must produce estimates of the two quantities mentioned.

In our design, each access point transmits beacons that are used by its neighbors for determining their relative frequency offsets. Ideally, all beacons should be received

without interference by the sender's neighbors during the estimation process. Scheduling the transmission of beacons without receiver-side interference in order to cover every link in a minimal number of slots is a well-known problem in adhoc networks. Our system has the advantage of a wired control plane which allows for solving this allocation problem in a centralized manner. Exact and approximate solutions to this problem are well-known.
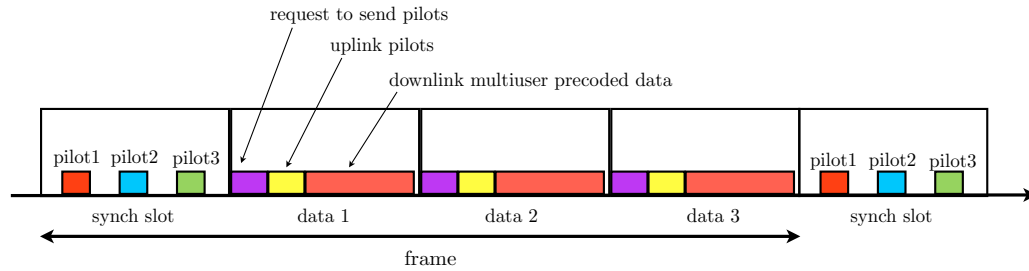
Figure 4.3: **Superframe.** The slot structure includes a synchronization period followed by a downlink transmission period. [RBP$^+$]

The required precision for each of the estimates mentioned is recorded in Table 4.3. The table also includes the rate of change of each parameter in the case of the oscillators used in our wireless platform. These rates of change have been established either experimentally (in the case of the thermal induced CFO drift) or by analyzing the clock characteristics presented in the datasheet.

Interpreting the values in the table motivates the design decisions behind the synchronization procedure. The first thing to notice is that for the high SNR regime corresponding to wireless links between neighboring access points in WiFi deployments (for example, above 20 dB), achieving the required CFO estimation precision requires a pilot overhead much larger than achieving the timing estimation precision [RBP$^+$] when using standard estimation techniques (e.g. PN-based timing offset estimation [TEF99a] and sine-based CFO estimation [Tre85, Kay89]). Moreover, the drift in the case of the CFO is essentially a random process which must be continuously measured an tracked, possibly using a Kalman filter. By contrast, the drift in the case of the timing offset is frequency-offset induced and occurring at a constant rate. The evolution of the timing offset can be readily predicted from the CFO value. For example, for our oscillators,

Parameter	Required Precision	Rate of Change
CFO	10 Hz	≈ 100 Hz/sec
SFO	5 Hz	$\approx .5$ Hz/sec
Timing offset	200 ns	1500 nsec/sec

Table 4.1: Estimated variables.

which have 1.5 ppm frequency stability, the expected drift within a second is 1.5 μsec. This corresponds to a carrier frequency offset of about 4 kHz. However, looking at Table 4.3, the carrier frequency offset can be estimated with a 100 Hz precision, therefore the timing drift over the same second interval can also be estimated with a ≈ 40 nsec precision. Therefore an accurate CFO estimator greatly reduces the number of measurements needed for timing offset estimation. It results from the above considerations that frequency offset estimation is the main problem to be solved and incurs the greatest pilot overhead.

We have decided to adapt a standard frequency estimation technique used by many OFDM-compatible transmission schemes, namely the Schmidl-Cox estimation procedure [SC97]. Standard Schmidl-Cox estimation uses the cross-correlation between two identical pilot sequences spaced one symbol duration apart in order to determine the phase change over the symbol duration. We introduce a simple modification that improves the accuracy of the estimator without increasing the pilot overhead. By increasing the spacing between the two sequences, the sensitivity of the phase change estimate to channel noise is severely reduced. The risk that we incur by increasing the spacing is that the phase might rollback without the system being able to detect it, producing a significantly lower frequency offset estimate than the correct value. In fact, the spacing of the sequences determines the maximum frequency range that the estimator can cover. The system designer must either be able to guarantee that the oscillators will not produce frequency offsets larger than this range or use a second, less accurate frequency estimator that works over shorter periods, for example over the period of a single sequence, in order to place the accurate estimate in the right frequency bin. We have chosen the later option in our design.

Figure 4.4 illustrates a CFO estimation period. The beacons transmitted by different APs are interleaved in order to take advantage of the spacing of the two sequences comprising them. Generally, the sequence length corresponds to a few OFDM symbols, which is smaller by an order of magnitude than a typical data packet duration. If the chromatic number of the interference graph is sufficiently low, the whole estimation procedure can be executed during a single packet transmission time, i.e. in less than a millisecond.
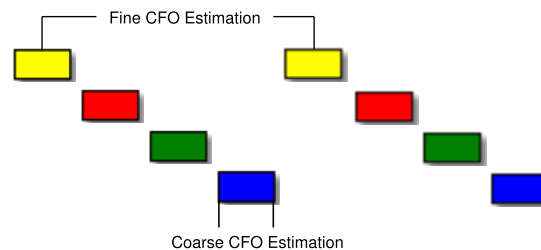
Figure 4.4: **CFO Estimation Period.** The beacons of different APs are interleaved in time. The increased spacing between the constituent sequences allows for a finer frequency estimate.

Once all the relative carrier frequency offsets measurements are available, the central controller solves a constrained least squares optimization problem in order to combine them into a uniform view. This procedure is optimal under the assumption that carrier frequency offset measurements are linear, which for small oscillator imprecision is a reasonable assumption.

For a more detailed description of the synchronization slot and a model of the oscillator-induced effects on the transmitted signals, we direct the reader to [RBP$^+$].

After estimating the timing and the frequency offsets, they can be compensated for each slot separately, Hosts maintain constant their relative timing offsets during the slot, allowing for small timing drifts. However, over a slot duration these drifts are significantly smaller than the duration of a cyclic prefix and, after the transmitters align their symbol transmission, also OFDMA operation is possible. To achieve MIMO operation, which also requires phase coherence, CFO-effects have to be also compensated. Remember that the value of the frequency offset depends on the subcarrier frequency, since it is actually the result of a time compression/dilation due to different sampling speeds. For

each subcarrier, for small frequency offset values, we can compensate the CFO by inducing phase changes in the transmitted soft symbols, i.e. rotating them to compensate the CFO-induced rotations [2]. In our system, receivers also compensate their frequency offsets, however in a general deployment this is not a requirement and this optimization can be sacrificed in order to preserve compatibility with legacy devices. The effect of maintaining constant the timing offsets and compensating for the frequency offsets is that the effective channel matrix in Equation 2.13 is preserved constant for the duration of a downlink slot.

In each slot, the effective matrix will still include a set of random transmitter-side phases and received-side phases. Therefore the matrix changes from slot to slot depending not only on the changes in the channel (which occur at the rate of the channel coherence time) but also on the two diagonal matrices that describe the mentioned phases. The effective channel can be estimated in each slot using downlink channel estimation involving downlink pilot transmission followed by feedback from the clients, uplink channel estimation involving beacons transmitted form the clients to the network-side transmitters or a phase estimation procedure among the transmitters similar to the synchronization procedure used for estimating carrier frequency offsets. While the testbed implementation used downlink channel estimates, the system design is compatible with uplink channel estimation as well. We have considered uplink channel estimation since it measures the effective channel matrix, i.e. it combines the antenna-to-antenna channel matrix estimation with estimating the two diagonal matrices. The advantage over downlink channel measurements is that while downlink channel estimation involves a feedback step in which the measurement are reported over the air in quantized form, with a high overhead, uplink estimation only transmits the pilots in analog form. When

---

[2]This approach does not effectively change the frequency of the transmitted subcarriers, but for a small ratio of CFO/subcarrier bandwidth it does not produce significant inter-carrier interference (ICI). Alternatives include choosing the CFO of the null subcarrier as a common CFO value and using a sine synthesizer to change the frequency of the transmitted signal by this amount while still applying the pre-IFFT phase rotations to compensate for the CFO differences between subcarriers or renouncing the IFFT completely and generating each subcarrier signal separately using a frequency synthesizer. The later approaches are suitable for LTE-compatible networks where the subcarrier separation is much smaller than in WiFi networks and ICI effects are more of a concern. In fact, in LTE networks the clients, which use low-precision oscillators prone to generating frequency offsets leading to significant ICI, correct their signals based on a base-station broadcasted frequency reference before transmission.

using the slot structure presented in Section 3.3, the overhead is only one OFDM symbol/client.

The basic assumption involved in realizing a system based on uplink estimation is that of reciprocity between the downlink and the uplink transmission channels (see [SYA$^+$12] for details). Specifically, the downlink channel coefficients and uplink channel coefficients for each subcarrier should be identical up to a complex scaling value which is constant across measurements.

When dealing with channel changes, we would like to know to which extent the changes are determined by actual antenna-to-antenna channel changes (the matrix $\mathbf{H}(n; t)$ or by electronics induced phase changes (the matrices $\Theta_m(n; t)$ and $\Phi_m(n; t)$). For this purpose, we estimate the timing offsets and random phase factors from the measured channel and compensate them in order to gain more information about the actual channel.

It turns out that estimating the channel offset over the different subcarriers is analogous to a well-known problem in communication theory, namely estimating the frequency of a noisy sinusoid. At high SNR, solutions to this problem have been provided by [Tre85,Kay89]. The following exposition adapts their treatment of the problem to this situation, mentioning along the way a few key differences.

Consider the end to end channel between transmitter m and receiver k:

$$\widetilde{\mathbf{H}}_{km}(n; t) = e^{j\left(\frac{2\pi}{NT_s}(\tau_m - \delta_k)n + (\phi_m(t) - \theta_k(t))\right)}\mathbf{H}_{km}(n; t) \tag{4.1}$$

It follows that

$$\angle\widetilde{\mathbf{H}}_{km}(n; t) = \left(\frac{2\pi}{NT_s}(\tau_m - \delta_k)n + (\phi_m(t) - \theta_k(t))\right) + \angle\mathbf{H}_{km}(n; t) \tag{4.2}$$

Assuming that the n-th subcarrier is measured at time t with a pilot received with power P and assuming a receiver noise power N, under the high SNR approximation [Tre85], the SNR figure for the measured angle is $\frac{2P}{N}$.

Linear regression offers us a way of estimating the parameters $\tau_m - \delta_k$ and $\phi_m(t) - \theta_k(t)$. Assume that the angles in Equation 4.2 have been unwrapped along the n variable.

In the following we will have to ignore the presence of the $\angle\mathbf{H}_{km}(n;t)$ factor, which is nonlinear. The standard estimators for the two parameters in the case of N_c subcarriers numbered $0, 1, \cdots, N_c-1$ placed symmetrically around the center frequency are [Tre85]:

$$\widehat{\tau_m - \delta_k} = \frac{12}{N_c(N_c^2 - 1)} \sum_{c=0}^{N_c-1} c\angle\widetilde{\mathbf{H}}_{km}(c;t) - \frac{6}{N_c(N_c + 1)} \sum_{c=0}^{N_c-1} \angle\widetilde{\mathbf{H}}_{km}(c;t) \tag{4.3}$$

and

$$\widehat{\theta_k(t) - \phi_m}(t) = \frac{1}{N_c} \sum_{c=0}^{N_c-1} \angle\widetilde{\mathbf{H}}_{km}(c;t) \tag{4.4}$$

Equivalently, Kay [Kay89] proposes the following form which at high SNR is also equivalent to the maximum likelihood estimator:

$$\widehat{\tau_m - \delta_k} = \sum_{c=0}^{N_c-2} w_c\angle\widetilde{\mathbf{H}}_{km}^*(c;t)\widetilde{\mathbf{H}}_{km}(c + 1;t) \tag{4.5}$$

with

$$w_c = \frac{\frac{3}{2}N_c}{N_c^2 - 1} \left\{ 1 - \left[\frac{c - \left(\frac{N}{2} - 1\right)}{\frac{N}{2}}\right]^2 \right\} \tag{4.6}$$

which avoids the need for phase unwrapping.

The variances of the estimated values can be easily computed [Tre85, Kay89] and meet the Cramer-Rao bound for this problem [Tre92, RB74]:

$$\mathrm{Var}(\widehat{\tau_m - \delta_k}) = \frac{1}{(N_c - 1)^2 \frac{P}{N}} \tag{4.7}$$

while

$$\mathrm{Var}(\widehat{\phi_m(t) - \theta_k}(t)) = \frac{1}{N_c} \frac{1}{\frac{2P}{N}} \tag{4.8}$$

Returning to the factor $\angle H_{km}(n;t)$, we can see that it induces the following biases into the estimates:

$$\text{Bias}(\widehat{\tau_m - \delta_k}) = \frac{12}{N_c(N_c^2 - 1)} \sum_{c=0}^{N_c-1} c \angle \mathbf{H}_{km}(c; t) - \frac{6}{N_c(N_c + 1)} \sum_{c=0}^{N_c-1} \angle \mathbf{H}_{km}(c; t) \quad (4.9)$$

and

$$\text{Bias}(\widehat{\phi_m(t) - \theta_k}(t)) = \frac{1}{N_c} \sum_{c=0}^{N_c-1} \angle \mathbf{H}_{km}(c; t) \quad (4.10)$$

These biases make the estimates channel-dependent and prevent us from using the timing and phase estimate in order to locally correct for the random phases of the transmitters. However, if we know that the channel $H_{km}(n; t)$ are constant in time, we can readily see that the biases will be constant and the transmitters can compute exactly their timing and phase offsets. It results that, with a fixed channel, the transmitters can use a set of previously computed downlink coefficients and just compensate their own phase changes, which are now known. With a constant channel to the user, the transmitters can thus avoid repeating the precoding procedure.

