Adaptation in Convolutionally-Coded MIMO-OFDM Wireless Systems through Supervised Learning and SNR Ordering

Robert C. Daniels, *Student Member, IEEE,*
Constantine M. Caramanis, *Member, IEEE,*
and Robert W. Heath, Jr. *Senior Member, IEEE*

Abstract

MIMO-OFDM (multiple-input multiple-output *and* orthogonal frequency division multiplexing) wireless systems use link adaptation to exploit the dynamic nature of wireless environments. Link adaptation maximizes throughput while maintaining a target reliability by adaptively selecting the modulation order and coding rate. Link adaptation is extremely challenging, however, due to the difficulty in predicting error rates in OFDM with binary convolutional codes, bit-interleaving, MIMO processing, and real channel impairments. This paper proposes a new machine learning framework that exploits past observations of error rate and the associated channel state information to predict the best modulation order and coding rate for new realizations of channel state without modeling the input-output relationship of the wireless transceiver. Our approach is enabled through our new error rate expression that is parametrized only by post-processing signal-to-noise ratios (SNR), ordered over subcarriers and spatial streams. Using ordered SNR we propose a low-dimension feature set that enables machine learning to increase the accuracy of link adaptation. An IEEE 802.11n simulation study validates the application of this machine learning framework in real channels and demonstrates the improved performance of SNR ordering as it compares to competing link-quality metrics.

Index Terms

Adaptive Modulation and Coding, Link Adaptation, MIMO-OFDM, Supervised Learning, SNR Ordering

The authors are with the Wireless Networking & Communications Group in the Department of Electrical and Computer Engineering at the University of Texas at Austin, 1 University Station C0803, Austin, TX 78712-0240 (email: {rdaniels,cmcaram,rheath}@ece.utexas.edu, phone: +1 512.232.2014, fax: +1 512.471.6512).

This work was supported by Freescale Semiconductor and the DARPA IT-MANET program, Grant W911NF-07-0028. The material in this paper, in part, was presented at the *IEEE Global Telecommunications Conference 2009 [1].*

I. INTRODUCTION

Advances in wireless communications, such as multiple-input, multiple-output (MIMO) systems and orthogonal frequency division multiplexing (OFDM), provide the infrastructure for high spectral efficiency and consequently high data rate in the physical layer. The data traffic in higher layers also requires that reliability constraints are met to reduce overhead, reduce latency, and maximize network throughput [2]–[4]. Current wireless systems, however, operate in dynamic wireless channels where the quality of the link depends on explicit wireless channel characteristics such as transmission range, obstacles in the wireless medium, and the interaction of the multiple paths from the transmitter to receiver [5]. To meet the competing demands of high data rate and high reliability, the physical layer is configured to maximize the data rate while satisfying the reliability constraints imposed by higher layers. Unfortunately, the physical layer configuration or *link adaptation* given the changing state of the wireless channel is difficult, and as a result the physical layer in current systems is rarely configured optimally [6]–[9].

In this paper, following the convention of IEEE 802.11n, IEEE 802.16e, and 3GPP-LTE, we consider link adaptation in the form of *adaptive modulation and coding* (AMC) where the symbol modulation order, error control coding rate, and spatial multiplexing order are assigned to maximize data rate under frame error rate (FER) reliability constraints imposed by higher layers [10]–[13]. To provide a better understanding of the relationship between MIMO-OFDM system performance and AMC, there has been considerable research activity recently [14]–[16]. Unfortunately, jointly modeling the effects of MIMO processing, OFDM modulation, convolutional-coding, and bit interleaving is challenging. Consequently, approximations have been made in both the performance of interleaving with convolutional codes [15], [17], [18] and the FER through pairwise error probability [14], [16]. Moreover, for wireless systems with radio-frequency (RF) and analog circuits that interface to the wireless medium, mitigating circumstances such as amplifier nonlinearity, transmission frequency instability, and non-Gaussian additive noise limit tractable mathematical modeling of the complete wireless system [19]. Hence, adaptation based solely on the analysis of mathematical models struggles to accurately account for all the factors that determine performance.

Therefore, for effective link adaptation a flexible framework is desired to enable AMC using as few assumptions on the mathematical model of the physical layer as possible such that approx-

imations do not have to be made (e.g. ideal interleaving, wireless channel distribution, etc.) and mitigating circumstances can be accounted for (e.g. amplifier nonlinearity, non-Gaussian additive noise, etc.). Machine learning, or "the programming of computers to optimize a performance criterion using example data" [20], facilitates this approach to decision making since machine learning considers the physical layer as nothing more or less than transfer between system state and data observation [21], [22]. Consequently, we propose a paradigm shift in AMC for MIMO-OFDM systems through a machine learning framework that provides a mapping between modulation order, coding rate, and spatial multiplexing order and the channel state.

Application of machine learning algorithms to AMC in MIMO-OFDM is nontrivial since the performance directly depends on properties of the example data from which it learns. In this paper we consider *supervised learning* where mappings between the channel state and the optimal AMC parameters are contained in the example data. Supervised learning captures the channel state through low-dimensional mappings onto a *feature space*. If the feature space dimension is too large, machine learning algorithms cannot identify accurate mappings without an exceedingly large number of channel observations. If the feature space does not accurately reflect the performance of the physical layer, machine learning algorithms cannot determine the relationship between channel state and system performance. Therefore, the effective application of machine learning algorithms for AMC will only be enabled if a low-dimension mapping from the channel state accurately parametrizes the frame error rate of MIMO-OFDM systems.

Naively, we might proceed by using established link-quality metrics to define our feature space including those based on subcarrier SNR variance [23], mutual information metrics [24], average exponential mappings of post-processing SNR [6]–[8], [25], average post-processing SNR [26], minimum post-processing SNR, minimum subcarrier capacity, average subcarrier capacity [27], and weighted mappings of several link-quality metrics [27]. These link-quality metrics have been limited to one or two dimensions to enable simplicity in look-up-table based AMC. Unfortunately, although we desire to keep dimensions as small as possible, past research has shown it is difficult to define a one- or two-dimensional link-quality metric that can accurately account for the frequency- and spatial-selective channels that MIMO-OFDM experiences [6], [8], [9], [28], [29]. This is also reiterated in Section V-C of this paper. Look-up-tables are also limited by the need to discretize the values of link-quality metrics used to create mappings to the frame error rate for a fixed convolutional coding rate, QAM modulation size, and spatial multiplexing order.

Supervised learning allows for simple AMC procedures based on example data with flexibility in the feature space for link-quality metrics with greater than two dimensions.

In this paper we propose a new multi-dimensional feature space through the post-processing SNR (after equalization), ordered over subcarriers and spatial streams to enable supervised learning for AMC. This feature space is motivated by a new expression for FER in MIMO-OFDM systems that is solely a function of ordered post-processing SNR. Moreover, because the ordering operation over post-processing SNR adds significant correlation between SNR values, we reduce the feature space to very low dimensions (≤ 4 dimension in IEEE 802.11n) and approximate the ordered SNR profile. Although we have discovered a connection between the FER and ordered post-processing SNR, there is no clear connection between ordered subcarrier SNR and the correct set of modulation order, coding rate, and spatial multiplexing order. Therefore, this paper proposes a non-parametric supervised learning algorithm, namely k-nearest neighbor, that harnesses example data with ordered post-processing SNR features and creates mappings to AMC parameters. IEEE 802.11n performance simulations in this paper show that k-nearest neighbor supervised learning with ordered post-processing SNR feature sets outperforms competing link-metrics and provides accurate AMC.

Prior Work: Applications of machine learning to AMC in wireless networks have been investigated in the past using supervised learning [30] and stochastic learning models [31], [32] that operate over single antenna channels where the performance is completely parametrized by a single SNR value. While adequate for the systems they model, prior work is difficult to extend since MIMO-OFDM systems experience increased complexity due to coding and interleaving over subcarriers, frequency-selective channels, and the need to choose the number of spatial streams along with modulation order and coding rate. Although the work in this paper is considered new, it has been partially treated in a conference paper by the same authors [1]. The conference paper, however, due to its length constraints, does not supply the full system model, the derivation of a new FER expression that inspires feature sets based on ordered post-processing SNR ordered, or the general subcarrier ordering feature set. Additionally the entire performance evaluation of IEEE 802.11n systems with our supervised learning framework is new in this paper. This includes a search over the entire space of ordered SNR feature sets up to four dimensions to optimize performance in channels of varying frequency selectivity and a performance comparison of alternative feature sets. Essentially, substantial pieces of the

machine learning framework, the motivation and development of the feature space, and the entire validation of the performance are all missing in the conference paper.

Outline: The remainder of this paper is organized as follows. Section II provides the model for the MIMO-OFDM system under AMC investigation. In Section III we discuss the FER of MIMO-OFDM based on system assumptions and construct an expression for FER (with details in Appendix A) based on post-processing SNR, ordered over subcarriers and spatial streams. In Section IV we design a classifier based on k-nearest neighbor to enable supervised learning for MIMO-OFDM AMC. Section IV also displays the design of a low-dimension feature space from ordered post-processing SNR values, advised by the FER expression in Section III. Using this supervised learning algorithm we complete a comprehensive investigation of AMC for IEEE 802.11n with several competing link-quality metrics in Section V. We conclude with some suggestions for future research directions in Section VI.

Notation: $\mathbf{A}^*$ denotes Hermitian transpose of matrix $\mathbf{A}$ and $[\mathbf{A}]_{a,b}$ is the matrix element in the ath row and bth column. We represent a sequence with index k from a to a' as $\{\cdot\}_{k=a}^{a'}$. Circularly symmetric complex Gaussian vectors with mean $\boldsymbol{\mu}$ and covariance $\mathbf{R}$ are indicated by $\sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{R})$.

II. System Model

This section defines the MIMO-OFDM system model, summarized in Fig. 1, which represents the communication procedure of wireless signals in a practical MIMO-OFDM wireless link such as IEEE 802.11n. Small modifications to this system model can be made to align with 3GPP-LTE and WiMax standards. This is discussed wherever appropriate. The symbols used throughout the system are shown in Table I. Each physical layer frame of the MIMO-OFDM transceiver is processed as follows[1]. The source bits for a single frame $\{b[n]\}_{n=0}^{KN_{\mathrm{O}}N-1}$ are delivered to the convolutional encoder in stage 1. In stage 2, the convolutionally encoded bits $\{c[n]\}_{n=0}^{KN_{\mathrm{O}}N/\mathcal{C}-1}$ with coding rate $\mathcal{C}$ are sent to the spatial bit parser. Each spatial branch $a \in \{1, 2, \ldots, N_s\}$ of the transmitter receives an equivalent number of bits (we assume each spatial branch has the same modulation order and coding rate). The spatial parsing operation not only assigns bits

[1]We assume positive integer values of $N_{\mathrm{O}}N/\mathcal{C}$, $N/(\mathcal{C}N_s)$, and $N/(\mathcal{C}N_s \log_2 M)$. In practice this is accomplished through bit padding of the input frame. We also assume the convolutional encoder returns to zero state after encoding the last bit in the frame through sufficient zero padding.

