
Author final version (often known as postprint)

Link to published version (if available): 10.1109/PIMRC.2010.5671652

Link to publication record in Explore Bristol Research

PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: http://www.bristol.ac.uk/pure/about/ebr-terms.html

Take down policy

Explore Bristol Research is a digital archive and the intention is that deposited content should not be removed. However, if you believe that this version of the work breaches copyright law please contact open-access@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline of the nature of the complaint

On receipt of your message the Open Access Team will immediately investigate your claim, make an initial judgement of the validity of the claim and, where appropriate, withdraw the item in question from public view.
Unitary Beamforming Techniques Enhancing VoIP Call Admittance Capacity

Marios Nicolaou, Angela Doufexi, Simon Armour and Yong Sun

Department of Electrical & Electronic Engineering,
University of Bristol, Bristol, United Kingdom,
*Toshiba Research Labs Europe
M.Nicolaou@bristol.ac.uk

Abstract—Fourth Generation Networks will invariably adopt MIMO-OFDMA techniques, in order to cope with increased data rate demands and improved Quality of Service (QoS) requirements. OFDMA is a popular multiple access candidate, since it facilitates multi-user diversity by enabling two dimensional multiple access, in time and frequency domains. Random Beamforming enables exploitation of spatial multi-user diversity and spatial capacity multiplexing gain. Layered Random Beamforming is able to achieve further multiple access benefits by enabling different users to be scheduled on different spatial layers of the same time-frequency resource. This paper examines the QoS performance of SU-MIMO and MU-MIMO precoding schemes for a VoIP system under a realistic traffic and channel environment. A novel Link Adaptation algorithm operating under PER constraints for delay sensitive VoIP traffic that promises increased data rates, flexibility and improved QoS is proposed and examined in this paper.

I. INTRODUCTION

The emergence of new wireless communications systems in recent years that require support for a variety of applications (e.g. video streaming, voice, data etc.) has drastically changed the scope and design principles of wireless systems. The need for higher data rates in conjunction with improved QoS control has been identified by both the IEEE 802.16 committee [1] and Long Term Evolution (LTE) of the Third Generation Partnership Project (3GPP) [2]. In order to achieve these requirements, MIMO and OFDMA have been identified as the most promising PHY and multiple access candidates in the downlink [3].

In a multi-user environment, better spectrum utilisation can be achieved by a channel aware scheduling process, assigning resources to users according to their instantaneous channel strength in time and/or frequency. Opportunistic Beamforming (OB), proposed in [4], applies a randomly generated beamforming pattern at the transmitter to effectively exploit this multi-user diversity (MUD) in combination with transmit beamforming. Only the Signal to Noise Ratio (SNR) is required for uplink feedback information rather than full Channel State Information (CSI). MIMO specific precoding schemes have been specially proposed for the 3GPP-LTE standard, in the form of unitary codebook beamforming [5]. Two modes of operation based on the spatial resource allocation capability are defined: Single-User MIMO (SU-MIMO) and Multi-User MIMO (MU-MIMO). Analysis in [6] has highlighted the notable performance benefits in the joint exploitation of spatial multiplexing gain as well as spatial and spectral multiuser diversity gain achieved via the MU-MIMO approach.

Support for QoS is a major part of the WiMAX and LTE MAC layer design. QoS control is maintained by a connection-oriented MAC architecture, where a serving Base Station (BS) controls all the uplink (UL) and downlink (DL) connections. A number of classes with different QoS requirements are specified for WiMAX. These include the following: Unsolicited grant service (UGS), Real-time polling service (rtPS), Non-real time polling service (nrtPS), Best Effort (BE) and Extended real-time variable rate (ERT-VR) service. This paper is concerned with servicing Voice over IP (VoIP) service, which is supported by either UGS or ERT-VR. Real time VoIP traffic has tight packet delay and additional Packet Error Rate (PER) constraints. LTE adopts a priority based scheme based on QoS Class identifiers which can be directly related to the WiMAX QoS classes.