While the clients to the users cannot be assumed to stay constant, the channels between the access points themselves are constant to a large degree, so a set of inter-access point estimations can be used to determine the relative phase changes and correct for them. If the changes are measured relative to a single other access point, we are back to the reference-based synchronization scenario. If the changes are measured between all pairs of access points, the transmitters will need to centralize their measurements, similarly to the procedure applied in the case of frequency estimates in order to compute the necessary adjustments.

Assuming we have a set of channel measurements taken at different times we can use the above estimates to bring them to a canonical form:

$$C(\widetilde{\mathbf{H}}_{km}(n; t)) = e^{j\left(\frac{2\pi}{NT_s}((\tau_m - \delta_k) - (\widehat{\tau_m - \delta_k}))n + (\phi_m(t) - \theta_k(t)) - (\widehat{\phi_m(t) - \theta_k}(t)))\right)} \widetilde{\mathbf{H}}_{km}(n; t) \quad (4.11)$$
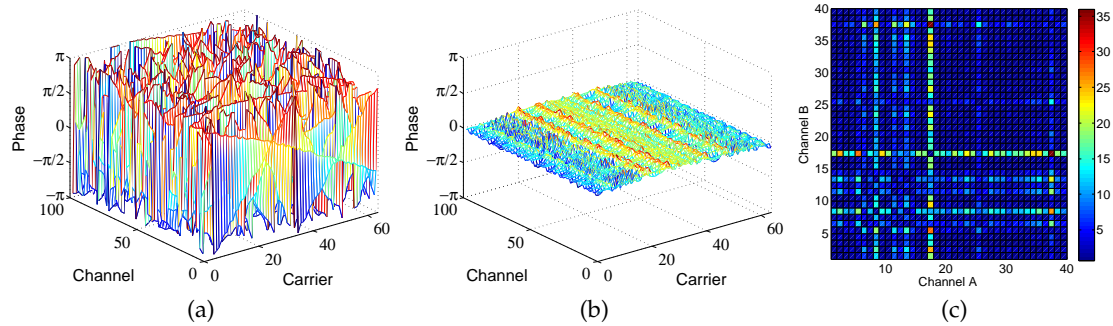
Figure 4.5: **Channel Estimation.** (a) Channel measurements (phase angle) before timing and phase-offset compensation. (b) Channel measurements (phase angle) after bringing the measurements to canonical form. (c) Distance between canonical forms for different pairs of random channels.

In order to evaluate the precision of our canonical form, we generated a set of random channel instantiations according to a 10-tap channel model with coefficients generated according to a Rayleigh-distribution with amplitude decreasing exponentially in the tap count and with random phases. Figure 4.5a illustrates, for a single channel, the measured phases for each subcarrier under a set of random timing and phase offsets, where the timing offsets are uniformly random from $-CP/5$ to $CP/5$ where CP is the length of the cyclic prefix and the phase offsets are uniformly random over $[0, 2\pi)$. The same channel measurements in canonical form are represented in Figure 4.5b. It can be seen that the compensation applied almost completely removes the timing and offset effects.

Encouraged by these results, we wanted to see whether the canonical form of the channel could be an indicator of whether the channel has significantly changed. For pairs of different random channels, with random timing offsets we have used the l^2 metric in order to quantify the distance between the phase vectors. The results are illustrated in Figure 4.5b, where for each pair of channels the likelihood has been obtained by averaging over a large number of timing offsets. While for the same channel the distance remains low, for a large number of pairs of different channels the distance does not increase to much larger values, suggesting that this method cannot be used to precisely distinguish between different antenna-to-antenna channel measurements.

■ 4.4 Performance Evaluation

Our transmission system is comprised of two WARP radios configured as access points, connected to the same server and another two WARP radios configured as clients. One of the two access points broadcasts synchronization beacons used by the other three hosts in correcting the frequency of incoming and outgoing signals. [3]

The experiments have been conducted in an anechoic chamber and the hosts have been placed as shown in Figure 4.6 at various positions. The two blue hosts on the right side serve as access points while the two red hosts placed near them serve as users. As mentioned, one of the access points transmits reference beacons used by the other hosts for obtaining timing and frequency offset estimates. The access points use the timing estimates not only to align their OFDM symbols, but also to identify and count them in parallel fashion, making it possible to start packet transmissions from different access points on the same symbol. In our experiment, each packet transmission takes place according to the following sequence of operations: first the downlink channel is estimated and the coefficients are collected at the server. The implementation uses 64 subcarriers and, for each subcarrier, the I and Q components of the channel coefficients are each sampled with 16 bit precision. It results that for every link in the system, the channel coefficients add up to 256 bytes. Each user collects a set of channel estimates for each AP, that is each user must measure the channel coefficients for two links, resulting in a total of 512 bytes of data transferred by each user to the server. The server, in turn computes the precoding coefficients and returns them to every access point in a message equal in size to the one received from an user, that is 512 bytes. Once all precoding coefficients have been broadcasted, which the access point can observe by sniffing the packets broadcasted over the wired network, the access points start transmitting coherently, starting on the same symbol.

---

[3]Due to the small scale of our setup, we have not implemented the entire synchronization protocol using least squares estimation. According to the specification of the full implementation, both access points should be broadcasting beacons, giving a second source of beacons to be used in the frequency offset estimation process. However, we believe that in this case the estimation results obtained with two beacon sources would closely match the ones obtained with a single source.
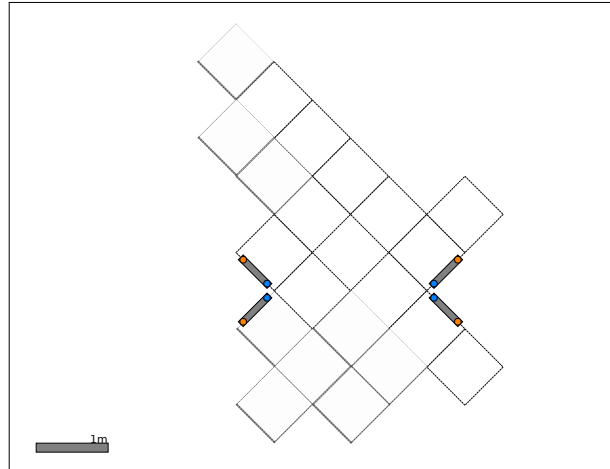
Figure 4.6: **Placement Diagram for 4x4 Experiments.** The positions of the radios in the anechoic chamber are marked in the above figure. The transmitters are marked in blue and the receivers are marked in red. In the 2x2 experiments all the nodes were placed in the positions marked on the right side of the diagram.

The users record the received data and attempt to decode it. If the channel coefficients have been determined correctly and the carrier frequency offsets are properly compensated, the transmitters should be able to accurately zeroforce the signals transmitted to the users, all but eliminating any interference between the signals destined to different users.

While our implementation includes idle times, in a practical deployment the passing of coefficients to and from the server and the downlink data transmissions could be pipelined and done in parallel. Due to the intricacies of running multiple consecutive experiments on the WARPs, we have not implemented this optimization.

Synchronization

The first set of experiments is concerned with the accuracy of our frequency offset estimation. We used simple sine signals, broadcast from an access point to an user, in order to get an upper bound on the frequency offset between the two. Figure 4.7 shows the distribution of the measured frequency offsets both when the transmission source was the same access point broadcasting synchronization beacons and when the transmission source was the other access point.
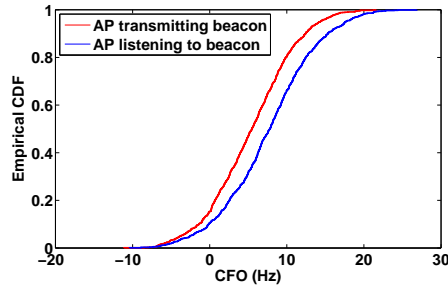
Figure 4.7: **Residual CFO distribution.** The empirical cumulative distribution function of the frequency offsets of the AP that sends the beacon and the AP than listens to it.

We next wanted to quantify the effects of the remainder CFO between the two access points on the phase coherence of the access points, namely measure the difference between their relative phase at the time of channel estimation and their relative phase at the time of data transmission, which occurs up to one millisecond later. For this purpose, we broadcasted pilots from the two access points both during the channel estimation period and, without making use of any precoding coefficients, during the downlink data transmission period. We measured the change in relative angle between the pilots from the two APs. Figure 4.8 shows the distribution of our results.
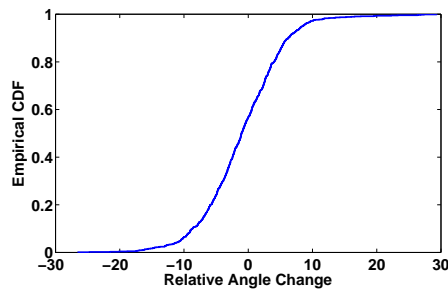
Figure 4.8: **Angle drift distribution.** The empirical cumulative distribution function of the relative angle change between the pilots of the two APs.

Finally, we have used the measured pilots in order to Zero-Force our data transmissions. We measured the channel SINR corresponding to each user channel in order to determine the accuracy of our signal separation. The resulting distribution, for the two users is plotted in Figure 4.9. The results are consistent with the angle differences measured in the previous experiment.
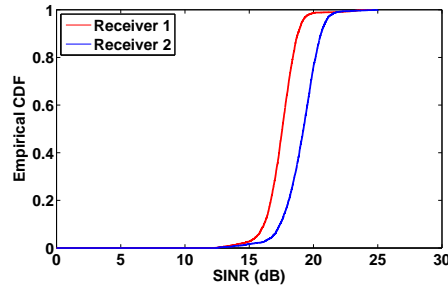
Figure 4.9: **SINR Distribution.** The cumulative distribution function of the channel SINR for each user.

The high SINR figures allow for sending relatively clean QAM-modulated signals to the two users, as illustrated in Figure 4.10.
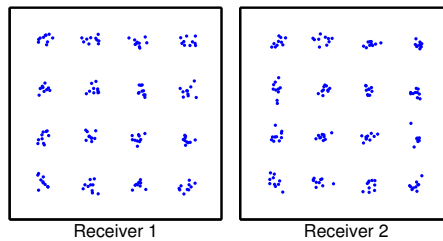
Figure 4.10: **Zero-Forcing Scattering Diagram 2x2.**   The scattering diagram at the receivers of two independent data streams concurrently transmitted from two APs.

In a similar experiment with four access points having each a receiver nearby we have measured the scattering diagrams found in figure 4.11. We have varied the transmitted power in order to assess the rate gains over point-to-point transmission.
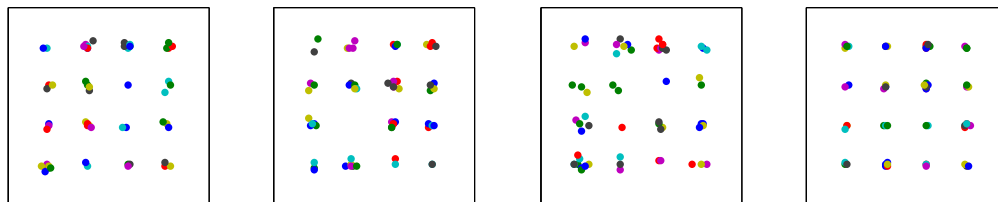
Figure 4.11: **Zero-Forcing Scattering Diagram 4x4.**   The scattering diagram at the receivers of four independent data streams concurrently transmitted from four APs.

The comparative medium rates are illustrated in figure 4.12. The average rate gain in sum rate is about 2.65, closely mimicking the behavior of the reference-based distributed MIMO system. We can conclude that the modifications that enable scaling up the size of the deployment do not affect the achieved performance.
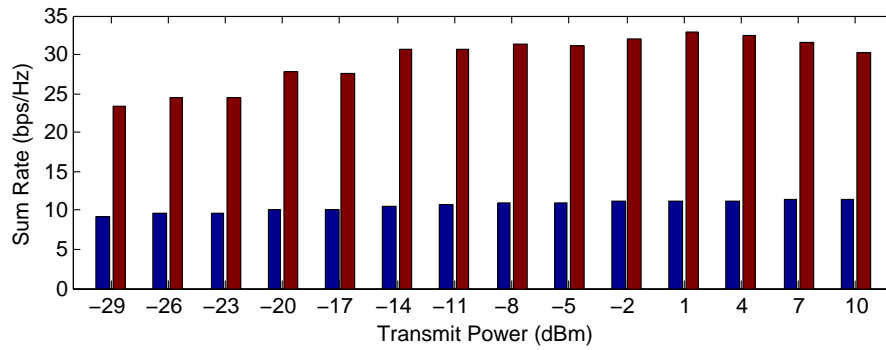
Figure 4.12: **Rates of Multiplexed MIMO Transmission vs Point-to-Point Transmission.** The sum rates obtained through multiuser transmission with four multiplexed streams are about 2.65 times higher than the average rates of point-to-point transmissions.

The reference-based system and the current system are similar in the way in which they perform channel estimation and precoding. An open question is how these factors affect the achievable performance and whether they leave space for further improvements in each of the two systems considered.

CHAPTER 5

Tag Spotting at the Interference Range

The current chapter introduces tags, wireless primitives aimed at message passing in adhoc networks. We discuss the construction of tags, conduct a theoretical analysis of their performance, present their overhead, evaluate them experimentally under different interference conditions. Through two sample applications in scheduling and congestion control we showcase the way in which tags enable a large number of capacity-sharing schemes based on information sharing. As opposed to regular message passing methods, we show that tags can reach beyond the data transmission range to the edge of the interference range of the transmitter.

■ 5.1 Introduction

Many of the challenges encountered in the design of wireless networks with multiple transmission and reception points stem from the quirks of wireless signal propagation. Using currently prevailing transmission techniques, wireless signals cannot be focused exclusively towards their intended recipient, making wireless an inherently shared medium. Wireless transmissions are local in their coverage, and, in general, no sender or receiver will have access to complete channel state information. Because of these characteristics, a wireless network is commonly modeled as a set of links among

which interference may occur depending on the particular choice of senders transmitting at the same time. The effects of wireless interference are far reaching, affecting all network layers, from physical layer and medium access to flow control and user satisfaction. They extend beyond the space of a single host or a single link, as flows that do not share any hosts or links in their paths might in fact find themselves competing for resources. Its direct consequence is unfairness leading to flow starvation and underutilization of available resources. A study of the exact mechanisms through which interference leads to unfairness reveals problems at multiple network layers. The most general statements of these problems frequently preclude finding a decentralized and optimal solution. However, interference is a local disruption, and therefore leaves hope that a local, if imperfect, solution may be found.