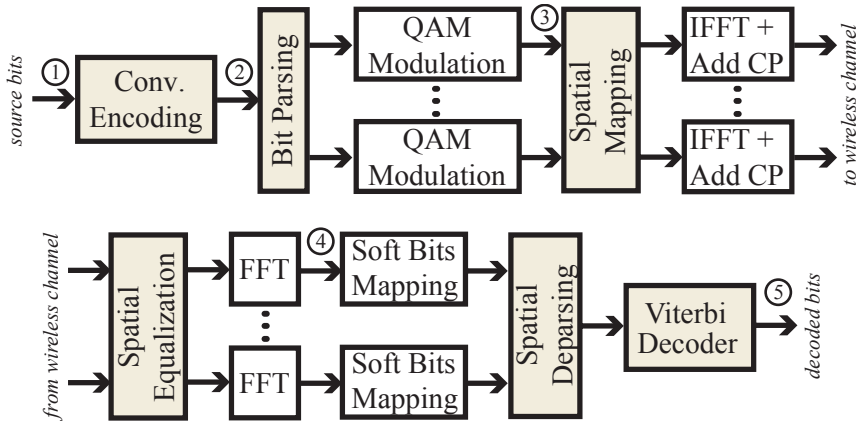

Fig. 1. Transmitter and receiver processing for MIMO-OFDM system with bit-interleaved convolutional forward error correction

TABLE I

NOTABLE STAGES AND MATHEMATICAL NOTATION OF MIMO-OFDM TRANSCEIVER

Stage	**Description**	**Data Quantity**
1	source bits	$b[n] \in \{0,1\}$, $n \in \{0,1,\ldots,KN_{\mathrm{O}}N-1\}$
2	encoded source bits	$c[n] \in \{0,1\}$, $n \in \{0,1,\ldots,\frac{KN_{\mathrm{O}}N}{\mathcal{C}}-1\}$
3	baseband, frequency-domain transmitted QAM symbols	$\mathbf{x}[m,n] \in \mathbb{C}^{N_s}$ $m \in \{0,1,\ldots,N_{\mathrm{O}}-1\}, n \in \{0,1,\ldots,\frac{N}{\mathcal{C}N_s}-1\}$
4	equalized baseband frequency-domain received symbols	$\mathbf{y}[m,n] \in \mathbb{C}^{N_s}$ $m \in \{0,1,\ldots,N_{\mathrm{O}}-1\}, n \in \{0,1,\ldots,\frac{N}{\mathcal{C}N_s}-1\}$
5	estimated source bits	$\hat{b}[n] \in \{0,1\}, n \in \{0,1,\ldots,KN_{\mathrm{O}}N-1\}$

from codewords to multiple spatial streams, but also interleaves bits over the spatial streams and subcarriers. In practical systems this function may not process the entire $KN_{\mathrm{O}}N/\mathcal{C}$-length bit sequence at once due to complexity, latency, and memory constraints. In this paper we consider spatial parsing functions which process L bits consecutively. This does not restrict the implementation of the spatial parsing function (interleaving) since every realizable spatial parsing function can be represented if $L \geq KN_{\mathrm{O}}N/\mathcal{C}$.

After the spatial parsing operation, the bits in each spatial branch $a \in \{1,2,\ldots,N_s\}$ enter the quadrature-amplitude modulator (QAM). Each modulated symbol in each spatial branch is a member of the M-QAM constellation. Therefore, $K = \log_2(M)$ bits are modulated at a time in each spatial branch. Using matrix notation we represent the set of M-QAM samples across all spatial branches as $\{\{\mathbf{x}[m,n]\}_{m=0}^{N_{\mathrm{O}}-1}\}_{n=0}^{N-1}$ in stage 3 where $\mathbf{x}[m,n] \in \mathbb{C}^{N_s}$. The spatial mapping

block transforms each N_s-dimensional complex vector into N_t complex dimensions that map distinctly to each transmit radio frequency (RF) chain ($N_s \leq \min\{N_r, N_t\}$) through a linear precoding matrix $\mathbf{F}[n] \in \mathbb{C}^{N_t \times N_s}$ for each subcarrier n, although the exact linear precoding matrix used in this paper is not specified.

The last digital processing at the transmitter is the inverse discrete Fourier transform (IDFT) and cyclic prefix addition. This final step in the transmitter provides OFDM with significant tolerance to intersymbol interference (ISI). In each of the N_t transmit branches, the IDFT algorithm processes N-length blocks[2]. Each block prepends the last N_{CP} time-domain symbols of that particular block. This operation enables the receiver to observe ISI cyclically for each of the N_{O} OFDM symbols.

At the receiver, the N_r-dimensional (for each of the N_r receive chains) time-domain complex samples are received before stripping the cyclic prefix and equalizing the channel. The equalized samples are mapped equivalently into the discrete Fourier domain and at stage 4 the frequency-domain representation of OFDM symbol m and subcarrier n is given by[3] $\mathbf{y}[m, n]$. After mapping the QAM symbols to soft bits per spatial branch and spatial branch deparsing, the receiver calculates the sequence of log-likelihood ratios corresponding to each of the input convolutionally-encoded bits. The optimal maximum-likelihood (ML) detection of MIMO signals, in general, does not use linear spatial equalization since soft bits are created from each N_r-dimensional received sample. If ideal SVD precoding is implemented in the linear precoding matrix, however, ML detection may be processed on each spatial stream separately [34]. The Viterbi decoder implements ML sequence detection to yield the final binary sequence $\{\{\hat{b}[m, n]\}_{m=0}^{N_{\text{O}}-1}\}_{n=0}^{KN-1}$ in stage 5 as an estimate of the input sequence.

The transceiver in Fig. 1 provides a linear time-invariant input-output relationship of frequency domain QAM symbols over the wireless channel. Given MIMO-OFDM symbol $m \in \{0, 1, \ldots, N_{\text{O}} - 1\}$, subcarrier $n \in \{0, 1, \ldots, N - 1\}$, $\mathbf{H}[n] \in \mathbb{C}^{N_r \times N_t}$ as the wireless channel, $\mathbf{G}[n] \in \mathbb{C}^{N_s \times N_r}$ as the linear equalization matrix, and $\mathbf{v}[m, n] \in \mathbb{C}^{N_r}$ as the thermal noise at

[2]In practice not all N subcarriers carry data. Some subcarriers are dedicated to pilots or guardbands. In multiuser standards like 3GPP-LTE and WiMax, we only consider the subcarriers that are dedicated to a single user.

[3]Spatial equalization and the IFFT are often combined into a single step with the IFFT preceding spatial equalization since equalization in the frequency domain offers favorable complexity for broadband frequency selective wireless channels [33]. Fig. 1 provides a more general framework for equalization.

baseband, then

$$\mathbf{y}[m,n] = \sqrt{E_s}\mathbf{G}[n]\mathbf{H}[n]\mathbf{F}[n]\mathbf{x}[m,n] + \mathbf{G}[n]\mathbf{v}[m,n]. \quad (1)$$

$\tilde{\mathbf{H}}[n] = \mathbf{G}[n]\mathbf{H}[n]\mathbf{F}[n]$ and $\tilde{\mathbf{v}}[m,n] = \mathbf{G}[n]\mathbf{v}[m,n]$ represent effective quantities to include the combined effects of spatial mapping and spatial equalization. All modulated signals will have unit average power such that $\mathbb{E}[\mathbf{x}^*[m,n]\mathbf{x}[m,n]] = 1 \ \forall \ m,n$. Consequently E_s designates the expected total transmitted signal power if $\|\mathbf{F}[n]\|_F^2 = 1$. We will consider complex Gaussian noise vectors $\mathbf{v}[m,n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ where each element has variance σ^2 and mean $0 \Rightarrow \tilde{\mathbf{v}}[m,n] \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{G}[n]\mathbf{G}^*[n])$. The wireless channel $\mathbf{H}[n]$ for subcarrier n observes the slow fading assumption, provided the channel does not change for all OFDM symbols within a frame. This enables the receiver to remove m as a parameter of the effective channel.

For multiple spatial stream transmission, the SNR is defined per-subcarrier and per-spatial stream. For analysis, linear receivers and linear precoders are convenient since the performance can be described with the post-procesing SNR per spatial stream $a \in \{1, 2, \ldots, N_s\}$ as

$$\gamma[a,n] \triangleq \frac{E_s |\, [\mathbf{G}[n]\mathbf{H}[n]\mathbf{F}[n]]_{a,a} \,|^2}{\sum_{a' \neq a} E_s |\, [\mathbf{G}[n]\mathbf{H}[n]\mathbf{F}[n]]_{a,a'} \,|^2 + \sigma^2 \sum_{a'=1}^{N_s} |\, [\mathbf{G}[n]]_{a,a'} \,|^2} \quad (2)$$

in subcarrier n [35]. Common spatial equalization matrices include zero forcing (ZF) where $\mathbf{G}_{\text{ZF}}[n] \triangleq (\mathbf{F}^*[n]\mathbf{H}^*[n]\mathbf{H}[n]\mathbf{F}[n])^{-1}\mathbf{F}^*[n]\mathbf{H}^*[n]$ and minimum mean-square-error (MMSE) where $\mathbf{G}_{\text{MMSE}}[n] \triangleq (\mathbf{F}^*[n]\mathbf{H}^*[n]\mathbf{H}[n]\mathbf{F}[n] + (\sigma^2/E_s)\mathbf{I})^{-1}\mathbf{F}^*[n]\mathbf{H}^*[n]$. For ZF receivers, post-processing SNR simplifies to $\gamma_{\text{ZF}}[a,n] = E_s/(\sigma^2 \sum_{a'=1}^{N_s} |[\mathbf{G}_{\text{ZF}}[n]]_{a,a'}|^2)$, with the desirable property of Gaussian inter-stream interference plus noise.

The key assumptions for this system model are as follows.

1) *Uniform Modulation*: Every symbol over each spatial stream and every subcarrier is modulated with the same constellation order for each transmitted frame. This corresponds to the mandatory modes supported in IEEE 802.11n, WiMax, and single codeword MIMO in 3GPP-LTE. Multiple codeword MIMO, with parallel encoding, interleaving, and modulation operations, is available in 3GPP-LTE as well. Each of the codewords are independent and therefore multiple codeword MIMO may be represented as multiple independent realizations of our current system model. In the optional modes of IEEE 802.11n different modulation orders may be considered for separate spatial streams. This is not covered in the current system model and is treated as future work. An explanation of the challenges is summarized in Section III-E.

2) *Linear Precoding*: All spatial mapping and precoding operations at the transmitter are accomplished through matrix operations (i.e. linear operations), although no assumption is made on the type of linear precoding used (open loop, channel informed, etc.). Nonlinear precoding techniques have been studied in the past [36], however they are not provided in current MIMO-OFDM standards. Moreover, nonlinear precoding methods complicate calculation of post-processing SNR, which is integral to the analysis in this paper.
3) *Linear Equalization*: Linear receivers reduce the complexity of decoding, especially for large dimension MIMO systems, by decoupling maximum likelihood sequence detection of convolutionally-coded bits and MIMO channel equalization. For systems that use SVD precoding at the transmitter, linear receivers achieve ML performance [34]. When SVD precoding is not used, however, linear receivers suffer performance degradation when compared to optimal ML over spatial streams. For simplicity we only consider linear receivers with the understanding that all development in this paper applies to alternative equalization procedures if SNR can be defined per subcarrier and spatial stream.
4) *Slow Fading*: The wireless channel is assumed static for all OFDM symbols in a single frame. This assumption is justified if the coherence time of the wireless channel is larger than the duration of the frame transmission. Slow fading is valid for indoor channels and many outdoor channels [37]. No assumptions are made about the fading correlation between successively transmitted frames.

III. Frame Error Rate Analysis

Prior to the application of supervised learning we analyze the performance of the MIMO-OFDM system defined in the previous section through *frame error rate* (FER) or the probability that the input binary sequence does not match the estimated binary sequence, i.e.

$$\text{FER}(\{\mathbf{H}[n]\}_{n=0}^{N-1}, E_s, \sigma^2) \triangleq \Pr\left[\{b[n]\}_{n=0}^{KN_0N-1} \neq \{\hat{b}[n]\}_{n=0}^{KN_0N-1} | \{\mathbf{H}[n]\}_{n=0}^{N-1}, E_s, \sigma^2\right] \quad (3)$$

for fixed spatial parsing and convolutional coding operations. Technically speaking, this is not a complete definition of FER. In practice, assuming the slow fading model, each frame potentially experiences a different channel realization. Hence, the actual frame error rate over the lifetime of the system is

$$\overline{\text{FER}} \triangleq \mathbb{E}_{\{\mathbf{H}[n]\}_{n=0}^{N-1}}\left[\text{FER}(\{\mathbf{H}[n]\}_{n=0}^{N-1}, E_s, \sigma^2)\right] \quad (4)$$

where statistical expectation is taken over the distribution of channels. In this paper, we will usually refer to FER using (3) and for notational simplicity we will abbreviate this with FER$(\cdot)$.