A delay oriented Link Adaptive (LA) scheme that incorporates an upper bound PER constraint (in order to avoid excessive delays arising from the retransmission of erroneous packets) is adopted. The LA approach considered in [7], whereby the non-deterministic nature of fading channels is accounted for, is extended to accurately represent multiuser channel statistics. This approach promises improved mode prediction accuracy over the conventional PHY mode selection approach based on the average SNR, achieving both increased throughput as well as QoS efficiency.

The remainder of the paper is organised as follows: Section II describes the PHY model of SU-MIMO and MU-MIMO OFDMA. Section III discusses the particular characteristics of VoIP traffic. The performance analysis of the SU-MIMO and MU-MIMO for VoIP is presented in Section IV. Section V discusses the LA approach and presents QoS results for VoIP. The paper concludes in Section V.

II. PHYSICAL LAYER MODEL

Unitary codebook based beamforming is capable of achieving spatial multiuser diversity gain and spatial multiplexing gain [8]. Unitary codebook based beamforming suggests that a predefined set of antenna beams is used which ensure a uniform sector coverage. With considerably lower uplink feedback requirements than the conventional eigenbeamforming approach, the transmitter applies a predefined set of precoding matrices to the transmitted signal and selects the best user based on the feedback of effective signal-to-interference and noise ratio (ESINR) indicating channel quality for each of the precoding matrices along with its corresponding index.

The proposed pre-coder design relies on the Fourier basis. The resolution of these beams is dependent on the overall codebook size. The codebook E, consists of the unitary matrix set, i.e. $\tilde{E}_{\tilde{g}} = \{\tilde{y}_{\tilde{g}}^{(0)}, \ldots, \tilde{y}_{\tilde{g}}^{(G-1)}\}$, where $\tilde{y}_{\tilde{g}}^{(s)} = \left[\tilde{y}_{\tilde{g},0}^{(s)} \ldots \tilde{y}_{\tilde{g},M-1}^{(s)}\right]$
is the $g$-th precoding matrix, and $v_m^{(s)}$ is the $m$-th precoding vector in the set. According to the Fourier basis,\
\[
v_m^{(s)} = \frac{1}{\sqrt{M}} \left[ v_0^{(s)} \ldots v_{M-1}^{(s)} \right]^T
\]

The same unitary matrix $V_k^s \times$ is applied to all users across all the sub-carriers of the same frequency resource. Different matrices are used across different frequency resources. The received signal after FFT and guard interval removal becomes (time index $t$ is omitted):
\[
Y_k,s = H_k,s V_k^s X_s + N_k,s
\]

where $k$ denotes a user index, $s$ denotes a sub-carrier index, $(\cdot)^T$ denotes the Hermitian function and $H_k,s$ is a matrix containing user $k$’s frequency responses of the channels between $M$ transmit and $N$ receive antennas at sub-carrier $s$. $D_{k,s}$ is a diagonal matrix including all the singular values of $H_k,s$ and $U_{k,s}$ and $V_{k,s}$ are the unitary matrices obtained by applying Singular Value Decomposition (SVD) to $H_k,s$. $X_s$ denotes an $M \times 1$ matrix containing the transmit signals at sub-carrier $s$ at the BS and $N_{k,s}$ represents the additive complex Gaussian noise with zero mean and variance $(\sigma_{k,s})^2$. If $V_k^s$ is equal to $V_k$, the $V_k^s V_k^s^H$ term becomes an identity matrix and the user $k$ is said to be in the true eigenbeamforming configuration.