Distributed algorithms often make use of local exchanges of information. This creates a need for a communication backplane capable of connecting each host to the set of hosts affected by its transmissions. This requirement is more cumbersome than it might seem at first sight, for successful data transmission at common data rates requires rather large signal to noise ratios. The capacity of links is however affected even by interferers reaching them at far lower signal levels. Connecting the recipients of interference with transmitters requires thus a communication backplane capable of operating over channels offering low signal-to-noise ratios. This, in turn, implies that constructing such a communication backplane will require designing a physical layer different from the standard physical layers used for high rate data transmissions. Further differences arise from the fact that, in a wireless environment designed to support primarily data transmissions over short links operating at strong signal levels, long range communication is at best opportunistic. Backplane communication receivers are therefore required to be able to discriminate actual backplane transmissions from high levels of background chatter. Moreover, in order to offer a significant improvement without further aggravating the interference problem, communications along the backplane should not create new interference constraints.

In this paper we propose a signaling scheme enabling the creation of a communication backplane which meets all the above requirements. Our scheme induces a low

per-packet overhead, is resilient to high levels of noise and interference, and minimizes the disruption of data transmissions due to the interference that it induces.

Our first contribution is the design of tags, members of a set of signals designed to be easily detectable and recognizable in the presence of high levels of noise and interference, in the absence of time and phase synchronization and with only approximate frequency synchronization. Their increase in range over regular data transmissions is obtained in part through added redundancy. Tag signals are modulated using multitone modulation over a time duration that is larger than the duration of a regular data-transmitting tone. A tag is a distinct superposition of several tones whose frequencies are chosen according to the codewords of a binary algebraic code. On the receiver side, tags are recognized using a receiver based on spectral analysis.

What about interference caused by tags on data payload? OFDM, the prevalent modulation for data transmission in today's wireless networks, used in this study as well, exhibits flat time and frequency power densities. We will prove that interference disruptions caused by tags are comparable to the ones caused by data transmissions.

Our second contribution is an analysis of the performance of tags at different noise and interference levels and when making different design choices. Starting from detection theory principles, we derive, under a sufficiently general propagation model, the detection likelihood/false alarm likelihood curves at different SNR levels. We are particularly interested in evaluating the trade-off between transmitting more information (i.e. more bits per tag) and the corresponding increase in the likelihood of false alarm or misclassification. Our findings are later on compared to experimental results.

Our third contribution is the implementation and testing of Tag Spotting through experiments performed using a software radio platform in a testbed comprising senders, receivers and interferers. This series of experiments makes use of a tag family capable of conveying about one byte of information. The results of our evaluation support the conclusion that communication through tags is effective at SNR[1] values as low as 0 dB

---

[1]Throughout this presentation we understand the noise part of the SNR figure to also include interference power, unless specifically noted otherwise.

and is robust to the effects of interference. In addition, through a series of channel simulations, we establish the possibility of constructing tags capable of conveying more bits of information and we survey the effects of design decisions on the system performance.

Our fourth contribution is the use of Tag Spotting in two applications, showcasing the performance improvements brought by the existence of a control plane able to reach all nodes within the interference range. Specifically, we use tags to efficiently implement a state of the art congestion control scheme for multi-hop networks which requires neighboring nodes, i.e, nodes that interfere with each other, to exchange control information in an effort to fairly share the available bandwidth. We also use tags in order to design and test a simple MAC-layer signaling mechanism meant to prevent the starvation of TCP flows in multi-hop wireless networks.

This chapter is organized as follows: In Section 5.2 we give an overview of related work and indicate some congestion control and scheduling mechanisms that would benefit from the use of tags. Section 5.3 introduces multitone modulation along with a simple model for estimating the spectral footprint of multitone signals and discusses techniques for limiting inter-carrier interference. Armed with the conclusions of Section 5.3, we proceed in Section 5.4 to describe in detail the structure of tags and construct a tag detector capable of distinguishing tags from interference. Section 5.5 discusses in detail the rationale behind each design decision present in tag construction. Section 5.6 experimentally evaluates, using a software radio platform, the communication range of tags as well as the rate of false detections. It is shown experimentally that tags can be reliably identified at SNR values as low as 0 dB while the likelihood of false detections can be sufficiently limited. The same section evaluates the impact of noise, tag transmissions, and data transmissions, on each other. Section 5.7 presents an analysis grounded on detection theory principles of the effects of different choices available in the process of designing the modulation of tags and, more importantly, of the trade-off between the data rate of tags (i.e. the number of bits transmitted) and the likelihood of false alarms. Two examples of using tags to facilitate the implementation and improve the performance of congestion control and scheduling are presented in Sections 5.8 and 5.8. Finally, Section 5.9 concludes the paper.

■   5.2   Related Work

1. **Communication and Detection Theory.** Tag Spotting is closely related to a classic
research topic in communication theory, namely information transmission at low
signal-to-noise ratios. The motivation of this research has varied from securing the
transmitted data such as in the case of spread-spectrum communication [SOSL94]
to protection against interference in the case of the widely used CDMA standard
[Vit95] and to achieving long-range transmission [RBMR63].

Tags employ a multicarrier spread-spectrum modulation. They are clearly related
to MC-CDMA [YLF93], however they use a non-coherent modulation and do not
use orthogonal codewords. Like multitone FSK [LM02], tags use a combination of
tones in order to transmit information.

The design of the tag detector presented in Section 5.4 is based on the detection the-
ory of multipulse signals with constant amplitudes and unknown phases. While
the classical detector for such a situation is well-studied and understood (see, for
example [MW95, Tre92]), it requires a precise estimate of the background noise
level in order to set appropriate detection thresholds. Interference from compet-
ing packet transmissions will confront tags with different levels of background
noise, making a precise and timely estimate impossible. Our detector is indepen-
dent of the level of background noise, requiring only a base SINR as prerequisite
for the accuracy of the detection. In the appendix we apply a theoretical analysis
similar to the one of the classical detector in order derive the detection/false alarm
trade-off curves of our own detector.

2. **Physical Layer Extensions.** In the wireless networking world, carrier sense [BM09]
can be seen as an example of a message passing mechanism operating beyond
the data transmission range. Closely related is the use of dual busy tones [HD02]
in order to signal channel occupancy. A recently proposed physical layer exten-
sion, CSMA with collision notification CSMA/CN [SRCN10], aims at reducing the
impact of collisions through an early termination signal sent by the receiver of the

colliding packet. The transmitter-based detector uses self-interference cancellation techniques in order to improve the SNR of the reciprocal channel and detects the termination signal using correlation, in a manner similar to [TEF99b]. However, as the authors of both these papers find out, a correlation based receiver cannot function without prior channel and frequency offset estimation, which prevents their use for broadcasts over arbitrary channels, as in the case of tags.

Carrier sense, dual busy tones and collision notifications are binary signaling mechanisms, not suited for transmitting numeric information, as required by message passing protocols. Another recent physical layer extension [WTL+10] aims at realizing a side-channel over spread-spectrum based protocols through perturbations of certain chips comprising a transmitted symbol. These perturbations are in turn compensated for by the normal error correcting codes employed in data transmission, thus allowing in most cases for the payload to be decoded correctly, and they are also detected by a special pattern analyzer, allowing for the transmission of side information. While the motivation and design constraints of this works are similar to ours, the design choices made in the creation of tags are very different. Most importantly, tags do not require the data transmission scheme to be spread-spectrum based.

The technology of software defined radios [Ett, GHMS09] has acted as an enabler for some of the recent advances in multiuser wireless network research. It allowed, for example, the experimentation of techniques such as zigzag decoding [GK08], interference cancellation [HAW08] or dynamic bandwidth adaptation [CMM+08]. Perhaps the most similar technique to the one presented in this paper is the one of smart broadcast acknowledgments, introduced in [DSGS09b], in which multitone modulation is used for the purpose of simultaneously conveying positive acknowledgments from multiple receivers. Other recent advances that also make use of software radios for evaluation purposes include multi-user beamforming [AASK10], fine grained channel access in which bandwidth allocation based on an increased number of carriers is coupled with frame synchronization used in

creating an effective uplink OFDM implementation [TFZ$^+$10] and frame synchronization used in obtaining diversity gains [RHK10].

3. **Congestion Control and Scheduling.** Prior works on congestion control for multi-hop wireless networks differ in the way in which congestion is reported to the source. One class of schemes sends implicit or imprecise feedback by dropping or marking packets [XGQS03, RJJ$^+$08] in the tradition of TCP congestion control [Jac88], or by regulating transmissions based on queue differentials [WJHR09] along the lines of back-pressure ideas [TE92]. Another class of schemes [RJJ$^+$08, TJZS07] explicitly computes available channel capacity and then sends precise rate feedback, in the spirit of wired network congestion control mechanisms such as XCP [KHR02] and RCP [DKZSM05].

In an effort to tackle the complexity of creating optimal schedulers, recent work on medium access for multihop networks has proposed distributed algorithms capable of approximating the optimal solution [Sto05] [LS04] [JPPQ03] [KMPS05] [JLS09]. A common theme here is the use of local, neighborhood-centered information in achieving a global solution. Our work is partly motivated by the recent development of a number of congestion control and scheduling schemes for multi-hop wireless networks that are based on the local sharing of information, such as, for example, [RJJ$^+$08] and [Chi05].. Local information is at the heart of several other wireless multi-hop problems: neighbor discovery [VTGK09], reliable routing [YKT03], capacity estimation [XGQS03] and signaling congestion and starvation. The mechanism proposed in this paper offers an efficient way to implement the neighborhood-wide sharing of control information in the schemes mentioned.

While these schemes append control plane information to data packets and rely on packet overhearing, it has been recognized that this information needs to reach all nodes within the carrier sense range of a node of interest. The information sharing mechanism proposed in this paper eases the implementation of many of these ideas and improves their performance (see Section 5.8), as control information will reach nodes outside the data transmission but within the carrier sense range.

■   5.3   Intercarrier Interference

An OFDM encoder uses an inverse discrete Fourier transform in order to encode a sequence of symbols into a set of tones over a finite time interval, from here on named either a frame or an OFDM symbol. Consider the sequence x of length $N = 2^k$ whose elements are chosen from a signal constellation, arriving for encoding at an IFFT (inverse fast Fourier transform) frame generator. The resulting signal will be generated according to the formula:

$$X(t) = \sum_{\forall k \in \{0, N-1\}} x_k e^{ik2\pi t}.$$

Passing to the continuous Fourier transform exposes a windowing effect and provides us with an insight into the spectral footprint of the generated signal. The discrete spectrum of an OFDM frame is illustrated in the upper half of Figure 5.1 and corresponds to the original encoded symbol sequence x. Since the summed exponentials are bounded in time, their continuous transforms are sinc functions and the spectrum of the encoded signal is the sum of these sinc functions, as depicted in the lower half of Figure 5.1. The orthogonality of different signals is preserved: the peak of each sinc function is aligned with the zeros of all other sinc functions.
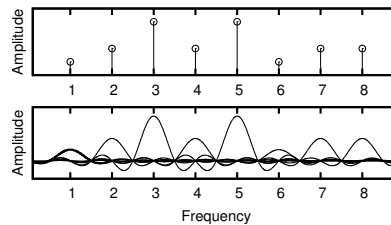
Figure 5.1: **Frequency domain view.** The discrete and continuous Fourier transforms of an OFDM frame.

Let's focus on non-coherent communication, i.e. we assume that the phase of each received signal is independently random, and try to derive a concentration result for the power leaked outside the intended transmission bandwidth. In Section 5.4 we will use this result in order to motivate that the modulation of tag signals exhibits a fast spectral decay and limited inter-carrier interference. The reasoning behind our approach
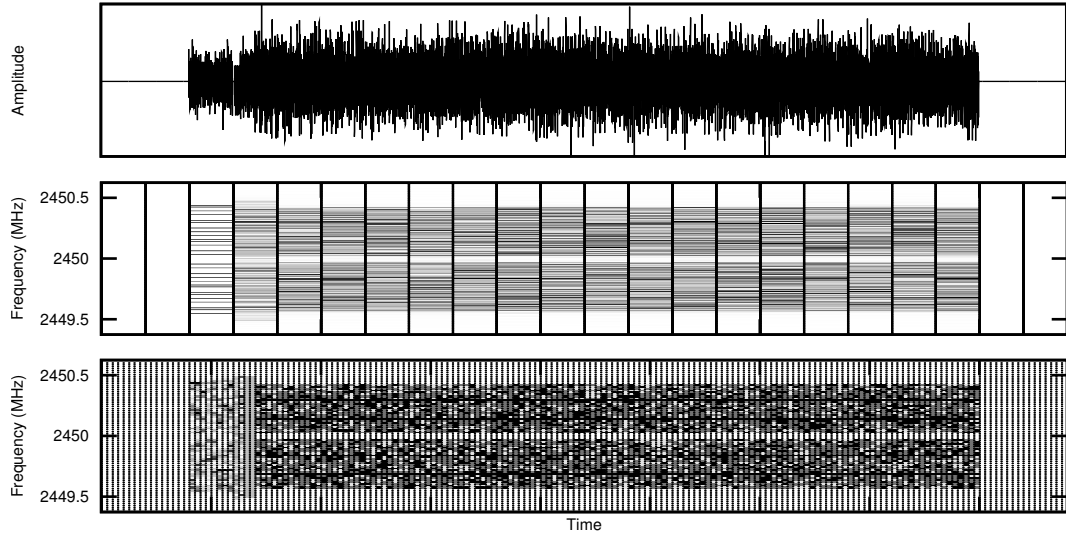
Figure 5.2: **Packet transmission.** The signal $s(t)$ (upper pane), a time frequency representation using 512 frequency bins (middle pane) and a time frequency representation using 64 frequency bins (lower pane).

is encompassed in Equations (5.1) and (5.2) below, which describe the effects of self-interference on OFDM signals and give the speed of their spectral decay. In the presence of a small frequency offset δ, the interference signal added by a carrier with unitary amplitude to a carrier placed k positions away is

$$l(k, \delta) \;=\; \int_{[0,1]} e^{2\pi i(k+\delta)t} dt \;=\; \frac{\sin(\pi\delta)}{\pi(k+\delta)} e^{i\pi\delta},$$

and in power terms,

$$p(k, \delta) \;=\; \mathrm{sinc}^2(\pi(k+\delta)) \;\leq\; \frac{1}{(k+\delta)^2}. \tag{5.1}$$

It can be easily verified that the magnitude of the interference among two carriers does not depend on the absolute difference of carrier frequencies but only on the number of carrier positions separating them.