In this section we investigate FER for MIMO-OFDM systems and arrive at an approximation of FER as a function of post-processing SNR, ordered over subcarriers and spatial streams. This is completed through the analysis of convolutional coding in MIMO-OFDM systems. The reader is encouraged to see Appendix A for details of the analysis. Because the FER approximation in terms of ordered SNR uses certain assumptions on the QAM constellation, the spatial parsing block size, the constellations on each spatial stream, and the spatial equalization procedure, this section also provides the consequences of relaxing these assumptions.

A. FER with Convolution Codes

The convolutionally-encoded sequence $\{c[n]\}_{n=0}^{KN_{\mathrm{O}}N/\mathcal{C}-1}$ is a function of the input binary source sequence $\{b[n]\}_{n=0}^{KN_{\mathrm{O}}N-1}$ and the codeword generator matrix $\mathbf{B}$. Let $\mathbf{b}$ be the $KN_{\mathrm{O}}N$-dimensional binary vector and $\mathbf{c}$ be the $KN_{\mathrm{O}}N/\mathcal{C}$-dimensional binary vector formed from the sequences $\{b[n]\}_{n=0}^{KN_{\mathrm{O}}N-1}$ and $\{c[n]\}_{n=0}^{KN_{\mathrm{O}}N/\mathcal{C}-1}$, respectively. The codeword generator matrix $\mathbf{B}$ is a function of the $q \times r$ generator polynomials characteristic of the convolutional code of rate $\mathcal{C} = q/r$ for coprime $q, r \in \mathbb{N}$ and $r \geq q$. $\mathbf{B}$ is a binary matrix of dimension $\frac{KN_{\mathrm{O}}N}{\mathcal{C}} \times KN_{\mathrm{O}}N$ that relates the convolutional encoding with $\mathbf{c} = \mathbf{B}\mathbf{b}$ where all operations (multiplication and addition) are on the binary field $\mathbb{F}_2$.

We define an error event when the receiver incorrectly chooses codeword $\hat{\mathbf{c}} = \mathbf{B}\hat{\mathbf{b}}$ and $\mathbf{b} \neq \hat{\mathbf{b}}, \mathbf{c} \neq \hat{\mathbf{c}}$. The probability that an error event occurs is characterized by the Hamming distance between distinct codewords generated by $\mathbf{B}$ as well as properties of the wireless channel and physical layer. Without loss of generality (by linearity), the correct codeword is the all-zero binary vector $\mathbf{c} = \mathbf{0}$. The Hamming distance between the codeword $\mathbf{c}$ and $\hat{\mathbf{c}}$, given by $d_H(\mathbf{c}, \hat{\mathbf{c}})$, is equivalent to the weight (total number of non-zero elements) of the binary vector $\mathbf{c}$ (i.e. $w_H(\mathbf{c})$). With these assumptions and definitions the well-known upper bound for the frame error rate with soft-output Viterbi decoding given d_f as the minimum distance error event, N_d as the number of d-distance error events, and P_d as the probability of d-distance error events [17] with convolutional codes is

$$\mathrm{FER}(\cdot) \leq \sum_{d=d_f}^{\infty} N_d P_d. \tag{5}$$

The number of d-distance error events, N_d, is a function of the convolutional encoder and the probability of the error event for d-distance errors, P_d, is additionally a function of $\{\mathbf{H}[n]\}_{n=0}^{N-1}$, E_s, and σ^2.

B. FER Approximation in MIMO-OFDM through SNR Ordering

We first define $\gamma^{(t)} \in \{\gamma[1,0], \ldots, \gamma[N_s, N-1]\}$ as the tth smallest post-processing SNR from (2) for all subcarriers and spatial streams such that

$$\gamma^{(1)} \leq \gamma^{(2)} \leq \cdots \leq \gamma^{(N_s N)} \tag{6}$$

$$\gamma^{(1)} = \min_{a,n} \gamma[a,n] \tag{7}$$

$$\gamma^{(N_s N)} = \max_{a,n} \gamma[a,n] \tag{8}$$

for all $a \in \{1,2,\ldots,N_s\}, n \in \{0,1,\ldots,N-1\}$. In Appendix A we derive an FER approximation for our MIMO-OFDM system model

$$\text{FER}(\cdot) \approx \sum_{d=d_f}^{d_L} \sum_{\mathbf{c} \in \mathsf{C}_{d,L}} \prod_{l=1}^{d} Z\left(\gamma^{(l)}\right) \tag{9}$$

with $\mathsf{C}_{d,L}$ as the set of all possible weight d codewords generated by $\mathbf{B}$ with error events of length at most equal to L, d_L as the maximum codeword weight with L-length error events, and $Z(\cdot)$ as the Bhattacharya parameter defined in Appendix A. This FER approximation is a powerful observation and it enables characterization of the FER without knowledge of bit mappings to subcarrier locations in the spatial parsing process.

C. Spatial Parsing (Interleaving) Block Length

For the approximation in (9), derived in Appendix A, we assume a spatial parsing block length $L \leq N_s N$ as in IEEE 802.11n. Other MIMO-OFDM standards, such as 3GPP-LTE and WiMax, allow for the interleaving block length to be larger than $N_s N$. When $L > N_s N$ the block size for the spatial parsing function exceeds the total number of subcarriers for all spatial streams, hence the worst case post-processing SNR approximation used to derive (9) in Appendix A includes redundant ordered post-processing SNR terms. In either case, as long as L exceeds the truncation length of the convolutional code, the FER approximation is still parametrized by the ordered post-processing SNR over all subcarriers and spatial streams.

D. M-QAM Constellations

The FER approximation in (9) assumes BPSK constellations. The bounds in (14) and (15) of Appendix A that lead to this approximation can be modified to include QAM constellations of general order, assuming Gray coding and equal probability of bit errors if an incorrect constellation point is selected. The probability of bit-error for these general constellations do not lead to simple definitions of P_d and convenient expression in terms of $Z(\cdot)$, the Bhattacharya parameter. Nevertheless, the bounds for M-QAM constellations are similarly a function of the SNR per subcarrier. Hence, different constellations do not change the parameters needed to determine FER (for an example see [15], [38]). In the development of (9) in Appendix A we used approximations to reduce the dependence of FER bounds on the exact placement of bits from the interleaver or spatial parsing function. Similar approximations can be made for general order constellations, although they have yet to be validated as they were for BPSK constellations in [39].

E. Non-uniform Energy Allocation and Constellations

Non-uniform energy allocation over subcarriers and spatial streams is allowed in our system model, although not explicitly. Different power levels over subcarriers and spatial streams may be embedded in the precoding matrices $\mathbf{F}[n]$ for each subcarrier. Non-uniform constellations, however, are not covered under the current system model. Some systems, such as the optional MCS modes of IEEE 802.11n, allow for mixed QAM dimensions over different spatial streams [28]. Non-uniform constellations will further increase the complexity of FER bounds since P_d is determined by the probability of bit error for different sized constellations. This complicates the approximations made in Section A-C to enable feature sets for supervised learning. Consequently, this is an important issue to resolve for future work.

F. Alternative Equalization

The FER approximation in (9) assumes ZF equalization, although other spatial equalization methods may be implemented in practice such as minimum mean-square-errror (MMSE) linear receivers and ML detection over spatial streams (and near-ML approximation). MMSE equalizers do not produce Gaussian inter-stream interference plus noise terms, however, Gaussian approximations are often used [40]. Consequently, we expect the FER approximation in (9) to apply

directly to MMSE receivers with post-processing SNR calculations from (2). Post-processing SNR, however, is not easily defined for full ML spatial equalization. One solution is to use the singular values of the channel matrix in the FER bounds since they characterize the post-processing SNR of ML spatial equalization when ideal SVD precoding is used at the transmitter [34].

IV. Supervised Learning for MIMO-OFDM AMC

In this section machine learning is applied to AMC in the MIMO-OFDM system described in Section II. In Section IV-A an introduction to machine learning is provided as well as the connection between machine learning and AMC. Next, in Section IV-B AMC is translated into a classification problem. In Section IV-C the complete procedure for AMC through machine learning is provided. Finally, in Section IV-D a feature space is designed through consideration of the FER analysis in Section III to enable accurate classification for MIMO-OFDM AMC.

A. Background

Machine learning broadly covers all systems that learn or improve their performance with data observations. In this paper, machine learning provides a framework for *classification*. The goal of classification is to place a label, indexed by a non-negative integer or *class*, on an element of interest. To facilitate machine learning and classification, a quantitative measurement, or *feature set* for the element is created. A *classifier* then determines which class the feature set belongs to by exploiting knowledge obtained from past data observations. By convention the feature set is a vector in real space $\mathbb{R}^p$. Therefore, the classifier essentially maps vectors in $\mathbb{R}^p$ to integers in $\mathbb{N}$. The field of machine learning considers all aspects of the classification problem including choice of the feature set, design of the classifier, and obtaining data observations [22].

To design a classifier that separates the feature space correctly, we need to obtain knowledge through data observation. Data observation may be explicit, such as with *supervised learning* where a *training set* of feature sets and correct class labels is available. Other forms of learning are possible, including unsupervised learning, where the training set includes feature sets without correct class labels. For the purposes of adaptive modulation and coding, we have chosen a supervised learning framework. Therefore, a training set that consists of channel realizations associated with the ideal coding rate, modulation order, and number of spatial streams must

be available. This training set may obtained through system measurements or simulations of the physical layer. To create the training set the PER is simulated/measured for each coding rate, modulation order, and number of spatial streams by repeated transmission of packets. For measurements the resolution of PER in the training set is limited by the coherence time of the channel used to create training sets.

B. AMC as Classification

Adaptive modulation and coding in our MIMO-OFDM system model is the process of selecting the QAM modulation order, M, convolutional coding rate, $\mathcal{C}$, and number of spatial streams, N_s, for a given realization of channel state information. In practice, there is only a finite set of allowable M, $\mathcal{C}$, and N_s triplets, where each element of this set is named a *modulation and coding scheme* (MCS). The MCS set is limited by many factors including practical implementation of different $\mathcal{C}$ values, achievability of large M in wireless hardware, the spatial dimensions available ($N_s \leq \min\{N_r, N_t\}$), and redundancy of performance in various triplets.

Given that the MCS list is finite, we provide indices to this list with $i \in \mathcal{I} \subset \mathbb{N}$ where the cardinality of $\mathcal{I}$ is the number of available M, $\mathcal{C}$, and N_s triplets. The index i, hereafter referred to as the *class*, distinctly maps to a realization of M, $\mathcal{C}$, and N_s to define MCS$_i$. Each class also maps to a single data rate, $\mathcal{R}_i$, provided by the values of M, $\mathcal{C}$, and N_s in MCS$_i$. Therefore, in terms of classification, AMC is the process of selecting a class i, corresponding to MCS$_i$, to maximize $\mathcal{R}_i$ under a reliability constraint for different realizations of the channel state. Following convention, due to its importance in higher layer reliability, we consider the frame error rate reliability constraint [26], [27]:

> A class i is only selected if the corresponding FER of MCS$_i$ for the channel $\{\tilde{\mathbf{H}}[n]\}_{n=0}^{N-1}$ with transmit power E_s and receiver noise variance σ^2, defined as FER$_i(\{\tilde{\mathbf{H}}[n]\}_{n=0}^{N-1}, E_s, \sigma^2)$, is less than or equal to $\mathcal{T}$ (i.e. FER$_i(\cdot) \leq \mathcal{T}$).

Once the set of classes that satisfy the FER constraint are determined, class i is selected to correspond to the MCS$_i$ with highest rate $\mathcal{R}_i$. Therefore, classification selects

$$\arg\max_i \{\mathcal{R}_i : \text{FER}_i(\cdot) \leq \mathcal{T}\} \tag{10}$$

to maximize the performance for a given channel realization. If no MCS satisfies (10) the most reliable MCS will be selected. If two different classes satisfy the FER constraint, but yield the

same rate, a rule is needed to determine the class for a given channel. In the next section, Section IV-C, which defines our classifier for AMC with supervised learning, we always choose the more robust MCS corresponding to a lower number of spatial streams.