The system adopts a linear MMSE receiver, which has interference suppression capability. For a 2x2 MIMO system, the MIMO channels can be decomposed into 2 separate spatial layers. The received signal $Y_k,s$ is multiplied by the MMSE filter $G_{k,s}$:
\[
G_{k,s} = \left( H_k,s V_k^s \right)^H \left( H_k,s V_k^s + SNR^{-1} I \right)^{-1} \left( H_k,s V_k^s \right)^H
\]

For data stream $q$ at sub-carrier $s$, the user $k$ computes the $\text{ESINR}_k^q$:
\[
\text{ESINR}_k^q = \frac{E_s}{\left| A_{k,s} \right|^2 \sigma_{k,s}^2} - 1
\]

where $A_{k,s} = \left( H_k,s V_k^s \right)^H \left( H_k,s V_k^s + SNR^{-1} I \right)^{-1}$. $E_s$ denotes the average symbol energy and $(\cdot)$ indicates the element located in row $q$ and column $j$.

Unitary codebook based beamforming defines two modes of operation. In SU-MIMO OFDMA, users are selected according to their average $\text{ESINR}$ across the spatial layers. Each user then occupies all spatial layers of a frequency resource in conjunction with frequency division multiple access (different users can be assigned across the different frequency resources over the frequency domain). A more efficient spectrum utilisation can be achieved via MU-MIMO, whereby the spatial dimension is exploited by incorporating the capability of allocating different spatial layers of the same frequency resource to different users. MU-MIMO, effectively a form of Spatial Division Multiple Access (SDMA), is expected to achieve a greater overall system performance gain than SU-MIMO.

In a SU-MIMO system, for a frequency resource $c$, denoting the index of the starting sub-carrier by $b$ and the finishing sub-carrier by $a$, the average rate of user $k$ is given by:
\[
R_k = \frac{1}{a-b+1} \sum_{j=b}^{a} \sum_{q=1}^{Q} \log_2 \left( 1 + \text{ESINR}_k^q \right)
\]

The BS allocates each frequency resource to a user according to the selected resource allocation algorithm.

For the MU-MIMO scheme, different spatial layers can be allocated to different users to realise an additional spatial multiuser diversity gain. However, this approach increases the feedback by the number of spatial layers (minimum of the number of transmit and receive antennas) compared to the SU-MIMO scheme. Based on the ESINR calculated from eq. (4), the user $k$ calculates the rate of each spatial layer $q$ on a frequency resource basis:
\[
R_k^q = \frac{1}{a-b+1} \sum_{j=b}^{a} \log_2 \left( 1 + \text{ESINR}_k^q \right)
\]

For every frequency resource, the BS allocates transmission resources on a per spatial layer basis.

Subsequent simulations assume a codebook size, $G=2$ for the precoding techniques, returning a good tradeoff between downlink spectral efficiency and uplink feedback overhead.

III. VoIP TRAFFIC MODELLING

A. Traffic Characteristics

VoIP traffic can be transmitted as a UGS or ERT-VR service. UGS uses a fixed amount of bandwidth for the duration of the call, with no periodic bandwidth request or polling service. Therefore, the MAC overhead and uplink access delay can be minimised. However transmitting VoIP through UGS is not bandwidth efficient as a fixed amount of bandwidth is allocated continuously even during the silent period of the VoIP call. When transmitting VoIP through ERT-VR, voice activity detection (VAD) [9] enables a more efficient BW utilisation, as resources are assigned for only the duration of talk spurts. The ERT-VR VoIP transmission with VAD can be modelled as a variable bit rate (VBR) system with active and silent periods.

A two state Markov process is used to model VoIP, representing the active voice and silent periods respectively [10]. The alternating periods of activity and silence are exponentially distributed with average durations $1/\alpha$ ($\beta$ being the parameter of the exponential law of the active period) and $1/\beta$ respectively ($\alpha$ being the parameter of the exponential law of the inactive period). Each VoIP session can be either in the active or inactive state. During the active state, fixed sized packets of 32 bytes are generated at a constant interval of 20ms [10].