Consider now a carrier at carrier position zero and an infinite block of carriers starting at carrier position k. The average power leaked by this block of carriers into the zero-positioned carrier can be bounded according to Equation (5.1) to be:

$$p_b(k) = \sum_{c=k}^{\infty} \frac{1}{c^2} = \frac{\pi^2}{6} - \sum_{c=1}^{k-1} \frac{1}{c^2}, \tag{5.2}$$

where we assume that symbols on different carriers are independent. The series in Equation (5.2) is rapidly converging.

Based on the observations above we conclude that replacing a single carrier through a block of random-phase, lower-bandwidth carriers that occupies the same part of the frequency band will sharpen the spectrum of a multitone signal and accelerate its spectral decay. Moreover, the insertion of a relatively small number of null-carriers will significantly reduce interference between two neighboring active regions of the spectrum. Creating the replacement carrier block calls for an increase in the number of carriers, and a corresponding increase in the time length of the OFDM frame, as described fully in Section 5.4.

■ 5.4 Tag Spotting

1. **High Level Design Overview.** Tag Spotting uses a set of signals (tags) that are easily detectable in low SNR conditions. The design of tags is determined by a number of constraints. Firstly, due to their short timespan, the presence of multiple indistinguishable sources and the presence of varying levels of interference, the tag detector cannot perform accurate channel estimation, or achieve time, phase and fine frequency synchronization. Secondly, in order to protect competing data transmissions from further levels of interference, tags must abide by a maximal spectral power constraint, which prevents the use of a peaky transmission scheme such as multiple frequency shift keying (MFSK). Finally, a tag detector must perform identically in the presence of added interference, as long as the SINR remains unchanged.

To address these constraints, we employed a noncoherent communication scheme that spreads a tag's energy over the frequency domain. The number of carriers was increased in order to sharpen the signal's spectral footprint and ease spectral analysis through discrete Fourier transform even in the absence of complete frequency synchronization. The symbol sequences encoded over the different carriers were chosen according to an algebraic code, thus adding extra redundancy.

The current section presents the details of tag construction. The following section will present the reasoning behind the design decisions made throughout the construction.

2. **Multitone Structure.** Figure 5.2 presents a packet transmission in time domain representation as well as in two different time-frequency representations. The two time-frequency representations, pictured in the lower two panes, are realized by performing the spectral analysis of successive blocks of 512 samples (middle pane) or 64 samples (lower pane). As we will see, these lengths are natural choices for describing the structures of tags and packet payloads respectively. In the same representations, each vertical column corresponds to an analyzed block, while the horizontal line pattern present in each such column illustrates the block's power spectral decomposition. In order to make the representation more meaningful, we have stripped tags and data frames of their cyclic prefixes and we aligned the boundaries of the analysis periods with the boundaries of tags and data frames.

Tags are encoded using 512 OFDM carriers while packet information makes use of 64 wider data carriers. Therefore, in time domain, a tag spans a period equivalent to eight regular data frames, while in frequency domain each "wide" data carrier corresponds to eight "thin" tag carriers. The transmission begins with a tag, whose spectrum can be observed in the third column of the middle pane, and continues with data frames containing synchronization data followed by the packet's payload. To both tags and data frames we append a cyclic prefix which increases their respective lengths by $\frac{1}{4}$-th.

The spectrum of a tag is constructed, at a basic level, by transmitting a symbol from a 0-1 (on/off) constellation in the frequency space corresponding to each of the wide carriers. Wide carriers are in turn grouped into groups of two neighboring carriers, and in each group only one of the two carriers will be active. This last constraint renders the modulation inside each carrier group to be a form of binary frequency shift keying. For tag construction purposes, eight of the wide carriers have been designated null carriers, while the remaining 56 carriers give rise to 28 two-carrier groups. Every tag can thus be naturally mapped to a 28-bit string in which each bit marks the choice of state in one of the groups.

The structure of a tag in frequency domain is depicted in the lower pane of Figure 5.4, while a two-carrier group is depicted in Figure 5.3.

On the receiver side, the tag detector operates in the presence of small, tolerable but unknown frequency offsets, which may cause power spillage from the active wide carriers to the inactive carriers. As illustrated in Figure 5.3, we chose to send the entire signal power allotted to an active wide carrier using the central four of the eight thin carriers corresponding to this particular wide carrier. We also chose to encode the tones sent on these thin carriers using different random phases. Looking at Equation (5.2) it can be readily seen that the above choices reduce the amount of power leaked onto inactive wide carriers. A straightforward computation assuming a frequency offset distributed uniformly between zero and two thin carrier widths reveals that the expected power leaked is, in expectation, about 2.3% of the total power. Figure 5.4 further illustrates this aspect by presenting the spectrum of a received tag in the presence of a frequency offset equal to 10% of a regular carrier width. By comparing the distribution of the received signal power in the frequency bins corresponding to wide carriers (middle pane) with the structure of the transmitted tag (lower pane), it can be seen that the received power is concentrated in those bins which correspond to active wide carriers. As a note, the use of random phases in signal construction has one further advantage: it allows sampling tags from a larger signal set in order to limit their peak to average power ratio.

3. **Encoding.** As it is common in communication system design, modulation is supplemented by a coding layer. The purpose of this layer is to create a subset of maximally differentiable signals, and to reduce the number of hypotheses tested.

The tag signal construction detailed above produces 2^{28} different basic tags, too many for efficient detection and insufficiently distinguishable from each other. A further restriction makes use of a $(28, 60, 13)$ [2] nonlinear code proposed by Sloane and Seidel [SS70]. Out of the basic tags only those having binary representations corresponding to the 60 codewords of this code are preserved. We discuss the use of other codes in Section 5.6.
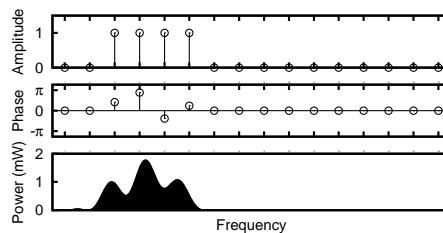
Figure 5.3: **Spectral Plot.** The discrete spectrum of a two-carrier group (amplitude and phase representation) and its continuous power spectrum.

Regular data frames dissipate the transmitted power over 48 carriers while tags make use of only 28 carriers. Since the average power spectral densities of used carriers in tags and data transmissions are equal, it results that the transmitted power in the case of tags is lower than the transmitted power in the case of data, as it can be observed in Figure 5.2.

4. **Constructing a Tag Detector.** Let $T = \{t_1, t_2, \ldots, t_{60}\}$ denote the set of all tags and C_i be the set of all data carriers activated when transmitting tag t_i. Let r_f denote the power of the received signal in the frequency bin corresponding to the

---

[2]The notation for codes used here is in the form (N, M, D) where N is the binary codeword length, M denotes the number of distinct codewords and D denotes the minimal Hamming distance between any two codewords.

f-th carrier. Our detector does not assume the channel phase response to be uniform and can therefore be used in a wideband scenario. We compute the following quantity which we will name from now on tag strength:

$$\frac{\sum_{f \in C_i} r_f^2}{\sum_{\forall f} r_f^2}. \tag{5.3}$$

Tag strength is compared against a fixed threshold γ and in case the threshold is exceeded a possible tag observation is recorded. [3]

Detection intervals have the same length as a tag from which the cyclic prefix has been removed and are spaced one tag cyclic prefix length apart. It results that successive analyzed intervals have significant overlap. Every transmitted tag will completely cover at least one detection interval. The detector processes every interval by first computing the Fourier transform of the contained signal and then computing, based on the resulting spectrum, the strength of each tag according to Equation (5.3). In order for a tag recognition event to be recorded, the corresponding tag strength must, firstly, exceed the threshold value and, secondly, be maximal among all tag strengths (for all tags) derived from detection intervals that overlap the current interval. A further detection metric effective in filtering off-band interference is computed for each detection interval by weighting the power levels in different carriers through the carrier's position in the frequency band, summing the resulting values and afterward dividing the result through the total interval power. As long as the resulting "center of mass" is placed in the central quarter of the frequency band, the tag observations are considered valid, otherwise they will be attributed to off-band interference.

---

[3]This equation is similar to the one of the low-SNR multipulse detector with n samples for a signal with unknown phase $\theta(t)$ varying at each pulse: $s(t) = A * \cos(2\pi f t + \theta(t))$, at a given SNR value $\alpha = \frac{A}{2N}$. Denote by H_S the hypothesis that signal s has been sent and by H_n the hypothesis that no signal has been sent. That detector is based on the equation $\log \frac{p(r|H_S)}{p(r|H_n)} \approx \sum_{i=1}^{N} r(t_i)^2 > \gamma' \frac{N}{\alpha}$ where γ' is a constant (see [MW95, p. 293]); in our case, the correction factor $\sum_{\forall f} r_f^2$ can be seen as an approximation of $\frac{A^2}{4} + N = (\alpha + 1)N$, meant to remove the linear dependence of the threshold on the noise power N, allowing thus for added noise-like interference.
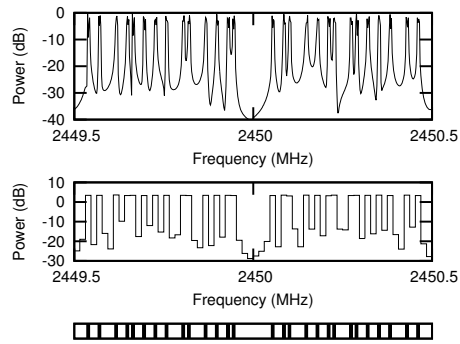
Figure 5.4: **CFO Effects.** The spectrum of a received tag in the presence of a frequency offset. Upper pane: the 512 frequency bins Fourier transform of the corresponding detector interval. Middle pane: the 64 frequency bins representation used in the detection decision. Lower pane: the structure of the transmitted tag.

In order to reduce the number of intervals analyzed and the likelihood of false alarms, a simple carrier sense scheme is employed. The receiver maintains a running estimate of channel noise and processes only those intervals for which the SNR exceeds -1 dB.

5. **Overhead.** Adding a tag to a packet incurs a transmission time overhead. Assume for now that only data packets are tagged and that a typical data packet has a payload of about 1500 bytes. Encoded using the parameters presented in Section 5.6, the payload will span 125 data frames, to which six synchronization frames are appended. A tag spans the equivalent of eight data frames and therefore its overhead is about 6.1% in terms of the normal packet duration. In some control schemes some of the data messages will not require tags to be piggybacked, allowing for a lesser overhead.

■ 5.5 Motivating the Design Choices

The previous section has presented in considerable detail the structure of tags. While the above description is complete, the decisions taken in the construction of tags may well seem arbitrary. The purpose of the current section is to motivate every design decision taken in tag construction.

Multitone Structure

The design space of tag signals is frequency space. At the lowest level, tags use the same form of modulation in frequency space, using orthogonal signals over a finite time interval, that is used in regular data transmission.

1. **OFDM.** In an opportunistic reception system that searches for the short occurrences of tags, the price of exact frequency synchronization and of exact channel estimation should be avoided. Tags are therefore transmitted and received without performing channel estimation or frequency synchronization. This raises a challenge in solving inter-carrier interference.

   Remember that the choice of OFDM for data transmission is linked to the propagation behavior of wireless signals. Since sine signals are the eigenfunctions of the wireless channel, the use of orthogonal sines along with an appropriate cyclic prefix is meant to prevent any channel-caused interference among the different transmitted symbols. The limited timespan of the transmission interval precludes the use of actual sines or any other signals with narrow support in the frequency space. As it was discussed in Section 5.3, the actual signals used in the transmissions have a slowly-decaying spectral footprint. In OFDM data transmission, the inter-carrier interference which the slow spectral decay entails is avoided through exact frequency synchronization, using the fact that in frequency space the zeros of the base sine-like signals align with the peaks of all other signals in the base set. In contrast, for tags, the packing of noncoherent carriers into larger building blocks reduces the amount of power leaked among the frequency bins corresponding to thick carriers. This reduction in leaked power allows the system to function as intended even when the receiver is not frequency-locked onto the transmitter. The lack of frequency synchronization together with the lack of an estimate of the channel phase response at different frequencies also leads to uncertainty regarding the phase of the transmitted signal. The use of a noncoherent encoding allows us to overcome the lack of knowledge of the channel phase response without further complications. It could be argued that these channel characteristics should be

measured in advance. However, our receptions are at best opportunistic, and the channel could be any one of a multitude of fast changing channels between any pair of hosts. Certainly, obtaining an exact estimate of the channel response and of the frequency offset, at the low SNR levels for which our system is designed, would greatly complicate the tag transmission problem.

2. **Fading.** Another characteristic of wireless channel transmission, frequency-selective fading, provides the rationale for the use of groups of two carriers as a encoding unit: since neighboring data carriers are likely to experience similar fading and since any of the codewords makes use of either one or the other of the two carriers in a two-carrier group in order to encode a bit value, fading over a two-carrier group will not induce a bias towards any of the hypotheses that a particular codeword has been transmitted. The received power and the likelihood of detection may well decrease due to fading. However, when considering a given overall signal to noise ratio, i.e. computed over all the carriers, the detection probabilities over fading and non-fading channels are quite similar, as shown in Section 5.7, while the false alarm probabilities are the same. We can conclude that this particular design decision manages to overcome most of the difficulties that fading introduces in tag detection.