AMC using (10) does not necessarily maximize the average data rate over the lifetime of the network. Each realization of channel state maps to the same class throughout the wireless network lifetime (only one class is assumed to be correct). Therefore, AMC selection using (10) will always be conservative because every channel realization must meet the FER target. The FER target, however, does not constrain the FER associated with each channel to be less than the target, just the channels in the window of time used to calculate the FER statistic. This conservative MCS selection problem is also found with related work on AMC [24], [26], [27]. The optimal data rate that satisfies the FER constraint is achieved by selecting an entire sequence of MCS values over the window of time used to calculate the FER statistic.

Machine learning provides a framework to approach optimal AMC with FER constraints through reinforcement learning. Essentially, if the classification algorithm knows that the current measurement of FER is below the target, it can bias the FER constraint in (10) above the actual target, to potentially increase average data rate. Our work is a first step towards this optimal procedure in MIMO-OFDM systems since we provide the MCS classification infrastructure with a frame error rate constraint.

C. k-Nearest Neighbor AMC

In Section IV-B we translated AMC into a classification problem. Now, we define the classifier used in this paper. The choice of the classification algorithm depends largely on the structure of classes in the feature space. Unfortunately, for the feature sets chosen in this paper, namely single-dimensional link-quality metrics and ordered post-processing SNR feature sets (to be discussed in the next section), little is known about the functional mapping between feature sets and MCS. As a consequence, non-parametric classification algorithms are preferred [22], [41]. Therefore, we have chosen k-nearest neighbor (k-NN) due to its ability to provide accurate class estimates without knowledge of a functional mapping between the feature sets and the class. k-NN does suffer the cost of increased complexity and memory storage compared to parametric classification algorithms. In this paper, however, we are primarily concerned with validation of supervised learning for AMC. Therefore, we treat the problem of finding other classification

algorithms that reduce complexity and memory storage as future work.

The k-NN system is trained with W distinct realizations of the feature set and its associated class (i.e. the training set). We define a realization index $w \in \{0, 1, 2, \ldots, W-1\}$ for each feature set in the training set. The feature set corresponding to index w in the training set, $\mathbf{z}_w \in \mathbb{R}^p$, is assigned to a class $i \in \mathcal{I}$ according to (10) using the realization of channel state from which $\mathbf{z}_w$ was evaluated. The corresponding MCS for each element of the training set follows the mapping

$$\{\mathbf{z}_w\}_{w=0}^{W-1} \mapsto \{i(w)\}_{w=0}^{W-1} \tag{11}$$

where $i(w) \in \mathcal{I}$ is the ideal AMC classification according to (10).

Empirical results show that supervised learning does not work well when the set of classes, $\mathcal{I}$, contains MCS with different spatial stream orders. In general, feature sets that are calculated from post-processing SNR measurements (such as all feature sets considered in this paper) have different meaning depending on the number of spatial streams. For example, with one spatial stream, there are N distinct SNR measurements corresponding to each subcarrier. For two spatial streams, however, there are $2N$ distinct SNR measurements corresponding to each subcarrier and spatial stream. Therefore, statistics gathered from post processing SNR with mixed spatial stream dimensions are difficult to compare. Consequently, we apply N_s separate classification procedures over N_s separate training sets. Each training set only considers MCS with the same number of spatial streams. We define the classes corresponding to MCS with a spatial streams as $\mathcal{I}_a$. Likewise we define $\{\mathbf{z}_{w,a}\}_{w=0}^{W_a-1}$ and $\{i(w,a)\}_{w=0}^{W_a-1}$ as the feature set realizations and associated classes for the training set with a spatial streams.

The algorithm for k-NN AMC is defined below and the procedure for supervised learning with N_s spatial streams is illustrated in Fig. 2. k-NN AMC predicts the class for a new realization of the channel state information using a query $\mathbf{q} \in \mathbb{R}^p$ representing the feature set for this channel realization. The depth of the search, k, and distance metric $d(\cdot, \cdot)$ remain arbitrary.

Algorithm *k-NN for AMC*

1. **for** $a \leftarrow 1$ **to** N_s
2. $\quad n_i \leftarrow 0 \ \forall \ i \in \mathcal{I}_a$
3. $\quad$ **for** $\ell \leftarrow 1$ **to** k
4. $\quad\quad w_\ell \leftarrow \arg\min_w \{d(\mathbf{z}_{w,a}, \mathbf{q}) : w \notin \{w_1, \ldots, w_{\ell-1}\}\}$
5. $\quad\quad n_{i(w_\ell)} \leftarrow n_{i(w_\ell)} + 1$

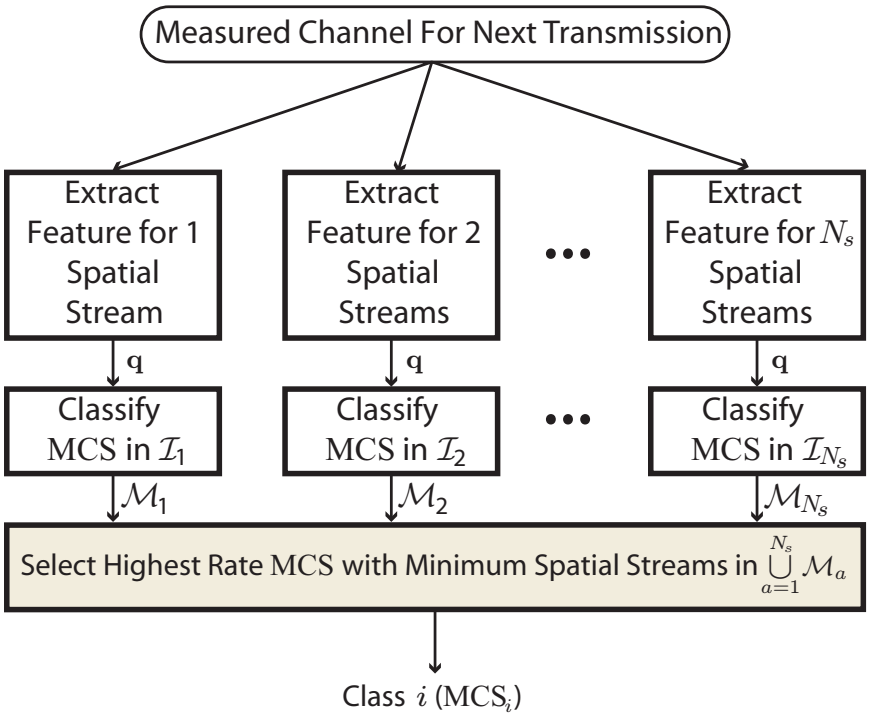

Fig. 2. Supervised Learning AMC procedure that selects a single MCS across multiple spatial streams

6. $\mathcal{M}_a = \min\{\arg\max_i\{n_i : i \in \mathcal{I}_a\}\}$
7. **return** $\min\{\arg\max_i\{\mathcal{R}_i : i \in \cup_{a=1}^{N_s}\mathcal{M}_a\}$

Implementation of the algorithm uses the distance metric to determine the "neighbors" in the training set. Therefore, given a query from a channel observation the classifier searches through the training set to find the k nearest neighbors in terms of the distance metric. The class assigned to the query is the class that occurs most often among the k-nearest neighbors in the training set [42]. This paper uses a Euclidean distance metric which implies that all dimensions of the feature set are equally weighted. Other distance metrics may be considered to improve performance, depending on the structure of the classes in the feature space [22].

One unique property of this algorithm is the rule for breaking ties. If two distinct classes $i' \neq i''$ both in $\mathcal{I}_a$ return the same number of neighbors to the query for a spatial streams, the tie is broken using the rate. That is, without loss of generality, i' is selected if $\mathcal{R}_{i'} < \mathcal{R}_{i''}$. If i' is selected through classification in $\mathcal{I}_{a'}$ and i'' is selected through classification in $\mathcal{I}_{a''}$ $(a' \neq a'')$ and $\mathcal{R}_{i'} = \mathcal{R}_{i''}$, which often occurs when more than one spatial stream is considered, then we

will always select the lower index, i.e. i' is selected since $a' < a''$. Hence, we are biased towards lower rate selection and a lower number of spatial streams since empirical studies show this often implies higher communication reliability. A similar selection strategy was chosen in [27].

D. Feature Space Selection

In Section IV-C we designed a supervised learning framework for AMC through k-NN classification. To provide accurate classification, however, an appropriate feature set must be found. The choice of an appropriate feature space is critical due to the phenomenon in supervised learning known as "the curse of dimensionality" [43]. That is, each dimension that is added to the feature space, requires increasingly more elements in the training set to uniformly sample the feature space. For example, a 1-dimensional feature requires a factor 10^{18} less samples than a training set with 10-dimensional features and equal resolution [44]. It is necessary to find a low-dimension set that also provides an accurate mapping to the MIMO-OFDM FER to achieve accuracy in the classification from (10). In IEEE 802.11n, for example, we cannot include all the channel subcarriers in the feature space. For IEEE 802.11n systems with 2 transmit and 2 receive antennas, using the full channel in the feature space would require 416 dimensions, resulting in complete failure for classification with training sets of reasonable size.

Prior work on AMC for MIMO-OFDM has resulted in several one- or two-dimensional link-quality metrics extracted from channel state that attempt to parametrize FER for look-up-table based AMC. Initially, our naive implementation of supervised learning using these link-quality metrics did not produce satisfactory performance in predicting MCS values. These link-quality metrics, in general, only provide coarse information about the channel quality over all subcarriers and spatial streams. Empirical results have shown, however, that more information about the quality of the channel per-subcarrier and spatial stream in frequency and spatial selective channels is needed to perform MIMO-OFDM AMC. This is confirmed through the simulations of Section V in this paper. Because the established link quality metrics are not suitable to predict the best MCS, we proceed with the creation of a feature set for supervised learning in MIMO-OFDM based on post-processing SNR, ordered over subcarriers and spatial streams.

Given the approximation in (9), we know the set of ordered SNR, $\left\{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(N_s N)}\right\}$, uniquely parametrizes FER in MIMO-OFDM systems. Unfortunately, the dimension of the set of ordered SNR, $N_s N$, is still too large to represent the feature set for supervised learning. One

key observation, however, is the significant correlation between the ordered SNR values due to the sorting process and the bounds on achievable SNR. We exploit this correlation through a simple low-dimensional mapping, $\mathcal{P}$, from the $N_s N$-dimensional ordered post-processing SNR set to the p-dimensional ordered post-processing SNR feature set, defined

$$\mathcal{P}\left(\left\{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(N_s N)}\right\}\right) \triangleq \left\{\gamma^{(n_1)}, \gamma^{(n_2)}, \ldots, \gamma^{(n_p)}\right\} \tag{12}$$

with indices $n_1, n_2, \ldots, n_p \in \{1, 2, \ldots, N_s N\}$. Essentially, to calculate our feature space we first sort the post-processing SNR over all subcarriers and spatial streams and then select p of these sorted SNR values. In Section V we show a procedure for discovering the best indices, $\{n_1, n_2, \ldots, n_p\}$, for a given training set. If the correct subcarrier indices are chosen, an accurate approximation of the set $\left\{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(N_s N)}\right\}$ is obtained through only p dimensions.

The assumptions that yield the FER approximation in (19) as a function of ordered SNR assume a frequency selective channel. We suggest, however, that an ordered post-processing SNR feature set will also suitably parametrize the FER of a frequency flat channel. All subcarriers for each spatial stream will observe the same SNR. Hence, the frame error rate is a function of a single post-processing SNR for each spatial stream from (15). In other words, any ordering within a spatial stream will result in the same performance. If an ordered subcarrier feature set selects at least one subcarrier from each spatial stream, this should adequately parametrize FER in frequency flat channels. Therefore, post-processing SNR, ordered over subcarriers and spatial streams, provides a low-dimension feature set through $\mathcal{P}$ that characterizes FER in frequency-flat and frequency-selective MIMO channels.