Figure 1: Two-state Markov chain for VoIP activity transitions

In Table 1 a summary of the traffic distribution parameters for VoIP is presented. The maximum tolerable packet delay for VoIP is set to 30ms. Packets that exceed this delay
constraint are assumed to time-out and hence are removed from the buffer queue. QoS outage for VoIP is registered for users whose average packet timeout ratio over a session exceeds 4% [10]. The aggregate system’s QoS performance is measured in terms of the satisfied customer ratio (i.e. users not exceeding the max. allowable drop ratio with respect to the total number of admitted VoIP users). On a per-user basis, the average packet drop ratio is used to quantify perceived QoS.

The effect of PER in QoS has been extensively studied in [11]. A high PER can result in significant degradation of the perceived speech quality due to the increased number of required packet retransmissions. An upper PER bound of 10^-7 is set for the VoIP system under consideration to avoid excessive delays due to packet re-transmissions. Packet losses at the physical layer (PHY) are assumed negligible compared to timeouts. Hence they are not taken into account for the packet timeout ratio calculation.

Table 1: VoIP Traffic Distribution Parameters

<table>
<thead>
<tr>
<th>Component</th>
<th>Distribution</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active state duration (1/µ)</td>
<td>Exponential</td>
<td>Mean =0.4 s</td>
</tr>
<tr>
<td>Inactive state duration (1/α)</td>
<td>Exponential</td>
<td>Mean =0.6 s</td>
</tr>
<tr>
<td>Packet Inter-arrival rate within a burst</td>
<td>Fixed</td>
<td>τ = 5/packet/s</td>
</tr>
<tr>
<td>Probability of transition from active to inactive state</td>
<td>N/A</td>
<td>μ=0.6</td>
</tr>
<tr>
<td>Probability of transition from inactive to active state</td>
<td>N/A</td>
<td>λ=0.4</td>
</tr>
</tbody>
</table>

B. Packet Scheduler Structure

A key component in providing QoS lies at the MAC layer, where the packet scheduler operates. A QoS oriented packet scheduler proposed in [12] is adopted in this paper. The packet scheduler, operating at the BS, consists of three distinct components: the Packet Classifier (PC), the Buffer Management Block (BMB) and the Packet Scheduler (PS). The PC classifies incoming packets from different users into QoS classes based on their user ID and forwards them to the BMB. For the particular VoIP scenario, it is assumed that an independent buffer management policy exists exclusively for VoIP traffic. The PS transmits packets to users according to their priority obtained from the channel status of each user.

VoIP packets arriving at the buffer remain stored until the scheduler assigns resources for the transmission of each packet, or until they reach their maximum corresponding lifetime in which case they are dropped from the queue. Real-time, VoIP traffic incorporates stringent packet delay constraints and thus the maximum queuing time \( T_q \) must be constrained. Additionally, once a packet has been selected for transmission, a finite time \( T_p \) associated with the transmission duration of the packet has to be incorporated in the VoIP packet latency. A maximum tolerable packet delay of 30ms for VoIP is assumed in subsequent simulations. Once the Head-of-Line (HOL) packet has been scheduled for transmission, a number of resources are assigned to the MT according to the adopted scheduling strategy. If adequate throughput is assigned to a packet over the tolerable packet lifetime, a successful transmission of this packet is registered. If a packet fails to acquire the required throughput over the remaining lifetime, the packet times out and all throughput that has been allocated for this packet will be lost. Therefore, excessive packet timeouts not only result in a degradation of the perceived QoS but also give rise to poor utilisation of the available radio resources. An upper bound for tolerable packet timeout ratio is set at 4% [10]. Users exceeding this ratio are assumed to be in QoS outage.

IV. PERFORMANCE ANALYSIS OF RANDOM AND LAYERED RANDOM BEAMFORMING

In order to evaluate the performance of a LA system employing SU-MIMO and MU-MIMO in combination with OFDMA and serving VoIP traffic, both numerical and simulated performance is presented in this section. All the OFDMA systems considered require uplink channel quality information in the form of quantised ESINR on every physical resource block.