Constructing a Tag Detector

1. It is worth mentioning here a significant difference between the main purpose of a tag receiver and the purpose of a communication system receiver. While a communication system receiver is meant to accurately distinguish between a number of hypotheses corresponding to signal transmissions under the assumption that an actual transmission has occurred, the tag receiver listens for the most part to noise and background chatter. The main task of a tag receiver is therefore to detect, with sufficient confidence, a tag transmission when one occurs and, if possible, to correctly identify the transmitted tag. Tag detection is therefore a detection problem more than a communication problem and the design of the detector reflects this fact. The probabilities of false alarm allowed in the case of tags are well under the

typical probabilities of misclassification allowed in a communication system, since the occurrence of tag transmissions are assumed a priori to be rather rare events. False alarms weigh in more heavily when compared to the total number of detected tags. Due to the fact that tag detection is essentially a detection theory problem, we choose the detection metric (probability of false alarm versus the probability of detection) to be the main measure of tags performance. The secondary metric considered will be the probability of misclassification of a transmitted tag. The experimental section will reveal that this probability is negligible due to the high threshold required for a positive tag detection, even when using a rather large number of codewords.

2. **Detecting Patterns.** In general, tag observations occur over short intervals of time and channel conditions change too frequently for the receiver to obtain and update an accurate noise and interference power estimate. The only assumption made in the following is that the spectral envelope of the noise and interference signals is flat, an assumption that can be justified in the case of data networks using OFDM-based encoding. We design therefore our modulation scheme and our receiver to use as a detection indication not the sheer amount of power received but rather the concentration of the received power into pre-determined frequency bins. The receiver detects a transmission event whenever the concentration of the received power (the ratio of the power received in the designed frequency bins to the overall received power) exceeds a certain threshold. Therefore, the receiver searches not just for the presence of a signal but for a certain spectral shape. The fact that a power ratio measurement is used as a detection metric guarantees an universal receiver in a wider sense: the probability of detection for a threshold value chosen as the receiver parameter will only increase with increasing SINR. A standard detector is denoted as universal when a similar guarantee exists in terms of the SNR.

3. **Tags and FSK.** The attentive reader might have noticed that a simpler encoding scheme might have provided a similar detection/false alarm performance trade-off without the use of an algebraic code. Frequency shift keying simply concentrates the available power into the frequency space corresponding to one of the available carriers, thus offering similar received power characteristics. However, FSK has a large power spectral density, due to the fact that all the transmitted power is effectively concentrated in one point of the frequency spectrum, which makes it undesirable in a network environment, where we would like to guarantee a certain flat envelope for the frequency spectrum of our transmission, with a fast decay outside the data band. Our choice of modulation limits the transmitted power at any given frequency, resulting thus in a flat spectrum, similar to the one corresponding to OFDM data transmissions. The experiments in Section 5.6 verify that the typical interference effects of tags on competing data transmissions are not worse than the interference effects caused by normal packet data transmissions sent at a similar overall power level, which would not be the case if tags were modulated using FSK.

4. **Choosing the number of active carriers.** In Section 5.4 it was mentioned that tags are a specific form of multiple frequency shift keying (MFSK), a modulation which makes use of multiple noncoherent carriers transmitted simultaneously as a single symbol. A key parameter of MFSK is the number of transmitted carriers. Having chosen the MFSK modulation and the power ratio detector as the basic building blocks of our system, we must find next a value of this parameter that offers reasonable detection performance while also allowing for the construction of a large set of tags. As a definition of performance, we seek to minimize the probability of false alarms while preserving a certain probability of detection. In Section 5.7 this measure of detection performance is evaluated, as a function of the number of carriers transmitted, in the case of a detector that is searching for a single tag. It is revealed that, for a detector operating at a target SNR of 0 dB, the optimal allocation of power uses a bit less than half of the available carriers. Therefore, the decision to use exactly half of the carriers in the modulation does

not impact the detector's performance, while the same decision eases the use of an algebraic code.

Encoding

1. **Families of Codewords.** The decision to use exactly half of the carriers in the construction of tags is motivated by the details of constructing a tag family. In particular, we look at the relationship between tags and the binary codes used in their construction, namely codes over the same number of bits as the number of thick carriers. The simple binomial expansion indicates that the binary strings whose weight is half this number of bits are most numerous. We expect that for small Hamming distances the codes of this particular fixed weight would have more members than codes constructed using any set of binary strings of a different fixed weight. Our construction, which converts codeword families constructed over a number of bits equaling half the number of thick carriers into families of codewords of fixed weight over twice that length, is almost optimal, as it can be checked using tables of optimal known codes of fixed weight. Moreover, this construction allows for the use of well-known algebraic codes in tag construction and therefore leaves open the possibility of developing fast algorithms for identifying the most likely codeword in the case of large codeword families, in a manner similar to the identification, in communication systems, of the most likely transmitted codeword based on receiver soft symbol values.

2. **Construction Procedure.** Based on this description of the intricacies of tag construction we are now able to give a general procedure for constructing a set of tags that meet a desired set of performance criteria. Firstly, a codeword family that guarantees a low enough probability of misclassification at the target SNR while offering a sufficient number of different codewords must be selected. The probability of misclassification is computed in this step using Monte Carlo simulation, as presented in Section 5.7. Secondly, again using Monte Carlo simulation, as presented in Section 5.6, the tag designer must determine the probabilities of false alarm for the chosen tag family. The probability of detection can be more readily

computed since it does not depend on the choice of codeword family but only on the number of carriers that a codeword uses. This process must be iterated until a suitable trade-off between the number of codewords available, the probability of detection and the probability of false alarm is achieved. When satisfactory results cannot be achieved, the tag designer may increase the time length of tag through the use of a larger FFT window and a correspondingly increased number of thin carriers, adding more redundancy to the signals.

■   5.6   Evaluation

Experimental Setup

1. **(a) System Parameters** The experiments were conducted using two USRP boards [Ett], one transmitter and one receiver, in an indoor environment without line of sight and with multiple concrete walls between the sender and the receiver. The GNU Radio software suite was used for signal transmission and recording. The carrier frequency chosen was 2.45 GHz and the bandwidth of the system was set to 1 MHz. In order to obtain a linear channel suitable for OFDM transmission without distortion effects introduced by the transmitter mixer and the receiver, the signal has been oversampled by a factor of four on both transmitter and receiver sides.

   The data transmission part of the system uses the Schmidl and Cox algorithm [SC97] for packet detection and initial CFO estimation and the Tufvesson [TEF99b] algorithm for block boundary detection. Channel gain is measured using the preamble, while phase response is tracked using four pilot carriers and a linear phase interpolator. The symbol constellation used was 16-QAM. Packet payload was encoded using a rate $1/2$ punctured Trellis code. The resulting link speed is 2.4Mbps.

(a) Experimental results in the presence of channel noise.

(b) Experimental results in the presence of data-like interference.

(c) Experimental results in the presence of tag-like interference.

(d) Average time between false alarms at different thresholds.
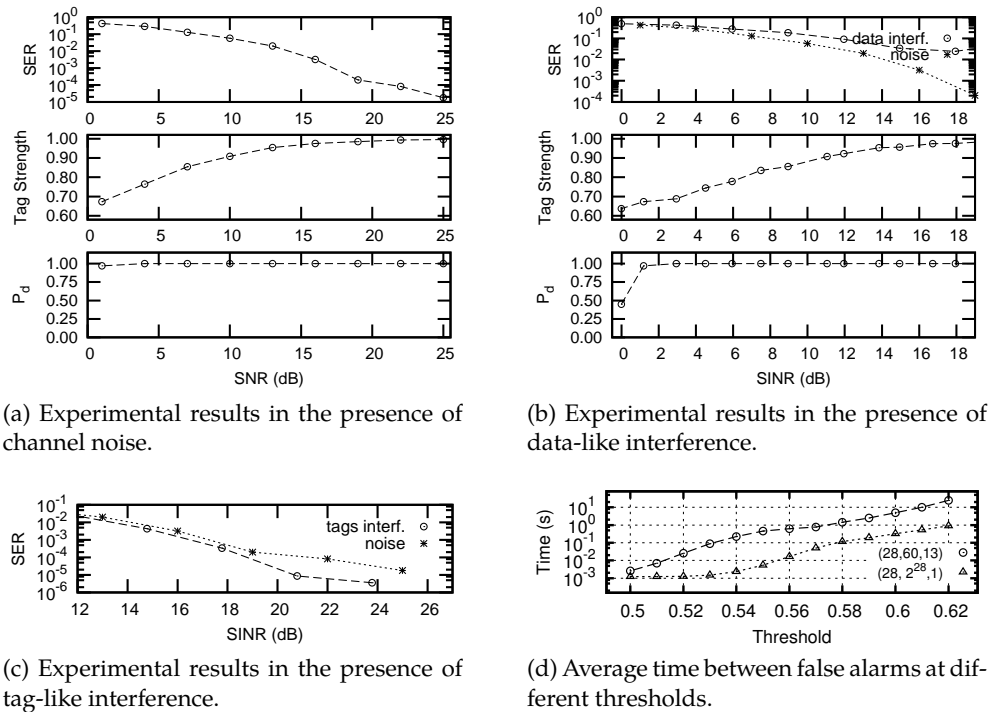
Figure 5.5: **Experimental Results.** The probabilities of detection and false alarm in different interference scenarios.

While the design of an efficient OFDM transmission and reception system is not the focus of this presentation, having a viable system was a precondition for showing that any interference effects produced by our set of header tags are comparable to interference effects caused by data transmissions and by environmental noise.

2. **(b) Experiment Description.** We performed three series of experiments intended to evaluate the impact of decreasing the signal to noise ratio on the effectiveness of tag spotting, the impact of rising interference power on tag spotting and the disruption caused to data transmission by interference in the form of tags. In order to determine the likelihood of false alarms, we have conducted a further series of experiments using half minute-long recorded signal sequences containing ambient radio noise pertaining to standard 802.11b/g transmissions in an office building occupied by numerous wireless networks in order to measure the detector's robustness to different kinds of radio interference. We have also evaluated through simulations the likelihood of misclassifications.

3. **(c) Metrics**

Tag Strength is the quantity defined in Equation (5.3), the primary metric for deciding whether a tag observation will be recorded. It is a measure of the ratio of power contained in the frequency bins allotted to a given tag and the total received power. In order for a tag observation to be recorded one of the necessary conditions is that the tag strength must exceed a threshold value γ. In all experiments presented $\gamma = 0.62$ was used. This choice of threshold accomplishes two goals: it is high enough to correspond to a low rate of false alarms, as verified through the experimental results presented in the current section (see Figure 5.5d) and it is low enough to allow detection at the target SNR values. Using the naïve assumption that noise (and interference) power contribute equally to the power levels detected in the different frequency bins, the SNR value that corresponds to this threshold can be derived to be about 0.4 dB.

Symbol Error Rate (SER) is measured for the payload of all correctly identified packets, that is, packets for which the packet detection, block boundary start estimation and CFO estimation succeed. It is the primary metric for estimating the effects of various noise and interference levels on data transmission. This metric was considered more fundamental than the bit error rate(BER), which is heavily dependent on the type of coding employed, a system design parameter that varies largely in current designs.

Probability of Detection (P_d) is defined as the probability that a header tag will be correctly detected and identified at different SNR and SINR levels. It is the primary metric for the success of tag spotting.

Probability of False Alarm (P_f) is defined as the likelihood that, in any given detection interval, noise and interference will cause a spurious tag detection and identification in the absence of a tag transmission.

Probability of Misclassification (P_m) is the likelihood of incorrect tag identification in the presence of a tag transmission.

4. **(d) Practical Considerations.** The tag detector presented in Section 5.4 is unable to compensate for frequency variations between the sender and the receiver. The structure of tags and the detection method described makes it possible to tolerate frequency offsets of up to two thin carriers, or about 4 kHz, without perceivable performance impacts. We have found that the clock jitters of the USRP radios are well within this limit, however different radios have initial frequency offsets of up to 200 kHz, necessitating a supplementary calibration step before each experiment. We assume the effects of Doppler spread to be minimal, i.e. a near-static scenario.

In a practical scenario the clock components could be replaced with more accurately designed/packed counterparts, and therefore we conclude that constructing self-sufficient tag detectors is possible.

Experimental Results

1. **(a) Impact of Noise.** The first series of experiments tries to quantify the range effectiveness of tag spotting in the presence of different levels of noise, in an interference-free environment.

The transmitter was configured to send sequences of 100 packets with random header tags. On the receiver side the transmitted sequence was decoded and the sequence of detected tags was compared to the original transmitted sequence, in order to obtain an estimate of the detection probability P_d. The decoded symbol payload of received packets was compared with the known symbol payload on the transmitter side in order to estimate the SER . The transmission's SNR was estimated for each detected packet using a low-pass filter-based average power estimator. The power level of the transmitter was varied between levels spaced 3 dB apart, resulting in different channel SNR values.

Figure 5.5a illustrates the results obtained. The upper pane shows the Symbol Error Rate (SER) for the payload as a function of the Signal to Noise ratio (SNR). The curve is typical for a receiver employing 16-QAM modulation, however the receiver appears to exhibit an error floor at the higher SNR values measured. At

SNR values of 20-25 dB, the system can sustain data transmission, when using a typical error-correcting code. This curve serves as a reference for the next experiments, in which noise-based disruptions will be replaced with data-like interference and tag-like interference.

The middle pane shows the Tag Strength as a function of the SNR. The curve decreases steadily as the SNR decreases, reaching the threshold value γ around 0 dB.