V. k-NN AMC FOR IEEE 802.11N

To illustrate the performance of our AMC algorithm we follow the MCS specifications from IEEE 802.11n [45] with 2 receive antennas and 2 transmit antennas. Consider the first 16 MCS for the IEEE 802.11n physical layer highlighted in Table II. Each class i for MCS$_i$ defines a convolutional coding rate, QAM modulation order, M, and number of spatial streams, N_s, to determine the physical layer data rate $\mathcal{R}_i$. This MCS table represents the modulation and coding schemes most commonly implemented in 2×2 IEEE 802.11n devices in 20 MHz channels. The simulation study that follows is composed of two parts. First, we determine a suitable feature space dimension for ordered SNR feature sets, the best subcarrier indices for that dimension,

TABLE II

IEEE 802.11N MCS LIST

MCS$_i$	N_s	M	Code Rate	$\mathcal{R}_i$
$i = 0$	1	2	1/2	6.5 Mbps
$i = 1$	1	4	1/2	13.0 Mbps
$i = 2$	1	4	3/4	19.5 Mbps
$i = 3$	1	16	1/2	26.0 Mbps
$i = 4$	1	16	3/4	39.0 Mbps
$i = 5$	1	64	2/3	52.0 Mbps
$i = 6$	1	64	3/4	58.5 Mbps
$i = 7$	1	64	5/6	65.0 Mbps
$i = 8$	2	2	1/2	13.0 Mbps
$i = 9$	2	4	1/2	26.0 Mbps
$i = 10$	2	4	3/4	39.0 Mbps
$i = 11$	2	16	1/2	52.0 Mbps
$i = 12$	2	16	3/4	78.0 Mbps
$i = 13$	2	64	2/3	104.0 Mbps
$i = 14$	2	64	3/4	117.0 Mbps
$i = 15$	2	64	5/6	130.0 Mbps

and the best k to select for channels of varying selectivity. Second, we analyze the system performance (FER and spectral efficiency) of k-NN AMC using the value of d, the indices, and the value of k determined from the first part of the simulation.

A. Simulation Assumptions

The system used for simulation studies in this section assumes the following.

- *Fixed Frame Length*: All frames are 128 Bytes in length. Frame error rate is a function of the number of bits per frame as well as the bit error rate. In practice many different frame lengths are used depending on the traffic source. For simplicity in simulations this paper does not consider different frame lengths.
- *Perfect Synchronization*: We neglect the pilot tones and training sequence. Errors in synchronization can lead to frame errors that are not associated with improper selection of the MCS. Therefore, channel frequency offset and symbol timing are not considered.

- *Perfect Channel Knowledge*: We assume that k-NN AMC knows the effective channel at the receiver perfectly. Channel estimation error results in two forms of performance degradation: symbol equalization error and feature set extraction error. Fortunately, classification error due to equalization performance degradation is minimal since the training set is able to capture this effect. Feature set extraction error, however, will result in misclassified features both in the training set and during the k-NN AMC process. The exact degree of this effect is unknown and an extensive analysis of this effect is treated as future work. Although not presented in this paper, simulations and measurements have shown that in an interference-free environment feature set extraction error (primarily due to thermal noise) is mollified by the SNR ordering operation as long as the extreme values of SNR (largest or smallest) are not used to characterize the SNR profile.
- *Zero-Forcing Linear Equalization*: Equalization in this paper uses a soft-output ZF structure in the frequency domain to compute log likelihood ratios for a Viterbi detector. As mentioned in Section III-F of the Appendix other equalization procedures are compatible with subcarrier ordering feature sets by redefining the post-processing SNR ratio.
- *No Space-Time Processing*: No space-time processing techniques such as space-time block coding, beamforming, or precoding occur during simulation. With space-time processing techniques, not only does the MCS have to be determined, but additionally a decision is made with regard to which precoding matrix or diversity technique to use. In this work we highlight MCS selection and the challenges presented for practical MIMO-OFDM systems.

The basic parameters of this IEEE 802.11n-based study of k-NN AMC are summarized in Table III. In IEEE 802.11n the number of data subcarriers (discrete Fourier transform (DFT) size minus pilot subcarriers and guard tones) $N_{ds} = 52$ for 20 MHz channels. Therefore, the complete ordered post-processing SNR set is $\left\{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(N_s \times 52)}\right\}$.

In Section V-B we determine the dimension of the ordered post-processing SNR feature set, the indices in the ordered post processing SNR profile to represent the ordered SNR feature set, as well as the best value of k in k-NN. In these simulations we use three channel distributions, all generated from uniform power delay profiles: (1) frequency-flat channels (1-tap) (2) moderately frequency selective channels (3-tap) (3) very frequency selective channels (8-tap). Eight taps represent the maximum frequency selectivity since the IEEE 802.11n cyclic prefix length is either 8 or 16 time-domain symbols in length for 20 MHz channels. For each of these 3 channel

TABLE III

SIMULATION PARAMETERS FOR k-NN AMC IN IEEE 802.11N

Property	Value
Frequency Flat Channel	1 tap $\in \mathbb{C}^{N_r \times N_t}$ each single-tap matrix element $\sim \mathcal{CN}(0,1)$
Moderately Selective Channel	3 taps, each tap $\in \mathbb{C}^{N_r \times N_t}$ each tap matrix element $\sim \mathcal{CN}(0,1)$
Very Selective Channel	8 taps, each tap $\in \mathbb{C}^{N_r \times N_t}$ each tap matrix element $\sim \mathcal{CN}(0,1)$
Transmit Energy (above sets only)	$10\log_{10}(E_s/\sigma^2) \in \{1, 2, \ldots, 28\}$
Realizations (W) (above sets only)	$1000 \times 28 = 2.8 \times 10^4$
Performance Analysis Channel	IEEE 802.11n Channel Model B 5000 Realizations total - 5 to 50 m
Classes (i)	$i \in \{0, 1, \ldots, 15\}$ - MCS in IEEE 802.11n
FER Target ($\mathcal{T}$)	0.1 (10% FER)
$N_r \times N_t$	$2 \times 2 \Rightarrow N_s \in \{1, 2\}$
k	25
$d(\cdot)$	Euclidean distance

distributions, we create a training set and test set each with 1000 realizations of the $N_r \times N_t$ channel matrix for each tap. Every matrix element for every tap is complex-Gaussian distributed with zero mean and unit variance. Furthermore, each of the 1000 channel realizations in the training and test sets are combined with 28 values of E_s/σ^2, resulting in 2.8×10^4 feature realizations in each training and test set. The range of E_s/σ^2 specified in Table III was chosen such that each training and test set of channel data observe a diverse MCS classification. The simulations are completed by first training the k-NN classifier on the training set for each of the 3 channels. The classifiers are then verified on a test set for each of the 3 channels. Although the test set and training set observe the same statistical distribution, they are each distinct realizations of the channel (training and testing are separate).

B. Selecting the Subcarrier Ordering Indices and k

To determine the best indices for ordered post-processing SNR feature sets we perform a brute force search over all possible ordered subcarrier feature sets up to four dimensions. This is completed as follows for a fixed feature space dimension, d, fixed number of spatial streams, and fixed neighbor size, k.

1) Given k and d select a set of ordered subcarrier indices not yet simulated.
2) For each channel and E_s/σ^2 realization in the training set compute the post-processing SNR in (2), sort according to (6), and select the ordered post-processing SNR values corresponding to the indices in 1) to construct the feature set, $\mathbf{z}$.
3) Associate this feature set with the correct MCS label for the channel realization in the training set.
4) Repeat 2) through 3) until all channels in the training set are exhausted.
5) For each channel and E_s/σ^2 realization in the test set compute the post-processing SNR in (2), sort according to (6), and select the ordered post-processing SNR values corresponding to the indices in 1) to construct a query, $\mathbf{q}$.
6) Find the k-nearest neighbors to the query in the training set. Apply the MCS label to the query according to the k-NN AMC algorithm in Section IV-C.
7) Repeat 5) through 6) until all channels in the test set are exhausted.
8) Count all errors k-NN AMC made in 5) and 6) for predicting the MCS of queries in the test set.
9) Repeat 1) through 8) until all possible ordered subcarrier indices are simulated. Choose the subcarrier indices with the lowest number of errors for each fixed dimension, d.

Table IV shows the results to this search for the 3 and 8 tap channel distributions with $k = 25$. The 1-tap channel was not included because its solution is trivial; any subcarrier feature set that includes post-processing SNR from each spatial stream will perform the same. Only four 4 dimensions are shown in Table IV since more than 4 dimensions did not improve performance substantially.

Using the indices found in Table IV with $d = 4$ to create ordered SNR feature sets, we also investigate the performance of k-NN AMC for different values of k. Fig. 3 shows the performance through the percentage of correct MCS labels assigned in the test set with the 1,

TABLE IV

BEST SUBCARRIER ORDERING SETS FOR k-NN AMC IN IEEE 802.11N (BRUTE FORCE SEARCH)

N_s	d	3 Tap $(n_1, \ldots, n_d)$	8 Tap $(n_1, \ldots, n_d)$
1	1	(13)	(17)
1	2	(8,31)	(10,31)
1	3	(6,16,35)	(6,17,39)
1	4	(5,10,23,40)	(5,12,24,38)
2	1	(25)	(27)
2	2	(11,52)	(14,46)
2	3	(7,24,61)	(7,23,56)
2	4	(6,13,24,56)	(6,8,27,50)

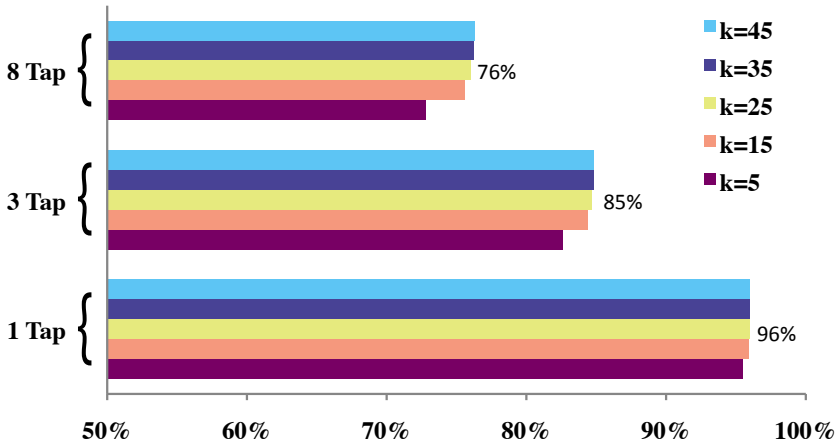

Fig. 3. Correct classification percentages in the test set of 1, 3, and 8 tap channels using the best 4-dimension feature sets from Table IV. The different values of k are displayed on the clustered bar graph. The order of the color code in each bar graph cluster matches the order in the legend.

3, and 8 tap channels when k is varied from 5 to 45. Initially, there is a significant performance improvement for increasing k, but around $k = 25$, the performance levels out for increasing values of k in any channel set. Generally, increasing the size of k reduces errors by limiting the effect of mislabeling that occurs during training set construction. Additionally, classification errors occur when queries from the test set are not close in the distance metric to any feature sets corresponding to channel realizations in the training set. Large k values can help to reduce the effect of such occurrences. Due to the limited size of training sets, however, increasing k

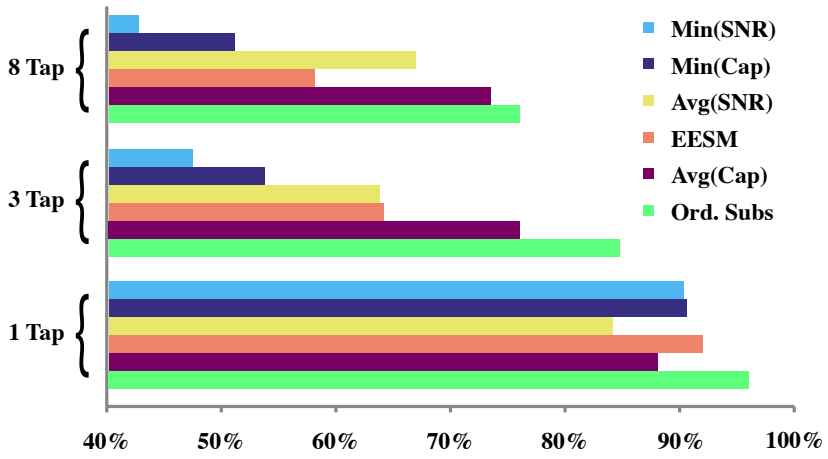

Fig. 4. Correct classification percentages in the test set with $1, 3$, and 8 tap channels and $k = 25$ for several feature sets . The ordered post-processing SNR feature sets use the best indices from Table IV with $d = 4$.The order of the color code in each bar graph cluster matches the order in the legend.

past a certain point will start to produce errors. It is advised to maintain k as small as possible to minimize processing time during k-NN searches.