A. System Setup and Channel Model

The key parameters used in the simulations are presented in Table 2. 6 PHY operating modes with different modulation and coding rate combinations are considered, as shown in Table 3, in accordance with [13]. The link throughput is calculated according to:

\[
LT = B \left(1 - \sum (PER)\right) \approx B(1 - PER) \tag{7}
\]

\( B \) being the number of encoded data bits per OFDM symbol and \( r \) the number of required re-transmissions.

Table 2: Parameters for the Proposed OFDMA System

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFT size</td>
<td>1024</td>
</tr>
<tr>
<td>Useful Subcarriers</td>
<td>768</td>
</tr>
<tr>
<td>Guard Interval Length</td>
<td>176</td>
</tr>
<tr>
<td>Subcarrier Frequency Spacing</td>
<td>10.94 KHz</td>
</tr>
<tr>
<td>Useful Symbol Duration</td>
<td>102.9 µs</td>
</tr>
<tr>
<td>MAC Frame Duration</td>
<td>5 ms</td>
</tr>
<tr>
<td>Channel coding</td>
<td>Punctured 1/2 rate, convolutional code, constraint length (7, [133,171] ) actual</td>
</tr>
</tbody>
</table>

Table 3: Transmission Modes

<table>
<thead>
<tr>
<th>Mode</th>
<th>Modulation</th>
<th>Coding Rate</th>
<th>Coded Bits per OFDM symbol</th>
<th>Data Bits per OFDM symbol (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BPSK</td>
<td>1/2</td>
<td>768</td>
<td>384</td>
</tr>
<tr>
<td>2</td>
<td>QPSK</td>
<td>1/2</td>
<td>1536</td>
<td>768</td>
</tr>
<tr>
<td>3</td>
<td>QPSK</td>
<td>3/4</td>
<td>1536</td>
<td>1152</td>
</tr>
<tr>
<td>4</td>
<td>16 QAM</td>
<td>1/2</td>
<td>3072</td>
<td>1536</td>
</tr>
<tr>
<td>5</td>
<td>16 QAM</td>
<td>3/4</td>
<td>3072</td>
<td>2304</td>
</tr>
<tr>
<td>6</td>
<td>64 QAM</td>
<td>3/4</td>
<td>4608</td>
<td>3456</td>
</tr>
</tbody>
</table>

An urban micro channel model is used based on the spatial channel model (SCM) [14], developed by ETSI 3GPP-3GPP2 to help standardise the outdoor evaluation of mobile systems.

B. Numerical Performance of unitary codebook beamforming

Figure 2 shows the theoretical spectral efficiency of SU-MIMO-OFDMA and MU-MIMO-OFDMA employing either the ESINR or ESINR metric using a greedy, rate maximisation scheduling as a function of SNR for K=20 users. The average throughput of SU-MIMO and MU-MIMO is calculated according to equations 5 and 6 respectively. For the MU-MIMO case, the average throughput is obtained by the sum of
the data rates across all spatial layers. MU-MIMO achieves a higher data rate than SU-MIMO as a result of the additional spatial layer MUD gain. As was shown in [15], the layered approach can even outperform the SVD OFDMA capacity bound.

Figure 2: Theoretical Spectral Efficiency for SU-MIMO and MU-MIMO

C. Simulation Performance of SU-MIMO and MU-MIMO-OFDMA

In order to verify the numerical analysis, Figure 3 shows the simulated PER performance for SU-MIMO and MU-MIMO-OFDMA for the PHY modes of Table 3, assuming a total number of users, K = 20.

A notable PER improvement is observed for MU-MIMO with regards to SU-MIMO. MU-MIMO optimises performance independently on each spatial layer, rather than improving the aggregate rate across all spatial layers, consequently giving rise to this physical layer performance improvement.