Finally, the lower pane shows the probability of detection as a function of the SNR. It can be seen that the probability of detection is close to one over the entire range considered.
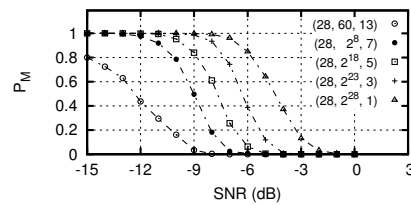
Figure 5.6: **Probability of miscalssification.** Probability of tag misclassification at different SNR levels.

2. **(b) Impact of Interference.** Figure 5.5b present the results of the same experiment in the presence of a second source transmitting an uninterrupted stream of payload-like data. As before the upper pane plots the Symbol Error Rate, the middle pane the Tag Strength, and the lower pane the probability of detection, all as a function of the SNR. The SER has a slightly different behavior in this case, due to the presence of a different type of interference, as can be seen when comparing the SER curve in the presence of data interference with the SER curve in the presence of just noise. The other quantities of interest, tag strength and the probability of detection P_d remain essentially unchanged. The probability of detection climbs a steep curve and quickly settles close to one. We conclude that the tag detector acts almost identically in the presence of pure noise or noise combined with temporary interference.

3. **(c) Impact of Tag Interference on Data.** Figure 5.5c presents the effect of tags on data transmissions. The SER curve is very close to the SER curve of Figure 5.5a, demonstrating that interference from tags does not increase the error likelihood beyond the error likelihood in the presence of comparable levels of noise.

4. **(d) Likelihood of False Alarm.** Figure 5.5d presents the dependence of the average time in-between false alarms on the threshold γ, when analyzing recordings of ambient WiFi traffic. Carrier sense has been disabled in this experiment and every input detection interval is analyzed. These results support our choice of detection threshold, since false alarms occur at a rate of less than once every 20 seconds.

5. **(e) Likelihood of Misclassification.** Figure 5.6 presents the dependence of the probability of misclassification on the receive SNR when tags are constructed using either the code mentioned in Section 5.4, a few extended BCH codes with smaller minimal distance or an unencoded modulation. No lower tag strength threshold was employed. This plot reveals that, for all these schemes, misclassification does not occur at the targeted signal levels. More details on the significance of this result will be given in the next section.

6. **(f) Choice of algebraic code.** The choice of algebraic code affects two quantities of interest, P_m and P_f. It has already been noted that, for a large class of codes, the probability of misclassification is negligible at the targeted signal levels. On the other hand, the probability of false alarm will increase as the number of code-words increases, as illustrated in Figure 5.5d, due to the presence of supplemental hypotheses. It results that there is a trade-off between the number of bits of information available per tag and the desired rate of false alarms.

Tag Range

The possible use of tags in wireless multi-hop networks prompts us to calculate the increase in range, expressed in distance terms, that transmission through tags brings over regular data transmission. It is well-known that power decay exponents are highly dependent on the environment. Measurements of signal propagation in the 2.4 GHz

(a) one-tag (wideband model)

(b) the $(28, 13, 60)$ code (wide-band model).

(c) the $(28, 1, 2^{28})$ code (wide-band model).

(d) one-tag (narrowband model)

(e) the $(28, 13, 60)$ code (narrow-band model).

(f) the $(28, 1, 2^{28})$ code (narrow-band model).
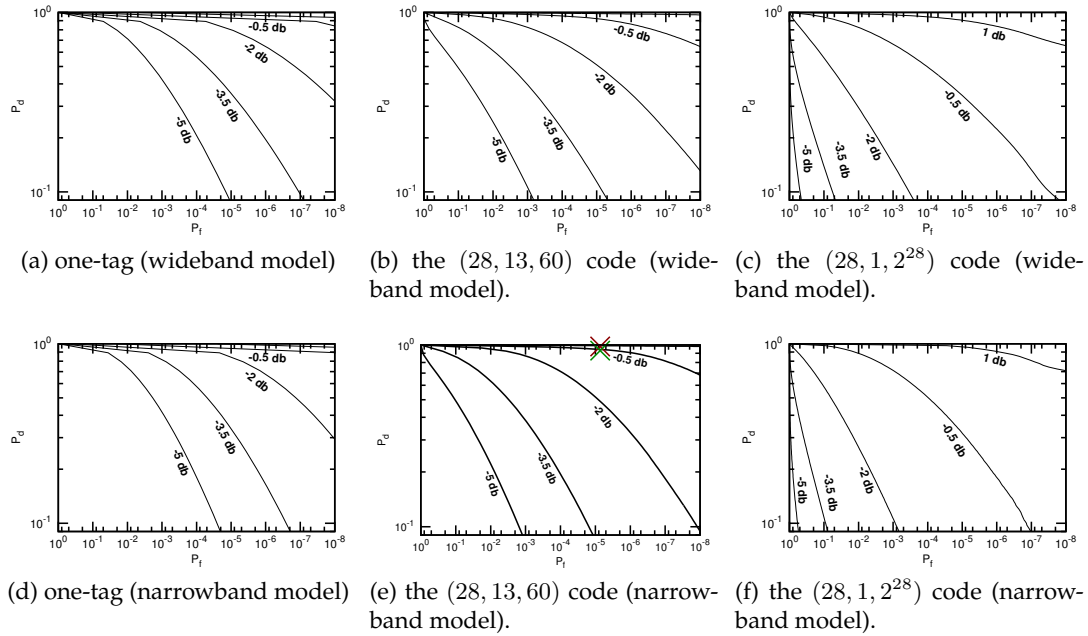
Figure 5.7: **Detection curves.** P_d and P_f for different choices of code, propagation model and detection threshold.

band described in [PO08] reveal a dependence of received power on distance of the form $p \propto \frac{1}{r^d}$ where the exponent d varies from 3 for line of sight propagation to about 6 for non-line of sight propagation. Regular data transmissions using the 16-QAM modulation present in our system necessitate a SNR value of about 20 dB while tags are detectable at a SNR value of 0 dB. It results that the ratio of the range of tag communication to the range of the data transmissions can be, depending on the power decay coefficient, anywhere between 2.15 and 4.65.

The carrier sense threshold is usually set about 10-20 dB above the noise level [BM09] [VRW05], which substantiates our claim that Tag Spotting can communicate information beyond the carrier sense range.

■   5.7   Performance Analysis

Section 5.4 has described the design of tags and introduced an universal detector[4] that does not require an estimate of the combined power of noise and background interference. However, the classic theoretical results on the detection and false alarm probability distributions are not readily applicable to our more complicated tag detector. In the following we will analyze, using rather conservative fading models, the performance achievable by a few particular tag families. At first we will consider the performance achievable when searching for a single tag signal, after which we will generalize to larger families of tags. The analysis will make use of two fading models, a narrowband fading model and a wideband fading model which assumes Rayleigh propagation. Due to the limited timespan of tags, these two models describe short-term fading effects only.

Single Tag. Let us consider a tag t. In what follows, we will denote by c the number of two-carrier groups used in the tag's construction. Assume that, for each thin carrier, the receiver noise has power n and its distribution can be modeled by a complex Gaussian random variable $\mathcal{N}(0, n)$. We further assume that for the active thin carriers the average signal power in the frequency band corresponding to any carrier is p. Under the assumption of narrowband transmission and without including the receiver noise contribution, the received signal obtained after demodulating one of the carriers can be modeled as a circularly uniform complex variable of constant amplitude $\sqrt{p}$, while in the case of wideband transmission the signal is modeled by a complex Gaussian random variable $\mathcal{N}(0, p)$. Assume that each thick carrier is composed of α thin carriers, out of which β are active. In the case of the system presented in the previous section $\alpha = 8$ and $\beta = 4$. Let P_t denote the set of thin carriers activated when tag t is transmitted, Q_t represent the thin carriers that, while they are not activated, belong to active thick carriers and R_t the thin carriers that belong to unactivated thick carriers. In accordance to the definition given when introducing the tag detector, we denote through $C_t = P_t \cup Q_t$ the set of all thin carriers belonging to activated thick carriers, regardless of whether they

---

[4]a detector for which, for any chosen probability of error P_f, the corresponding probability of detection P_d can only increase when the SINR is increased. A universal detector is particularly suited to our purposes, given the unpredictable nature of interference power.

are active or not. Let r_f be the complex values obtained after computing a fast Fourier transform of the real and complex components of a sampled tag signal, i.e. the received signal values corresponding to the various thin carriers. Let $|| \cdot ||$ denote the l^2 norm.

Under the assumption of independent Rayleigh fading, it results that for each active thin carrier the amplitude of the received signal is distributed according to a complex Gaussian random variable $\mathcal{N}(0, p + n)$. Choosing a threshold value γ for the quantity defined in Equation 5.3, we can write the probability of detection as:

$$P_d = P \left(\frac{\sum_{f \in C_t} ||r_f||^2}{\sum_{f \notin C_t} ||r_f||^2} > \frac{\gamma}{1 - \gamma} \right) \tag{5.4}$$

or

$$P_d = P \left(\frac{\sum_{f \in P_t} ||r_f||^2 + \sum_{f \in Q_t} ||r_f||^2}{\sum_{f \in R_t} ||r_f||^2} > \frac{\gamma}{1 - \gamma} \right) \tag{5.5}$$

It results from the previous paragraph and due to the independence of the circular random variables considered that the sums in the above equations can be written using the Chi-Square distribution with d components, denoted through χ_d^2. Namely $\frac{1}{p+n} \sum_{f \in P_t} ||r_f||^2 \sim \chi_{2\beta c}^2$, $\frac{1}{n} \sum_{f \in Q_t} ||r_f||^2 \sim \chi_{2(\alpha-\beta)c}^2$ and $\frac{1}{n} \sum_{f \in R_t} ||r_f||^2 \sim \chi_{2\alpha c}^2$.

In the case of narrowband fading, the amplitude of the received signal for each active carrier is constant. Therefore the first of these sums can be written using the noncentral Chi-Square distribution with parameter $\lambda = 2\beta c \frac{p}{n}$. We write $\frac{1}{n} \sum_{f \in P_t} ||r_f||^2 \sim \chi_{2\beta c}^2 \left(2\beta c \frac{p}{n} \right)$.

The probability of false alarm in the case of a single tag (and a single tested hypothesis) can be obtained by setting $p = 0$ in the above formulas. Therefore, writing the ratio of the two Chi-Squared random variables using a random variable f that follows the Fisher-Snedcor F-distribution [JSN95], $f \sim \mathcal{F}(2\alpha c, 2\alpha c)$,

$$P_f(t) = P \left(f > \frac{\gamma}{1 - \gamma} \right) \tag{5.6}$$

Figures 5.7a and 5.7d present the detector's behavior at different SNR values in the case of wideband and narrowband signals, respectively. [5]

Choosing the number of active carriers. Consider in the following a problem mentioned in Section 5.5, namely the optimal number of thick carriers q that should be activated during a tag transmission in order to maximize the performance of the detector. Let N denote the total number of thick carriers. In order to obtain closed-form results, we use a simplified model of tags in which we set $\alpha = \beta$, that is active thick carriers will use all thin subcarriers for transmission. Let f' be a random variable generated using the corresponding Fisher-Snedcor distribution, $f' \sim \mathcal{F}(2\alpha q, 2\alpha(N - q))$. Under this assumption we can simplify the formulas for the probability of detection and false alarm in the wideband case to: $P'_d(t) = P\left(\left(1 + \frac{p}{n}\right) f' > \frac{\gamma}{1-\gamma}\right)$ and $P'_f(t) = P\left(f' > \frac{\gamma}{1-\gamma}\right)$

It results that the probability distribution of the receiver response corresponding to detections is just a scaled version of the probability distribution corresponding to false alarms. We introduce a new performance measure in order to characterize the performance change due to the choice of q. Let γ_0 be the value of γ for which $P'_d = \frac{1}{2}$, at a SNR value of 0 dB. Figure 5.8 plots the behavior of P'_f, for a detector with a threshold γ_0, for different values of q, when $N = 56$. The quantity plotted represents the value of the tail of the probability of false alarm in the typical detection region.

Families of Tags. The next point in our analysis will be considering the situation in which the detector searches for multiple hypotheses. Assume therefore that the tag t is a member of a family of tags T, as described in Section 5.4. It can be readily observed that both in the narrowband and wideband cases the probabilities of detection remain unchanged. The probability of false alarm can be rewritten as:

$$P_f(T) = P\left(\max_{t \in T}\left(\frac{\sum_{f \in C_t} ||r_f||^2}{\sum_{f \notin C_t} ||r_f||^2}\right) > \frac{\gamma}{1 - \gamma}\right)$$

where $||r_f||^2 \sim \chi_2^2$.

---

[5]The SNR figures are computed using the power and noise figures for a thick carrier, that is $SNR = \frac{\beta p}{\alpha n}$.
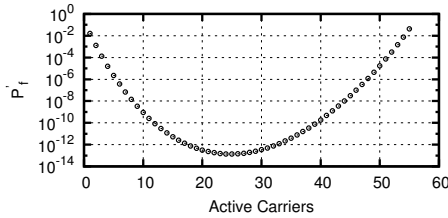
Figure 5.8: **False Alarms.** The probability of typical false alarms as a function of the number of active carriers. The total number of carriers used in tag construction is 56.

Figures 5.7b and 5.7e present the detector's behavior, evaluated through Monte Carlo simulation, at different SNRs in the case of wideband and narrowband signals, respectively for the $(28, 13, 60)$ code mentioned in Section 5.4. In particular, Figure 5.7e compares the theoretical predictions with the experimental results presented in Section 5.6. The green cross at the top of the figure presents the experimentally measured detection probability for tags in the presence of background noise only, at an SNR value of 1 dB, as shown in in Figure 5.5a, while the red cross presents the theoretical value at the same SNR. The experimental data and the theoretical curve, which indicates a probability of detection nearing one, are in agreement. Figures 5.7c and 5.7f illustrate the same detection curves in the case of the simple $(28, 1, 2^{28})$ code.

The power levels on the carriers are being summed up in the hypotheses in different ways. Since the variables that are summed up are chosen from the same set, the probability that the maximum present in the function will cross any chosen threshold γ' is significantly lower than what the sum bound on the individual probabilities of errors associated with the different codewords would predict.