C. Comparison to Alternative Feature Spaces

For completeness and validation of our subcarrier ordering feature set, we compare k-NN AMC with different feature sets. We use the indices from Table IV to construct 4-dimensional ordered post-processing SNR feature sets in the 1, 3, and 8 tap channels. Fig. 4 compares the performance (through the percentage of MCS values correctly selected) of ordered post-processing SNR feature sets to single-dimensional link quality metrics:

- minimum SNR $\rightarrow \mathbf{z} = \gamma^{(1)}$
- average SNR $\rightarrow \mathbf{z} = \sum_{t=1}^{N_s N} \gamma^{(t)}/(N_s N)$
- minimum capacity $\rightarrow \mathbf{z} = \log_2(1 + \gamma^{(1)})$
- effective exponential SNR mapping (EESM) $\rightarrow \mathbf{z} = \beta \log[\sum_{t=1}^{N_s N} \exp(-\gamma^{(t)}/\beta)/(N_s N)]$, and
- average capacity $\rightarrow \mathbf{z} = \sum_{t=1}^{N_s N} \log_2(1 + \gamma^{(t)})/(N_s N)$.

Each of these features is used in prior work as link-quality metrics to map from MCS to FER. Absent from this list are mutual information derived metrics. Mutual information metrics depend

explicitly on M, the QAM modulation order. Hence, this link-quality metric is not applicable to the k-NN AMC procedure that uses a single feature realization for all MCS within a single spatial stream. EESM features, when used in look-up-tables, are tuned for each MCS by adjusting the β parameter between $\beta = 1$ and $\beta = 25$ [46]. In our scenario, however, β tuning is unnecessary since we do not need the FER associated with each EESM metric to be aligned. During testing, tuning β between 1 and 25 did not change the number of classification errors.

The performance of ordered post-processing SNR feature sets in k-NN AMC outperforms all other potential feature sets for all three channel distributions. Although all link quality metrics perform well on 1 tap channel sets, ordered SNR feature sets perform significantly better. This is somewhat surprising because all of the single-dimensional link-quality metrics should uniquely parametrize frequency-flat channels. However, in the MIMO-OFDM system with $N_s > 1$ spatial streams, there are multiple metrics to consider, namely the post-processing SNR for each of N_s streams in the frequency flat channel. Therefore, single-dimensional link-quality metrics do not uniquely parametrize FER and suffer from classification errors.

Also notable is the performance of the different feature sets in 8-tap channels. The gap between the performance of subcarrier ordering and the alternative feature sets has narrowed. The performance of average capacity feature sets approaches ordered post-processing SNR feature sets. This is due to the curse of dimensionality. If more taps are observed in the channel (with a uniform power delay profile), more frequency selectivity is observed. Hence, more taps yield more diversity in the ordered post-processing SNR profile. The difference in performance between 3 and 8 tap channels, on the other hand, is minimal for the feature sets based on link quality metrics. Single-dimension feature sets are not as dependent on larger training sets to improve performance since they only gather coarse information from the ordered post-processing SNR profile.

D. Frame Error Rate and Spectral Efficiency

Next, we measure the performance of k-NN AMC as a function of feature set. Again, distinct training and test sets are constructed, this time with 5000 channel realizations in IEEE 802.11n channel model B over a transmission distance of $5 - 50$ meters [29]. k-NN is used with $k = 25$ to meet a target FER of 10% ($\mathcal{T} = 0.1$). Subcarrier ordering feature sets use four dimensions with indices chosen from the 8 tap channel scenario in Table IV. Fig. 5 shows the frame error

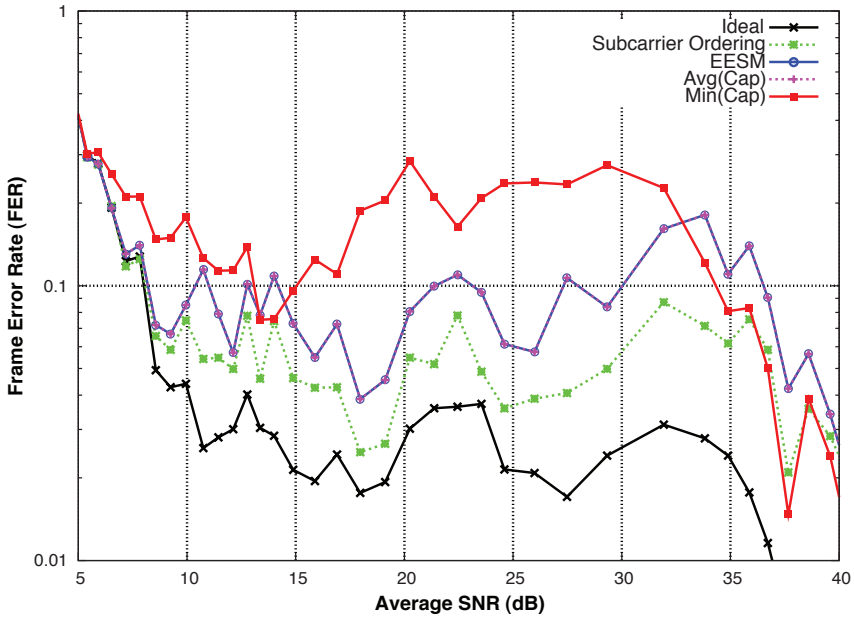

Fig. 5. Frame error rate as a function of average SNR (over subcarriers and spatial streams from equation (2)) when k-NN AMC is applied with different feature sets in IEEE 802.11n channel model B. FER target is 10%. Average capacity and EESM feature sets display the same frame error rate.

rate as a function of SNR averaged over subcarriers and spatial streams when k-NN AMC is implemented with several feature sets. For clarity in the plots, minimum SNR and average SNR feature sets were not included because their performance was inferior to the remaining feature sets. The ideal k-NN AMC procedure follows the adaptation criterion in (10).

The frame error rate constraint cannot be met when average SNR is less than ≈ 7 dB. This occurs when MCS_0 (the most robust MCS) does not observe the 10% FER constraint. Above 7 dB the classifier determines which feature sets allow k-NN AMC to meet the FER constraint. Although subcarrier ordering does not exactly match the FER performance of ideal MCS selection in (10), it is the only feature that does not exceed the 10% FER threshold. The next best feature sets were EESM and average capacity, which both showed identical performance for k-NN AMC in IEEE 802.11n channel model B.

The corresponding spectral efficiency of k-NN AMC in IEEE 802.11n channel model B is shown in Fig. 6. The spectral efficiency for each channel realization is defined as the data rate of the MCS selected divided by the bandwidth of the transmission. Therefore, the maximum spectral efficiency corresponds to MCS_{15} with 130 Mbps$/20$ MHz $= 6.5$ bps$/$Hz. Similarly the

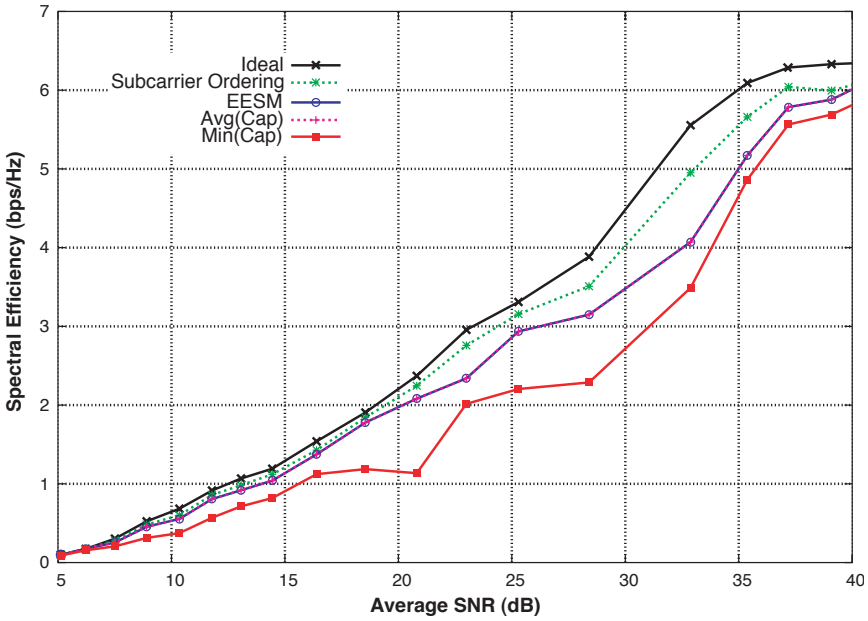

Fig. 6. Spectral efficiency as a function of average SNR (over subcarriers and spatial streams using equation (2)) when k-NN AMC is applied with different feature sets in IEEE 802.11n channel model B. FER target is 10%. Average capacity and EESM feature sets display the same spectral efficiency.

minimum spectral efficiency is $.375$ bps/Hz for MCS_0. The spectral efficiency in Fig. 6 is the spectral efficiency averaged over the all channels in the test set corresponding to each average SNR. Following the convention of [27], if the predicted MCS does not meets the FER constraint, the spectral efficiency is 0 bps/Hz. The spectral efficiency of k-NN AMC with ordered post-processing SNR feature sets distinctly outperforms the best performing feature sets based on link-quality metrics by as much as 1 bps/Hz.

To gain more insight into the performance for different feature sets, we plot a histogram of FER in Fig. 7. Ideal adaptation displays a much higher probability of selecting an MCS with FER that meets our target of 10%. Conversely, poor link-quality metrics (see minimum capacity) have a higher probability of improper MCS selection resulting in an FER that exceeds the target. There is a large percentage of FER values observed at FER $= 0.01$ because the simulation is not able to resolve FER < 0.01. There are also a large number of AMC selections that yield FER $= 1$. This occurs when an egregious error is made in the MCS selection, or when the SNR has fallen below the point at which any packets can be successfully transmitted.