Figure 3: PER performance for a) SU-MIMO-OFDMA, b) MU-MIMO-OFDMA under different PHY modes

Both numerical analysis and simulation have indicated the benefits in MU-MIMO in terms of throughput. Additionally, the exploitation of the spatial diversity gives rise to improved resource allocation fairness, due to the increased number of available radio resources distributed amongst a fixed number of MSs. Therefore, it is expected that a LA system employing MU-MIMO would be able to accommodate more users within a given bandwidth due to enhanced fairness and also provide PHY performance benefits due to a higher order diversity exploitation. PER curves indicate that a lower SNR can be tolerated for MU-MIMO for the same packet error constraints with respect to SU-MIMO. This can either be translated into a higher call admittance capacity, greater cell coverage or reduced power requirements [16]. The following section considers a link-level simulation of the considered VoIP system examining the degree to which QoS can be maintained as a function of the total system load. A Link Adaptation strategy, appropriate for VoIP, returning high mode selection accuracy and providing flexibility against fading and changes to the channel environment has been considered.

V. QoS PERFORMANCE WITH LINK ADAPTATION FOR VoIP

A. Proposed Delay Oriented Link Adaptation

Received signal quality in wireless systems is affected by various channel factors, namely the distance of the MS from the serving BS, path loss exponent, log-normal shadowing, fast Rayleigh fading and noise [17]. Systems that do not employ a LA strategy are constrained to use a fixed transmission scheme designed to maintain acceptable performance for worst case (or near worst case) conditions, resulting in a poor spectrum utilisation. By employing LA, the transmitted parameters can be altered to optimise performance under varying channel conditions.

Wireless channels often suffer from frequency selective fading as a consequence of the physical factors affecting transmissions. Studies in [7, 18] have shown that the conventional notion of performing LA based on the average PER performance fails, since the stochastic nature of individual fading channel realisations can vary considerably from the average, a result indicated in Figure 4 which shows that the instantaneous PER performance of a given fading realisation can vary by more than 10dB from the average.

Figure 4: PERs for individual channel samples and average PER, 16-QAM, ¾ rate

In [19] it has been argued that the causes of bit errors can be broadly attributed to the average received channel strength and the probability of resource allocation on fading instances. A zero-norm fading indicator metric is adopted according to eq. (8). As an indication of resource allocation on fading instances, the ratio of the frequency channel responses with gains lower than the average received response is adopted.

$$I_{fad} = \frac{1}{N_{SC}} \sum_{i} F(i)$$  \hspace{1cm} (8)

where $F(i) = 1$ if $|H_i| < |\overline{H}|$ and $F(i) = 0$ otherwise. $|\overline{H}|$ corresponds to the average response across only the assigned resources of each user.

A high $|\overline{H}|$ value and a low number of deep fades describe a channel with a low likelihood of bit errors. By mapping each fading realisation to a corresponding fading indicator...
value a prediction on the required SNR, for which a target PER requirement is met, can be made. Figure 5 shows that the relation between the fading indicator and the required SNR can be approximated with a high degree of accuracy by a cubic function. By generating these functions offline for each of the PHY modes, the proposed LA method can directly determine the optimum MCS mode based on the SINR feedback information. Any change in channel loading conditions, scheduling algorithm or channel environment will be directly reflected in the fading indicator value (a decrease for example would indicate worsening channel conditions). Provided, however, that the fading indicator value is mapped to the cubic function, the proposed LA remains immune to such variations. The mode selection efficiency for the proposed LA system has been measured to be in the range of 98%. The conventional LA approach achieves only a 28% efficiency [20].

VoIP traffic is characterised by asynchronous call arrivals and periodic packet arrivals during the active state as indicated in Table 1. Additionally, queued packets at the buffer add to the network loading. Therefore, the number of users simultaneously requesting service is variable. Unlike the conventional LA strategy that relies on a full buffer model, the fading indicator approach is independent of the number of users, or the scheduling approach. Any change in these parameters will be directly reflected on the fading indicator value, which can be directly mapped to a required SNR value for a target PER.