The simpler quantity $\max_{t \in T} \sum_{f \in C_t} ||r_f||^2$ can be bounded using an initial symmetrization step [LT06] and deriving, using generic chaining [Tal96], a Dudley-like inequality [Dud67] on the probability that the maximum exceeds any given threshold. The resulting bound limits the increase of the necessary threshold, for any fixed probability of false alarm, to a quantity of the form $O(\log(N))$ where N is the number of codewords used. For reasons of space we have not included the derivation of the bound.

For the code construction using the two-carrier groups presented in Section 5.4, a simple upper bound on the probability of false alarm can be derived by considering

the case of the simplest $(28, 1, 2^{28})$ code, which has the largest probability of false alarm of all possible codes since it includes all possible codewords . Consider a set of pairs of Chi-Square distributed random variables $(x_{i,1}, x_{i,2})$ and let x_i^M and x_i^m represent the maximum and minimum, respectively, in each pair. The probability of false alarm can be thus written, for the afore mentioned code, as:

$$P_f(T) = P\left(\frac{\sum_{\forall i} x_i^M}{\sum_{\forall i} x_i^m} > \frac{\gamma}{1 - \gamma}\right)$$

The formula above has been used in order to derive the detection curves for the code mentioned, which are presented in Figures 5.7c and 5.7f.

∎ 5.8 Applications

In this section we present the use of Tag Spotting in two applications aimed towards obtaining a fair resource allocation in multi-hop wireless networks. We would like to emphasize that attaining a fair distribution of resources when using a wireless medium is, in our opinion, a neighborhood-centric problem. We explore the structure, the granularity and the rate of information that hosts within a neighborhood should exchange in order to solve the fairness problem

Congestion Control

In a wireless setting, congestion is not always primarily experienced by the flow that causes it and a neighborhood-wide signaling mechanism becomes necessary.

With this in mind, we extend WCP [RJJ⁺08], a recent AIMD-based scheme from the congestion control literature, by using Tag Spotting for communicating congestion notifications, and we assess the achieved performance of both the original WCP protocol and its extended derivative through simulations. In the original WCP, there are two types of information broadcasted by each node in data and ack packets: a congestion bit that indicates congestion events to its neighbors and the maximum of the round trip times (RTTs) of flows traversing it, a metric used in achieving a max-min fair allocation. This maximum RTT is then used to pace the rate increases of the AIMD controllers which set
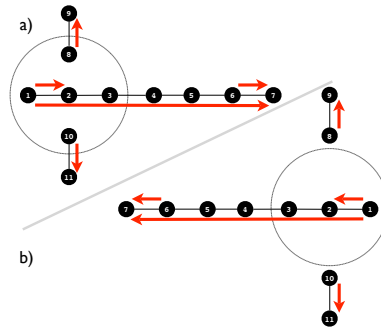
Figure 5.9: **Chain-cross topology.** All competing flows are separated by at most one transmission range (upper) and with some competing flows separated by more than the transmission range (lower).

the rates of the flows traversing the neighborhood. Both loss rates and delays experienced by competing flows passing through a congested neighborhood may vary widely, significantly more than in wired networks. This causes the senders' AIMD controllers to increase their rates at significantly different paces following a congestion event, unless a common loop duration is used. For more details on WCP, the interested reader is referred to [RJJ$^+$08].

We have simulated the performance of WCP using the Qualnet network simulator [Sca]. We have extended Qualnet's physical layer simulation in order to also handle the likelihood of tag detection using the detection probabilities measured in Section 5.6. The content of tags is composed of one congestion bit and a field that encodes, on a logarithmic scale with base $\sqrt[3]{2}$, the value, in milliseconds, the longest RTT of all flows traversing the tag emitter. We call the tag-based implementation of WCP, WCP-Tags. For the original WCP we have used a broadcasting mechanism that shares the same information as WCP-Tags, however the limit SNR for broadcast detection has been set at the same value at which successful payload data decoding occurs, since the original WCP broadcasts are inserted into the payload of data packets. All hosts use the regular 802.11 MAC for ad-hoc networks with default simulator values, with the only modification that the number of allowed MAC layer retransmissions has been doubled from its default value in order to decrease the rate of packet drops and increase the likelihood of tag reception.

(a) for the topology in Figure 5.9a. (b) for the topology in Figure 5.9b. (c) under the modified scheduling policy for the topology in Figure 5.9b.
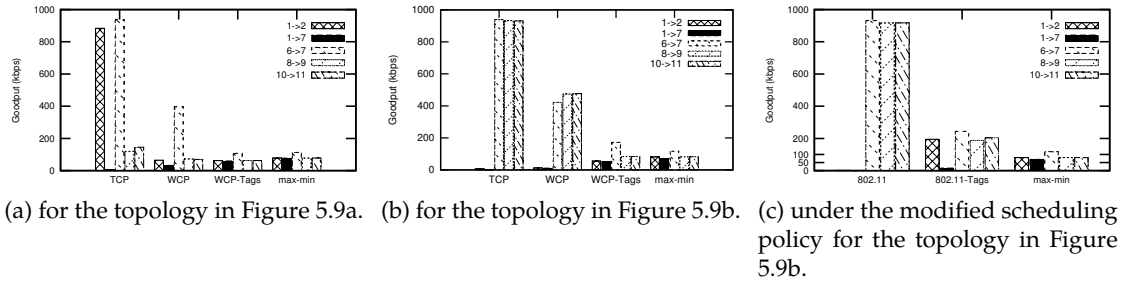
Figure 5.10: **Goodput.** Goodput results for the topology in Figure 5.9a

Figure 5.9a illustrates a textbook configuration for evaluating congestion control protocols in wireless environments. The two short flows on the outside of the central chain of nodes are within the transmission range of node 2, and we can therefore expect that the two variants of WCP will have similar performance. Figure 5.10a illustrates the rates obtained by TCP, original WCP ("WCP") and the tag-based implementation of WCP ("WCP-Tags"). It can be readily observed that, while TCP leads to starvation of the central flow, both WCP and WCP-Tags manage a fairer rate allocation.

Discussing the results of these experiments requires taking into account, above all, the fairness achieved and secondly the throughput. It is well known that supporting a long flow in a wireless multi-hop network is possible only at rates significantly lower than the maximal link speed [LBDC+01]. Any increase in the rate of the long flow in the figure will involve a drastic reduction in the rates of the other, shorter flows. To make this point more precise, we compute using brute force simulations and the theoretical framework in [JP09] the max-min rate allocation for these flows and compare it to the other allocations. It is evident that both the original WCP and WCP-Tags yield rates which are close to the max-min optimal rate allocation.

Figure 5.9b illustrates a variation of the previous topology in which the original WCP cannot effectively signal congestion between the involved hosts, due to the fact that some hosts are within the interference range but outside the data transmission range of each other. In particular, under the original WCP node 2 cannot inform nodes 8 and 10 that it is congested and the long flow is almost starved, similarly to what happens under TCP. In contrast, WCP-Tags does not starve the long flow as nodes 8 and 10 reduce their rates once they receive notifications through tags that node 2 is congested. The rate allocations

achieved by the three protocols as well as the max-min rate allocation for this topology are illustrated in Figure 5.10b. As before, WCP-Tags yields rates which are close to the max-min optimal allocation.

Scheduling

A large class of scheduling algorithms for wireless multi-hop networks are centered on ideas such as queue equalization using backpressure [Sto05], broadcasting local congestion indications [XGQS03] or creating a computationally tractable approximation of an optimal schedule [LS04] [JPPQ03]. Some of the benefits of schemes that use local communication will be illustrated in the current section through the evaluation of a rather simple scheduling scheme, designed as an extension to the ad-hoc mode of the the 802.11 MAC. The simple mechanism presented here targets some of the unfairness effects introduced by the 802.11 MAC which may lead to flow starvation.

Consider again the topology illustrated in Figure 5.9b. The results of Section 5.8 have already shown that using a standard 802.11 MAC in conjunction with TCP drives the longest flow in this topology into starvation. Our solution preserves TCP as the transport protocol but seeks to relieve such severely disadvantaged flows by enhancing the scheduling algorithm. The key to achieving this goal is an exchange of tags conveying a meaningful measure of starvation, namely the average delay of the packets currently enqueued for transmission.

The hosts observe all detected tags and decide that a host in their neighborhood is starved for medium access whenever they receive a tag conveying an average queueing time which is at least 32 times larger than their own average queueing time. In this case the tag receiver will enter a silence period of 15 milliseconds, allowing the starved host to gain access to the channel and transmit its packets. These numbers are not particularly optimized since our focus is to showcase the benefits of using tags in scheduling with a very simple addition to 802.11, rather than to provide a fully optimized and tested solution. All other medium access activity proceeds according to the normal 802.11 specification. We call this scheme 802.11-Tags. The only other MAC-layer modification

applied to both vanilla 802.11 and 802.11-Tags consists in doubling the number of MAC-layer retries performed in case of collision, just as in the previous section.

Interpreting the results in Figure 5.10c requires looking beyond the performance of individual links. In order for the long flow to achieve a transmission rate on the order of tens of kilobytes, all other flows must lower their rates far below the hundreds of kilobytes available on individual links. As shown in the previous section, a fair distribution of rates is associated with a drastic decrease of overall throughput. As expected, this simple MAC layer modification cannot achieve a max-min fair distribution of rates when used in conjunction with a standard AIMD rate controller like TCP at the transport layer.

As the results in Figure 5.10c show, the improved scheduler alleviates flow starvation. The shortest of the two starved flows in the figure reaches a rate similar to the ones of the other three short flows. The long flow in the figure increases its achieved rate from a level that cannot prevent connection timeouts and interruptions to a sustained rate of about 15 kbps.

■   5.9   Conclusion

This paper has proposed a mechanism for sharing control information in wireless networks, able to function in low SNR conditions and without introducing new interference constraints. It evaluated its performance through a theoretical analysis as well as experiments realized using software radios. It was found that the communication scheme presented is effective at SNR values as low as 0 dB, effectively covering the carrier sense range of a wireless host. Two algorithm implementations that make use of this scheme have been evaluated through simulations, confirming its applicability in protocol design.

Bibliography

[80212]   IEEE Draft Standard for IT - Telecommunications and Information Exchange Between Systems - LAN/MAN - Specific Requirements - Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications - Amd 4: Enhancements for Very High Throughput for operation in bands below 6GHz. *IEEE P802.11ac/D3.0, June 2012*, pages 1 –385, 11 2012.

[AASK10]   Ehsan Aryafar, Narendra Anand, Theodoros Salonidis, and Edward W. Knightly. Design and experimental evaluation of multi-user beamforming in wireless LANs. In *ACM MobiCom*, Chicago, IL, 2010.

[ABD$^+$90] E. Anderson, Z. Bai, J. Dongarra, A. Greenbaum, A. McKenney, J. Du Croz, S. Hammerling, J. Demmel, C. Bischof, and D. Sorensen. Lapack: a portable linear algebra library for high-performance computers. In *Proceedings of the 1990 ACM/IEEE conference on Supercomputing*, Supercomputing '90, pages 2–11, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

[Ala98]   S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Sel. Areas Commun.*, 16(8):1451–1458, 1998.

[BM09]   Micah Z. Brodsky and Robert T. Morris. In defense of wireless carrier sense. In *SIGCOMM*, 2009.

[BRM$^+$12] Horia Vlad Balan, Ryan Rogalin, Antonios Michaloliakos, Konstantinos Psounis, and Giuseppe Caire. Achieving high data rates in a distributed mimo system. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, Mobicom '12, pages 41–52, New York, NY, USA, 2012. ACM.

[BRM$^+$13] H. V. Balan, R. Rogalin, A. Michaloliakos, K. Psounis, and G. Caire. Airsync: Enabling distributed multiuser mimo with full spatial multiplexing, 2013.

[BTC06]   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.

[Chi05]   Mung Chiang. Balancing transport and physical layers in wireless multi-hop networks: jointly optimal congestion control and power control. *IEEE JSAC*, 23(1):104–116, 2005.

[cis12]   Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011 - 2016. Technical report, Cisco VNI, Feb 2012.

[CJKR10]   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.

[CJS$^+$10]   Jung Il Choi, Mayank Jain, Kannan Srinivasan, Phil Levis, and Sachin Katti. Achieving single channel, full duplex wireless communication. In *IEEE MobiCom*, Chicago, IL, 2010.

[CMM$^+$08]   Ranveer Chandra, Ratul Mahajan, Thomas Moscibroda, Ramya Raghavendra, and Paramvir Bahl. A case for adapting channel width in wireless networks. In *SIGCOMM*, 2008.

[Cos83]   M. Costa. Writing on dirty paper (corresp.). *IEEE Trans. Inf. Theory*, 29(3):439–441, May 1983.

[CRB$^+$12]   Krishna Chintalapudi, Bozidar Radunovic, Vlad Balan, Michael Buettener, Srinivas Yerramalli, Vishnu Navda, and Ramachandran Ramjee. Wifi-nc: Wifi over narrow channels. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, NSDI'12, pages 4–4, Berkeley, CA, USA, 2012. USENIX Association.

[CS03]   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.

[CT91]   Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley-Interscience, New York, NY, USA, 1991.

[DDS11]   Melissa Duarte, Chris Dick, and Ashutosh Sabharwal. Experiment-driven characterization of full-duplex wireless systems. *CoRR*, abs/1107.1276, 2011.

[DKZSM05]   N. Dukkipati, M. Kobayashi, R. Zhang-Shen, and N. McKeown. Processor sharing flows in the internet. In *Thirteenth International Workshop on Quality of Service (IWQoS)*. Springer, 2005.

[DS05]   G. Dimic and N.D. Sidiropoulos. On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm. *IEEE Trans. Signal Process.*, 53(10):3857 – 3868, Oct. 2005.

[DSGS09a]   Aveek Dutta, Dola Saha, Dirk Grunwald, and Douglas Sicker. SMACK: a SMart ACKnowledgment scheme for broadcast messages in wireless networks. In *ACM SIGCOMM*, Barcelona, Spain, 2009.