In Fig. 8 we observe the measured probability that MIMO modes are selected ($\Pr[N_s > 1]$)

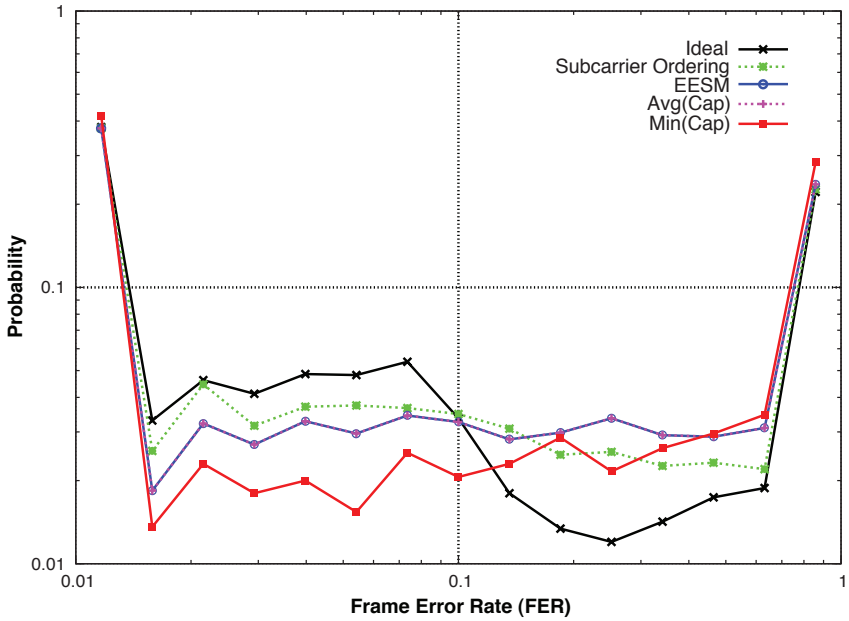

Fig. 7. The measured probability of FER over the entire distribution of channel realizations when k-NN AMC is applied with different feature sets in IEEE 802.11n channel model B. FER target is 10%.

as a function of average SNR. Minimum capacity features do not have knowledge of the entire SNR profile, but only the capacity that corresponds to the minimum value of post-processing SNR. A frequency-flat and frequency-selective sorted post-processing SNR profile may have the same minimum capacity metric, but vastly different FER characteristics for different MCS. This is especially true when $N_s > 1$ as more selectivity is added to the SNR profile. A similar observation can be made for EESM and average capacity metrics. They contain average information about the channel quality over subcarriers, however, they cannot distinguish different trends in the SNR profile. Consequently, subcarrier ordering is best suited to determine when MIMO modes should be used. This is a primary reason why the throughput of subcarrier ordering so closely matches that of ideal adaptation. For example, in Fig. 8, the region between 20 and 35 dB shows the most errors in MIMO mode selection. From the spectral efficiency curves in Fig. 6, this region also shows where feature sets based on link-quality metrics observe the largest gap between ideal MCS selection.

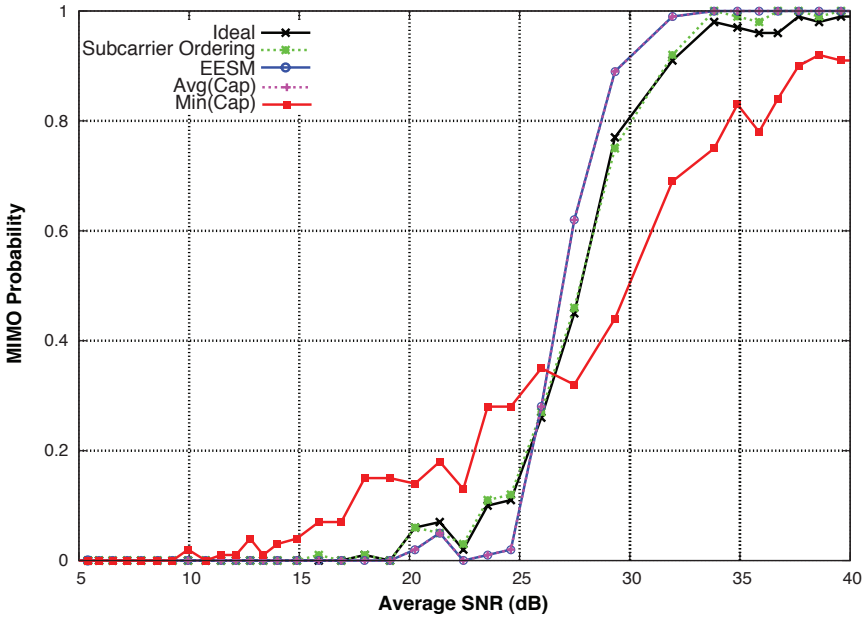

Fig. 8. The measured probability that MIMO is selected at each average SNR (over subcarriers and spatial streams using equation (2)) when k-NN AMC is applied with different feature sets in IEEE 802.11n channel model B. FER target is 10%.

VI. Conclusions and Future Work

Convolutionally-coded MIMO-OFDM systems provide challenging AMC procedures in spatial- and frequency-selective channels. In this paper we have shown that supervised learning techniques provide a framework for effective AMC in MIMO-OFDM. The key to the application of supervised learning techniques to AMC in MIMO-OFDM is subcarrier ordering. Subcarrier ordering provides an intuitive low-dimension feature set that also parametrizes a bound on FER in MIMO-OFDM convolutionally-coded systems, assuming interleaving is adequate over space and frequency.

One extension of this work is to consider online and reinforcement learning procedures that not only update the training set during the lifetime of the wireless network, but also change classification policies with information on the past network performance. Another potential extension of this work would be to consider a joint procedure for AMC and space-time processing selection (precoding matrix and space-time block coding selection). Whether or not supervised learning is also applicable to this scenario is not yet known. This will depend on whether a suitable feature space is available to select space-time processing procedures. Finally, the current

system model does not allow for non-uniform QAM modulation over different spatial streams. Subcarrier ordering for feature sets, as conceived in this paper, is insufficient because ordering over spatial streams destroys information about the spatial stream location.

APPENDIX A
FER EXPRESSIONS IN MIMO-OFDM

To improve readability of this paper, most of the development to represent FER as a function of ordered SNR is placed in the appendix. In Section A-A a FER bound for SISO-OFDM is created. This leads to a FER bound to MIMO-OFDM for ZF receivers in Section A-B. Finally, Section A-C describes an approximation of FER for MIMO-OFDM as a function of ordered SNR.

A. FER for SISO-OFDM

Consider single-input-single-output OFDM (SISO-OFDM) systems ($N_s = 1$) in block fading with binary phase-shift keying (BPSK). One critical component for performance analysis is binary interleaving, which occurs in the frequency domain in order to protect the system from error events. We characterize interleaving with the permutation function

$$\pi : \{0, 1, \ldots, KN_{\mathrm{O}}N/\mathcal{C} - 1\} \rightarrow \{0, 1, \ldots, KN_{\mathrm{O}}N/\mathcal{C} - 1\} \tag{13}$$

that acts on the indices of the binary sequence. Let the notation $\pi(n)$ denote the index of the interleaved sequence corresponding to index n of the sequence before interleaving.

Following a procedure similar to the frame error rate analysis in [39], we define the Bhattacharya parameters that characterize P_d in (5). For an AWGN channel with BPSK constellations and SNR γ, the Viterbi bound states $\mathrm{FER}(\cdot) < \sum_{d=d_f} N_d \exp(-d\gamma) = \sum_{d=d_f} N_d Z(\gamma)^d$ where $Z(\gamma) \triangleq \exp(-\gamma)$ is known as the Bhattacharya parameter for AWGN channels [47]. Essentially, the probability of selecting the wrong d-length codeword in independent realizations of white noise is a power of the Bhattacharya parameter. The product of the d Bhattacharya parameters establish an upper bound on the probability that a d-weight codeword is selected when an all-zero codeword was transmitted. This bound can also be directly translated to flat fading by incorporating the channel gain into the noise variance. In the SISO-OFDM system we make no assumption on the frequency selectivity of the channel as long as the multipath delay does not

exceed the cyclic prefix length. Hence, we observe N-fading channels corresponding to each subcarrier n, each parameterized by $\gamma[n] \triangleq E_s|\mathbf{H}[n]|^2/\sigma^2$, the SISO SNR. Therefore our SISO-OFDM FER bound with C_d as the set of all possible codewords generated by $\mathbf{B}$ with weight d is

$$\text{FER}(\cdot) < \sum_{d=d_f}^{\infty} \sum_{\mathbf{c} \in \mathsf{C}_d} \prod_{l=1}^{d} Z\left(\gamma\left[\text{mod}\left(\pi(\ell_l^{(\mathbf{c})}), N\right)\right]\right) \tag{14}$$

where $\{\ell_1^{(\mathbf{c})}, \ell_2^{(\mathbf{c})}, \ldots, \ell_d^{(\mathbf{c})}\}$ is the set of indices for codeword $\mathbf{c}$ that reference non-zero bits and $\text{mod}(n, N)$ returns the remainder of $\frac{n}{N}$. If all possible error events from weight d codewords are enumerated, as in the set C_d, the probability that soft-input Viterbi decoders suffer from each error event is the product of the Bhattacharya parameters associated with each post-processing SNR that the error event maps to.

B. FER for MIMO-OFDM with Zero-Forcing Receivers

The SISO-OFDM FER bound suggests that the system depends on the SNR for each of the N subcarriers, the distance properties of the convolutional code, as well as the interleaving permutation function π. For MIMO-OFDM with linear receivers, it is straightforward to calculate per-stream SNR values for each subcarrier as seen in equation (2). By construction, ZF equalization completely removes inter-stream interference. Gaussianity is preserved in the effective noise vector since it results from the linear combination of the elements in the thermal noise vector. Therefore, the procedure for computing the FER for MIMO-OFDM systems with BPSK constellations is essentially the same with ZF equalizers, although the notation becomes more complex as the interleaving function is replaced by a spatial parsing function s. Here we assume a general spatial parsing function s. In practice this function interleaves bits over spatial streams and subcarriers such that adjacent bits are placed randomly in space and frequency.

Assuming a general spatial parsing function with $L \geq KN_{\text{O}}N/\mathcal{C}$, let $s : \{0, 1, \ldots, KN_{\text{O}}N/\mathcal{C} - 1\} \rightarrow \{1, 2, \ldots, N_s\} \times \{0, 1, \ldots, KN_{\text{O}}N/(N_s\mathcal{C}) - 1\}$ represent the spatial parser in Fig. 1. Furthermore $s_1(n) \in \{0, 1, \ldots, N_s\}$ is equal to the spatial stream that the coded bit indexed by n is mapped to. Likewise $s_2(n) \in \{1, 2, \ldots, KN_{\text{O}}N/(N_s\mathcal{C})\}$ is equal to the parsed bit index in spatial stream $s_1(n)$. Fig. 9 demonstrates the operation of these functions. Using similar arguments as before, we conclude that the FER bound is

$$\text{FER}(\cdot) < \sum_{d=d_f}^{\infty} \sum_{\mathbf{c} \in \mathsf{C}_d} \prod_{l=1}^{d} Z\left(\gamma\left[s_1(\ell_l^{(\mathbf{c})}), \text{mod}\left(s_2(\ell_l^{(\mathbf{c})}), N\right)\right]\right) \tag{15}$$

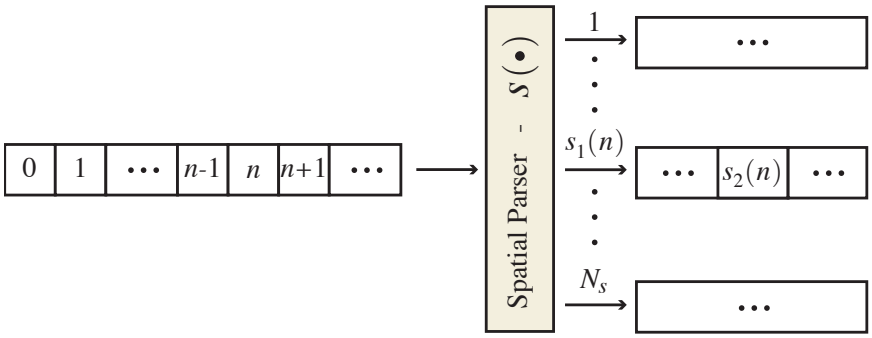

Fig. 9. Spatial parsing function notation

for MIMO-OFDM systems in block fading channels with BPSK constellations. Much like with SISO-OFDM, the frame error rate for MIMO-OFDM is a function of the convolutional encoder, the channels per subcarrier, as well as the spatial parser. The FER for MIMO-OFDM, however is somewhat different because of the spatial mapping algorithm and the presence of $N_s N$ fading channels (in contrast with N fading channels for SISO-OFDM). For the remainder of this paper no explicit distinction is made between FER bounds for SISO-OFDM and MIMO-OFDM systems by representing the SISO system as a special case of the MIMO system with $N_s = 1$.