Figure 5: Fading indicator values for a target PER=10^{-3}, 16QAM, ½ rate

B. Channel Environment

A SCM urban micro path loss model for Line-of Sight (LOS) urban micro scenario based on the Walfish-Ikegami-Model (COST-W1) is used in the following simulations [21]. A 2x2 MIMO antenna configuration is assumed. Users are randomly distributed within the cell, with a minimum range of 75m from the BS to the cell edge at 800m. A BS transmit power of 43dBm is assumed. Receiver Sensitivity is set to -120dBm.

C. QoS Results with Link Adaptation for SU-MIMO and MU-MIMO

As discussed in the previous section, MU-MIMO provides significant benefits over SU-MIMO in terms of throughput. The exploitation of spatial diversity can also improve the overall system fairness [4], since the likelihood of a user acquiring service increases due to the effective doubling of the available resources (for a 2x2 MIMO system). This section presents link-level simulation results under different channel loading conditions measuring the perceived QoS quality of each user independently based on the average packet timeout ratio and the aggregate QoS based on the satisfied customer ratio metric.

Figure 6 (a) examines the average latency of VoIP packets as a function of the total number of VoIP users. The MU-MIMO approach allows for significantly lower delays. This observation can be attributed to the properties of MU-MIMO, which provides a more timely service to users, due to the increased number of available radio resources and the enhanced resource allocation efficiency this approach provides. A direct correlation between packet latencies and packet timeout ratios is observed, by examining the findings of Figure 6 (b), which presents the average timeout ratio of the two precoding schemes as a function of the number of users. As the network becomes more congested, the resulting scarcity of available resources gives rise to increased buffer waiting times. It should be noted that since packets tend to remain at the buffer longer for SU-MIMO, the actual number of simultaneous service requests is likely to be considerably higher, despite assuming the same number of users with active VoIP sessions.

Figure 6 (c) shows the satisfied customer ratio, defined as the ratio of the total users not experiencing QoS outage, for SU-MIMO and MU-MIMO. It can be observed that the maximum number of VoIP calls that can be admitted with no QoS outage is significantly higher for MU-MIMO than for SU-MIMO. This observation is a direct consequence of the reduced packet waiting times achieved by MU-MIMO. An overall reduction in packet waiting times can be translated in an overall reduction in packet timeouts which consequently reflects to the overall QoS satisfaction ratio. In order to ensure high QoS satisfaction for SU-MIMO under consideration, more bandwidth should be allocated for VoIP traffic.

Figure 7: Cumulative Distribution Function of Packet delays
Figure 7 shows the cumulative distribution function of packet delays arising from SU-MIMO and MU-MIMO for a total number of users, K=20 and K=40. Results verify that the average packet delay is directly related to the traffic load and the degree of diversity exploitation. However, an additional result that can be extracted is that of a fairer distribution of packet delays arising from the adoption of MU-MIMO precoding with respect to SU-MIMO. This result is particularly important since it proves that users experiencing severe path losses and shadowing effects are not underserved compared to users experiencing stronger channel conditions, ensuring an overall fair QoS satisfaction across the network.

VI. CONCLUSIONS

This paper has examined the QoS performance of two MIMO precoding schemes proposed for 3GPP-LTE, in a VoIP traffic system under a realistic channel environment. A novel Link Adaptation strategy that considers the stochastic variability of fading channels has also been proposed. This approach has shown improved mode selection accuracy and higher flexibility towards dealing with the variable traffic loads arising from VoIP traffic. Tight delay constraints associated with real-time traffic can lower the call admittance capacity of a system. Simulation results in this paper have shown that efficient exploitation of multiuser diversity across the spectral and spatial domain in form of MU-MIMO allows for notable increases at the call admittance capacity of the system under specified bandwidth constraints without compromising QoS. In addition, a more uniform distribution of packet delays across users experiencing distinctly diverse channel conditions has been observed. These benefits have been attributed to the increased data rates and improved resource allocation, attained by MU-MIMO.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the financial support of EPSRC and of Toshiba Research Europe Limited (TREL).

REFERENCES