[DSGS09b] Aveek Dutta, Dola Saha, Dirk Grunwald, and Douglas Sicker. SMACK: a SMart ACKnowledgment scheme for broadcast messages in wireless networks. In *SIGCOMM*, 2009.

[Dud67] R. M. Dudley. The sizes of compact subsets of hilbert space and continuity of gaussian processes. *Journal of Functional Analysis*, 1(3):290 – 330, 1967.

[ESZ05] 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.

[Ett] Ettus Research. USRP board. `http://www.ettus.com/`.

[FE91] Jr. Forney, G.D. and M.V. Eyuboglu. Combined equalization and coding using precoding. *Communications Magazine, IEEE*, 29(12):25 –34, dec. 1991.

[FG98] G.J. Foschini and M.J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, 6:311–335, 1998.

[GHMS09] S. Gupta, C. Hunter, P. Murphy, and A. Sabharwal. Warpnet: Clean slate research on deployed wireless networks. In *Mobihoc*, 2009.

[GK08] Shyamnath Gollakota and Dina Katabi. Zigzag decoding: combating hidden terminals in wireless networks. In *ACM SIGCOMM*, Seattle, WA, 2008.

[GPK09] Shyamnath Gollakota, Samuel David Perli, and Dina Katabi. Interference alignment and cancellation. In *ACM SIGCOMM*, Barcelona, Spain, 2009.

[GWJ10] Tiangao Gou, Chenwei Wang, and S.A. Jafar. Aiming perfectly in the dark - blind interference alignment through staggered antenna switching. In *IEEE GLOBECOM*, Dec. 2010.

[HAW08] Daniel Halperin, Thomas Anderson, and David Wetherall. Taking the sting out of carrier sense: interference cancellation for wireless LANs. In *MobiCom*, 2008.

[HD02] Zygmunt J. Haas and Jing Deng. Dual busy tone multiple access (dbtma) - a multiple access control scheme for ad hoc networks. *IEEE Transactions on Communications*, 2002.

[Jac88] V. Jacobson. Congestion avoidance and control. In *SIGCOMM*, 1988.

[JAWV08] J. Jose, A. Ashikhmint, P. Whiting, and S. Vishwanath. Scheduling and pre-conditioning in multi-user MIMO TDD systems. In *IEEE ICC*, Beijing, China, 2008.

[JAWV11] J. Jose, A. Ashikhmin, P. Whiting, and S. Vishwanath. Channel estimation and linear precoding in multiuser multiple-antenna TDD systems. *IEEE Trans. Veh. Technol.*, 60(5):2102 –2116, Jun. 2011.

[JLS09] Changhee Joo, Xiaojun Lin, and Ness B. Shroff. Understanding the capacity region of the greedy maximal scheduling algorithm in multihop wireless networks. *IEEE/ACM Transactions on Networking*, 17(4):1132–1145, 2009.

[JP09]   A. Jindal and K. Psounis. Characterizing the Achievable Rate Region of Wireless Multi-hop Networks with 802.11 Scheduling. *IEEE/ACM Transactions on Networking*, 2009.

[JPPQ03]   Kamal Jain, Jitendra Padhye, Venkata N. Padmanabhan, and Lili Qiu. Impact of interference on multi-hop wireless network performance. In *MobiCom*, 2003.

[JSN95]   Norman L. Johnson, Kotz Samuel, and Balakrishnan Nikhil. *Continuous Univariate Distributions, Vol. 2*. Wiley, New York, 1995.

[Kay89]   S. Kay. A fast and accurate single frequency estimator. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12):1987–1990, 1989.

[KC06]   M. Kobayashi and G. Caire. Joint beamforming and scheduling for a MIMO downlink with random arrivals. In *IEEE ISIT*, Seattle, WA, Jul. 2006.

[KHR02]   Dina Katabi, Mark Handley, and Charlie Rohrs. Congestion control for high bandwidth-delay product networks. In *SIGCOMM*, 2002.

[KMPS05]   V. S. Anil Kumar, Madhav V. Marathe, Srinivasan Parthasarathy, and Aravind Srinivasan. Algorithmic aspects of capacity in wireless networks. *SIGMETRICS Perform. Eval. Rev.*, 33(1):133–144, 2005.

[LBDC$^+$01] Jinyang Li, Charles Blake, Douglas S.J. De Couto, Hu Imm Lee, and Robert Morris. Capacity of ad hoc wireless networks. In *MobiCom*, 2001.

[LM02]   Cheng Luo and Muriel Medard. Frequency-shift keying for ultrawideband - achieving rates of the order of capacity. In *In 40th Annual Allerton Conference on Comm., Control, and Computing*, 2002.

[LS04]   Xiaojun Lin and N.B. Shroff. Joint rate control and scheduling in multihop wireless networks. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 2, pages 1484–1489 vol.2, 2004.

[LSW05]   Amos Lapidoth, Shlomo Shamai, and Michele A. Wigger. On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information. In *43rd Annual Allerton Conference*, Monticello, IL, 2005.

[LT06]   Michel Ledoux and Michel Talagrand. *Probability in Banach Spaces*. Springer, 2006.

[Mar10]   T.L. Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *Wireless Communications, IEEE Transactions on*, 9(11):3590 –3600, november 2010.

[Mol05]   Andreas Molisch. *Wireless Communications*. Wiley-IEEE Press, 2005.

[MW95]   Robert N. McDonough and A. D. Whalen. *Detection of Signals in Noise*. Academic Press, Inc., 1995.

[OJTL10]   A. Ozgur, R. Johari, D.N.C. Tse, and O. Leveque. Information-theoretic operating regimes of large wireless networks. *Information Theory, IEEE Transactions on*, 56(1):427–437, 2010.

[OLT13]   A. Ozgur, O. Leveque, and D. Tse. Spatial degrees of freedom of large distributed mimo systems and wireless ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 31(2):202–214, 2013.

[PBS11]   Jonathan Perry, Hari Balakrishnan, and Devavrat Shah. Rateless spinal codes. In *ACM HotNets*, Cambridge, Massachusetts, 2011.

[PO08]   T.K. Paul and T. Ogunfunmi. Wireless LAN Comes of Age: Understanding the IEEE 802.11n Amendment. *IEEE Circuits and Systems Magazine*, 8(1):28–54, 2008.

[PS07]   J.G. Proakis and M. Salehi. *Digital communications*. McGraw-Hill, New York, NY, 2007.

[RB74]   D. Rife and R.R. Boorstyn. Single tone parameter estimation from discrete-time observations. *Information Theory, IEEE Transactions on*, 20(5):591–598, 1974.

[RBMR63]   H.K Robin, D Bayley, T.L Murray, and J.D Ralphs. Multitone signalling system employing quenched resonators for use on noisy radio-telprinter circuits. *Proceedings of the IEE*, 110(9):1554–1568, 1963.

[RBP+]   Ryan Rogalin, Ozgun Bursalioglu, Haralabos Papadopoulos, Giuseppe Caire, Andreas Molisch, Antonios Michaloliakos, Horia Vlad Balan, and Konstantinos Psounis. Scalable Synchronization and Reciprocity Calibration for Distributed Multiuser MIMO.

[RHK10]   Hariharan Rahul, Haitham Hassanieh, and Dina Katabi. SourceSync: a distributed wireless architecture for exploiting sender diversity. In *ACM SIGCOMM*, New Delhi, India, 2010.

[Ric]   Rice University. Rice university warp project.

[RJJ+08]   Sumit Rangwala, Apoorva Jindal, Ki-Young Jang, Konstantinos Psounis, and Ramesh Govindan. Understanding congestion control in multi-hop wireless mesh networks. In *MobiCom*, 2008.

[RKK12]   Hariharan Shankar Rahul, Swarun Kumar, and Dina Katabi. Jmb: scaling wireless capacity with user demands. *SIGCOMM Comput. Commun. Rev.*, 42(4):235–246, August 2012.

[SC97]   T.M. Schmidl and D.C. Cox. Robust frequency and timing synchronization for OFDM. *IEEE Transactions on Communications*, 45(12):1613–1621, 1997.

[Sca]   Scalable Network Technologies. Qualnet simulator: http://www.qualnet.com/.

[Sho06]   A. Shokrollahi. Raptor codes. *IEEE Trans. Inf. Theory*, 52(6):2551 –2567, Jun. 2006.

[SOSL94]   M.K. Simon, J.K. Omura, R.A. Scholtz, and B.K. Levitt. *Spread spectrum communications handbook*. McGraw-Hill New York, 1994.

[SPSH04] Q.H. Spencer, C.B. Peel, A.L. Swindlehurst, and M. Haardt. An introduction to the multi-user MIMO downlink. *IEEE Commun. Mag.*, 2004.

[SRCN10] Souvik Sen, Romit Roy Choudhury, and Srihari Nelakuditi. Csma/cn: carrier sense multiple access with collision notification. In *MobiCom*, 2010.

[SS70] N. J. A. Sloane and J. J. Seidel. A New Family of Nonlinear Codes Obtained from Conference Matrices. *New York Academy Sciences Annals*, 175:363–365, 1970.

[Sto05] Alexander L. Stolyar. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Syst. Theory Appl.*, 50(4):401–457, 2005.

[SYA$^+$12] C. Shepard, H. Yu, N. Anand, L. Li, T Marzetta, R. Yang, and L. Zhong. Argos: Practical many-antenna base stations. *SIGCOMM Comput. Commun. Rev.*, August 2012.

[Tal96] Michel Talagrand. Majorizing Measures: The Generic Chaining. *The Annals of Probability*, 24(3):1049–1103, 1996.

[TE92] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Transactions on Automatic Control*, 37(12):1936–1948, 1992.

[TEF99a] F. Tufvesson, O. Edfors, and M. Faulkner. Time and frequency synchronization for OFDM using PN-sequence preambles. In *IEEE VTC*, 1999.

[TEF99b] F. Tufvesson, O. Edfors, and M. Faulkner. Time and frequency synchronization for OFDM using PN-sequence preambles. In *IEEE Vehicular Technology Conference, VTC 1999 - Fall*, volume 4, pages 2203–2207, 1999.

[Tel99] Emre Telatar. Capacity of multi-antenna gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999.

[TFZ$^+$10] Kun Tan, Ji Fang, Yuanyang Zhang, Shouyuan Chen, Lixin Shi, Jiansong Zhang, and Yongguang Zhang. Fine-grained channel access in wireless LAN. In *ACM SIGCOMM*, New Delhi, India, 2010.

[TJZS07] K. Tan, F. Jiang, Q. Zhang, and X. Shen. Congestion control in multihop wireless networks. *IEEE Transactions on Vehicular Technology*, 56(2):863, 2007.

[Tre85] Steve Tretter. Estimating the frequency of a noisy sinusoid by linear regression (corresp.). *Information Theory, IEEE Transactions on*, 31(6):832–835, 1985.

[Tre92] Harry L. Van Trees. *Detection, Estimation, and Modulation Theory: Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Krieger Publishing Co., Inc., Melbourne, FL, USA, 1992.

[TV05]   D. Tse and P. Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, New York, NY, 2005.

[Ver98]   Sergio Verdu. *Multiuser Detection*. Cambridge University Press, New York, NY, 1998.

[Vit95]   Andrew J. Viterbi. *CDMA: principles of spread spectrum communication*. Addison Wesley Longman Publishing Co., Inc., 1995.

[VRW05]   A. Vasan, R. Ramjee, and T. Woo. ECHOS-enhanced capacity 802.11 hotspots. In *INFOCOM*, 2005.

[VTGK09]   Sudarshan Vasudevan, Donald Towsley, Dennis Goeckel, and Ramin Khalili. Neighbor discovery in wireless networks and the coupon collector's problem. In *MobiCom*, 2009.

[VV09]   Chinmay S. Vaze and Mahesh K. Varanasi. The degrees of freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT. *CoRR*, abs/0909.5424, 2009.

[WFVH04]   C. Windpassinger, R.F.H. Fischer, T. Vencel, and J.B. Huber. Precoding in multiantenna and multiuser communications. *Wireless Communications, IEEE Transactions on*, 3(4):1305 – 1316, Jul. 2004.

[WJHR09]   Ajit Warrier, Sankararaman Janakiraman, Sangtae Ha, and Injong Rhee. DiffQ: Differential Backlog Congestion Control for Multi-hop Wireless Networks. In *INFOCOM*, 2009.

[WP05]   R. Clint Whaley and Antoine Petitet. Minimizing development and maintenance costs in supporting persistently optimized BLAS. *Software: Practice and Experience*, 35(2):101–121, February 2005. `http://www.cs.utsa.edu/~whaley/papers/spercw04.ps`.

[WPD01]   R. Clint Whaley, Antoine Petitet, and Jack J. Dongarra. Automated empirical optimization of software and the ATLAS project. *Parallel Computing*, 27(1–2):3–35, 2001. Also available as University of Tennessee LAPACK Working Note #147, UT-CS-00-448, 2000 (`www.netlib.org/lapack/lawns/lawn147.ps`).

[WSS06]   H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inf. Theory*, 52(9):3936 –3964, Sept. 2006.

[WTL+10]   Kaishun Wu, Haoyu Tan, Yunhuai Liu, Jin Zhang, Qian Zhang, and Lionel Ni. Side channel: bits over interference. In *MobiCom*, 2010.

[XGQS03]   Kaixin Xu, Mario Gerla, Lantao Qi, and Yantai Shu. Enhancing TCP fairness in ad hoc wireless networks using neighborhood red. In *MobiCom*, 2003.

[YG06]   Taesang 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.

[YKT03]   Z. Ye, S.V. Krishnamurthy, and S.K. Tripathi. A framework for reliable rout-
           ing in mobile ad hoc networks. In *INFOCOM*, 2003.

[YLF93]   Nathan Yee, Jean-Paul Linnartz, and Gerhard Fettweis.   Multi-carrier
           CDMA in indoor wireless radio networks, 1993.