C. FER Approximation through Subcarrier Ordering

In (15) the probability of weight d codeword error given $\boldsymbol{\gamma} \triangleq \{\{\gamma[a,n]\}_{a=1}^{N_s}\}_{n=0}^{N-1}$ for codewords $\mathbf{c} \in \mathsf{C}_d$ is determined by the product

$$Z_d^{(\mathbf{c})}(s,\boldsymbol{\gamma}) \triangleq \prod_{l=1}^{d} Z\left(\gamma\left[s_1(\ell_l^{(\mathbf{c})}), \text{mod}\left(s_2(\ell_l^{(\mathbf{c})}), N\right)\right]\right). \tag{16}$$

This expression is parameterized only by the SNR at distinct subcarriers. Selection of the precise subcarriers of interest to determine codeword error probability is intrinsically linked to the realization of s, the spatial parsing function. Past investigation in [39] showed that the error event probability for weight d codewords is dominated by weight d codeword error corresponding to the maximum value of $Z_d^{(\mathbf{c})}(s,\boldsymbol{\gamma})$. In other words

$$\sum_{\mathbf{c}\in\mathsf{C}_d} Z_d^{(\mathbf{c})}(s,\boldsymbol{\gamma}) \approx \max_{\mathbf{c}\in\mathsf{C}_d}\left\{Z_d^{(\mathbf{c})}(s,\boldsymbol{\gamma})\right\} \tag{17}$$

which holds especially well for large values of d. Practical interleavers process L bits consecutively at a time to limit the complexity and processing delay for the interleaving process. Past

research on the truncation length with Viterbi decoding has shown only T_b consecutive bits are necessary to capture nearly all error events that occur in practice. Assuming $L \leq N_s N$ and $T_b \leq L$, which is sufficient for the convolutional codes in IEEE 802.11n and IEEE 802.16e with generators $(171, 133)_8$ [48], $\max_{\mathbf{c}\in\mathsf{C}_d}\left\{Z_d^{(\mathbf{c})}(s,\boldsymbol{\gamma})\right\}$ can be approximated with $\overline{Z}_d$ given

$$\overline{Z}_d \triangleq \max_{\substack{a_1,a_2,\ldots,a_d \\ n_1,n_2,\ldots,n_d}} \left\{\prod_{l=1}^{d} Z\left(\gamma[a_l, n_l]\right)\right\}. \tag{18}$$

where $a_l \in \{1, 2, \ldots, N_s\}$, $n_l \in \{0, 1, \ldots, N-1\}$ and $(a_l, n_l) \neq (a_\alpha, n_\alpha)$ for $\alpha \neq l$. The above approximation only allows $d \leq d_L \leq L$ where d_L is the maximum codeword weight in L-length error events. Moreover, the truncation length approximation requires restriction to codewords that result in errors events of length at most L, defined by $\mathsf{C}_{d,L}$, when the zero codeword is transmitted. The approximation in (18) is only valid if the operation of the spatial parsing function s adequately randomizes the location of bits in space and frequency. Approximation (18) also does not hold well for frequency flat channels since many or all terms of the summation in (17) have similar values. Assuming (15)-(18) we approximate

$$\text{FER}(\cdot) \approx \sum_{d=d_f}^{d_L} \sum_{\mathbf{c}\in\mathsf{C}_{d,L}} \prod_{l=1}^{d} Z\left(\gamma^{(l)}\right) \tag{19}$$

by observing the tightness of Viterbi bound at FER values of interest [17]. This approximation is completely parametrized by the set $\{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(d_L)}\} \subseteq \{\gamma^{(1)}, \gamma^{(2)}, \ldots, \gamma^{(L)}\}$ for fixed convolutional coding and spatial parsing operations.

REFERENCES

[1] R.C. Daniels, C.M. Caramanis, and R.W. Heath, Jr., "A supervised learning approach to adaptation in practical MIMO-OFDM wireless systems," in *Proceedings of the IEEE Global Communications Conference*. 2008, IEEE.

[2] M. Zorzi, R.R. Rao, and L.B. Milstein, "ARQ error control for fading mobile radio channels," *Vehicular Technology, IEEE Transactions on*, vol. 46, no. 2, pp. 445–455, May 1997.

[3] A. Iera, A. Molinaro, G. Ruggeri, and D. Tripodi, "Improving QoS and throughput in single- and multihop WLANs through dynamic traffic prioritization," *Network, IEEE*, vol. 19, no. 4, pp. 35–44, July-Aug. 2005.

[4] Huai-Rong Shao, H. Singh, and Chiu Ngo, "MAC-enabling technologies for high-throughput wireless LAN," *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, vol. 1, pp. 173–177, Jan. 2006.

[5] Andrea Goldsmith, *Wireless Communications*, Cambridge University Press, August 2005.

[6] Y.W. Blankenship, P.J. Sartori, B.K. Classon, V. Desai, and K.L. Baum, "Link error prediction methods for multicarrier systems," *Proceedings of the IEEE Vehicular Technology Conference, Fall*, vol. 6, pp. 4175–4179 Vol. 6, Sept. 2004.

[7] M. Lamarca and F. Rey, "Indicators for PER prediction in wireless systems: A comparative study," *Proceedings of the IEEE Vehicular Technology Conference, Spring*, vol. 2, pp. 792–796 Vol. 2, May-1 June 2005.

[8] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 4, pp. 2306–2311 Vol. 4, Sept. 2005.

[9] Sébastien Simoens, Stéphanie Rouquette-Léveil, Philippe Sartori, Yufei Blankenship, and Brian Classon, "Error prediction for adaptive modulation and coding in multiple-antenna OFDM systems," *Signal Processing*, vol. 86, no. 8, pp. 1911–1919, 2006.

[10] S. Catreux, V. Erceg, D. Gesbert, and R.W. Heath, Jr., "Adaptive modulation and MIMO coding for broadband wireless data networks," *IEEE Communications Magazine*, vol. 40, no. 6, pp. 108–115, Jun 2002.

[11] Nee, Richard, Jones, V., Awater, Geert, Zelst, Allert, Gardner, James, Steele, and Greg, "The 802.11n MIMO-OFDM standard for wireless LAN and beyond," *Wireless Personal Communications*, vol. 37, no. 3-4, pp. 445–453, May 2006.

[12] Jeffrey G. Andrews, Arunabha Ghosh, and Rias Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking (Prentice Hall Communications Engineering and Emerging Technologies Series)*, Prentice Hall PTR, March 2007.

[13] Erik Dahlman, Stefan Parkvall, Johan Skold, and Per Beming, *3G Evolution*, Academic Press, July 2007.

[14] Chang Kyung Sung, Sae-Young Chung, Jun Heo, and Inkyu Lee, "Adaptive bit-interleaved coded OFDM with reduced feedback information," *Communications, IEEE Transactions on*, vol. 55, no. 9, pp. 1649–1655, Sept. 2007.

[15] Fei Peng, Jinyun Zhang, and W. E. Ryan, "Adaptive Modulation and Coding for IEEE 802.11n," in *Proceedings of the IEEE Wireless Communications and Networking Conference*, Kowloon, Mar.11–15, 2007, pp. 656–661.

[16] Jialing Li and A. Stefanov, "Exact pairwise error probability for block-fading MIMO OFDM systems," *Vehicular Technology, IEEE Transactions on*, vol. 57, no. 4, pp. 2607–2611, July 2008.

[17] Rolf Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, Wiley-IEEE Press, 1999.

[18] M.R. McKay and I.B. Collings, "Capacity and performance of MIMO-BICM with zero-forcing receivers," *IEEE Transactions on Communications*, vol. 53, no. 1, pp. 74–83, Jan. 2005.

[19] Behzad Razavi, *RF microelectronics*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1998.

[20] Ethem Alpaydin, *Introduction to Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, October 2004.

[21] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, Springer, August 2001.

[22] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer, August 2006.

[23] M. Lampe, H. Rohling, and W. Zirwas, "Misunderstandings about link adaptation for frequency selective fading channels," *Proceeedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, pp. 710–714 vol.2, 15-18 Sept. 2002.

[24] S. Tsai and A. Soong, "Effective-SNR mapping for modeling frame error rates in multiple-state channels," in *3GPP Standardization Document*. 3GPP2, 2003, number 3GPP-C30-20030429-010.

[25] Ericsson, "System-level evaluation of OFDM - further considerations," in *3GPP Contributions*. TSG-RAN WG1, November 2003, number 35 R1-031303.

[26] P.H. Tan, Y. Wu, and S. Sun, "Link adaptation based on adaptive modulation and coding for multiple-antenna OFDM system," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 8, pp. 1599–1606, October 2008.

[27] Yang-Seok Choi and S. Alamouti, "A pragmatic phy abstraction technique for link adaptation and mimo switching," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 6, pp. 960–971, August 2008.

[28] S. Kant and T. L. Jensen, "Fast link adaptation for IEEE 802.11n," M.S. thesis, Aalborg University, February 2007.

[29] TGn Channel Models Special Committee, "TGn channel models," IEEE 802.11-03/940r1, November 2003.

[30] Yunqian Ma, "Improving wireless link delivery ratio classification with packet SNR," in *Proceedings of the IEEE International Conference on Electro Information Technology*, May22–25, 2005.

[31] M.A. Haleem and R. Chandramouli, "Adaptive stochastic iterative rate selection for wireless channels," *Communications Letters, IEEE*, vol. 8, no. 5, pp. 292–294, May 2004.

[32] A. Misra, V. Krishnamurthy, and R. Schober, "Stochastic Learning Algorithms for Adaptive Modulation," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, Toulouse, May14–19, 2006, vol. 4.

[33] D. Falconer, S.L. Ariyavisitakul, A. Benyamin-Seeyar, and B. Eidson, "Frequency domain equalization for single-carrier broadband wireless systems," *IEEE Communications Magazine*, vol. 40, no. 4, pp. 58–66, Apr 2002.

[34] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.

[35] R.W. Heath, Jr., S. Sandhu, and A. Paulraj, "Antenna selection for spatial multiplexing systems with linear receivers," *IEEE Communications Letters*, vol. 5, no. 4, pp. 142–144, Apr 2001.

[36] A.A. D'Amico and M. Morelli, "Joint Tx-Rx MMSE design for MIMO multicarrier systems with Tomlinson-Harashima precoding," *Wireless Communications, IEEE Transactions on*, vol. 7, no. 8, pp. 3118–3127, August 2008.

[37] Theodore Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2001.

[38] Rogerio C. Manso, "Performance analysis of M-QAM with Viterbi soft-decision decoding," M.S. thesis, Naval Postgraduate School, Monterey, CA, March 2003.

[39] S. Nanda and K.M. Rege, "Frame error rates for convolutional codes on fading channels and the concept of effective Eb/N0," *IEEE Transactions on Vehicular Technology*, vol. 47, no. 4, pp. 1245–1250, Nov 1998.

[40] D. Seethaler, G. Matz, and F. Hlawatsch, "An efficient MMSE-based demodulator for MIMO bit-interleaved coded modulation," *Global Telecommunications Conference, 2004. GLOBECOM '04. IEEE*, vol. 4, pp. 2455–2459 Vol.4, Nov.-3 Dec. 2004.

[41] Tom M. Mitchell, *Machine Learning*, McGraw-Hill Science/Engineering/Math, March 1997.

[42] Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification (2nd Edition)*, Wiley-Interscience, November 2000.

[43] Kevin S. Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft, "When is 'nearest neighbor' meaningful?," in *Proceedings of the 7th International Conference on Database Theory*, London, UK, 1999, pp. 217–235, Springer-Verlag.

[44] Richard E. Bellman, *Dynamic programming*, Princeton University Press, 1957.

[45] IEEE 802.11n Working Group, *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications - Draft* 5.0*: Enhancements for Higher Throughput*, part 11 standard edition, 2007.

[46] Hao Liu, Liyu Cai, Hongwei Yang, and Dong Li, "EESM based link error prediction for adaptive MIMO-OFDM system," *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 559–563, April 2007.

[47] A. Viterbi, "Convolutional codes and their performance in communication systems," *IEEE Transactions on Communications*, vol. 19, no. 5, pp. 751–772, Oct 1971.

[48] I.M. Onyszchuk, "Truncation length for Viterbi decoding," *IEEE Transactions on Communications*, vol. 39, no. 7, pp. 1023–1026, Jul 1991.