Multi-Antenna Techniques for Evolved 3G Wireless Communication Networks: An Overview

Michael Mao Wang, Andy Wang, Alexei Gorokhov, Tamer Kadous, and Min Dong
Corporate Research & Development, Qualcomm, San Diego, USA
Email: wangm@qualcomm.com

***Abstract*—Multi-antenna techniques offers significant capacity benefits. But realizing these gains in practical wireless systems remains a very challenging problem. MIMO-OFDM has been used in wireless LAN systems with great success. However, due to the largely varying spatial profiles and velocity patterns of the access terminals, integrating MIMO techniques in a wireless WAN system has been a difficult design problem. This paper overviews the multi-antenna techniques used in evolved 3G wireless communications systems, such as 3GPP2 Ultra-Mobile Broadband (UMB), IEEE Mobile Broadband Wireless Access (MBWA) and 3GPP LTE, which are poised to become the successor of 3G systems, *e.g.*, 3GPP2 1xEVDO and WCDMA, as the next generation of high speed wireless system. This paper uses MIMO in various forms to suit the deployment scenario. Although we use UMB as a design paradigm in this paper, most of the results apply to OFDM systems in general.**

***Index Terms*—OFDMA, Multi-antenna transmission, MIMO, spatial multiplexing, rank predication, SDMA, precoding, MMSE receiver, MMSE-SIC receiver, Ultra Mobile Broadband (UMB).**

I. Introduction

Orthogonal Frequency Division Multiple Access (OFDMA) and multiple input multiple output (MIMO) have been among the most promising techniques for wireless communications to achieve high data rates and many wireless industry standard bodies are investigating the MIMO-OFDM schemes which are the most efficient in a wireless wide area network, *e.g.*, a cellular network, environments [1]. This paper provides a technical overview of the multi-antenna techniques employed by the emerging evolved 3G wireless technologies such as 3GPP2 Ultra-Mobile Broadband (UMB) [3], IEEE Mobile Broadband Wireless Access (MBWA) [4], 3GPP LTE [8], WiMAX [9], and wireless LANs [10], for enhanced system performance. For the ease of discussion, we will use UMB as a design paradigm in the following discussion. But the design applies to any OFDM systems in general.

The design objectives of UMB [3]-[7] are to provide performance enhancements over existing cellular systems, while maintaining competitive edge over current 3G systems (*e.g.*, 3GPP WCDMA, 3GPP2 EVDO, *etc.*). The areas of performance improvements include the introduction of higher peak data rates, better spectral efficiency, lower latency, improved terminal battery life, higher capacity, and enhanced user experience for delay-sensitive applications.

While multiple antenna technologies offer many promising advantages over conventional systems, they also pose a set of unique challenges that a practical MIMO-OFDM based cellular network must be able to overcome. The rest of the paper is devoted to describing these challenges and the solutions that the evolved 3G standards offer. Note that the algorithms presented in the rest of the paper are only used to illustrate the concept and are not necessarily the recommended algorithms. In Section II, we review the basic MIMO spatial multiplexing structures, the limitations and the rank prediction solution. In Section III, we further discuss the solution to the basic MIMO limitations, *i.e.*, the precoding technique. Section IV deals with multi-user MIMO or spatial division multiple access (SDMA). Section V discusses MIMO pilot design. Finally, Section VI concludes this paper.

II. MIMO Spatial Multiplexing

Spatial multiplexing offer a linear increase in data rate through multiple-input multiple-output systems, *i.e.*, transmitting multiple, independent data layers (streams) within the bandwidth of operation [11][12][13]. Spatial multiplexing gain can be realized by simply transmitting data layers from each of the transmit antennas, thus maximizing the average data rate over the MIMO system. Assume that a vector of M independent modulation symbols $\mathbf{s} \in \mathbb{C}^{M_{\mathrm{T}}}$ is transmitted over the $M_{\mathrm{T}} \times M_{\mathrm{R}}$ MIMO system where M_{T} is the number of transmit antennae and M_{R} the number of receive antennae. The receiver obtains $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}$ where $\mathbf{y} \in \mathbb{C}^{M_{\mathrm{R}}}$, $\mathbf{H} \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{T}}}$ is the channel matrix, and $\mathbf{n} \in \mathbb{C}^{M_{\mathrm{R}}}$ is the AWGN noise. The multiple transmitted data layers interfere with one another at the receiver. The optimum symbol detection method is ML where the receiver compares all possible combination of symbols which may have been transmitted with what is observed, *i.e.*, $\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|$. However, the complexity of ML detection grows exponentially with the number of data streams. A linear minimum mean-square error (MMSE) receiver is commonly used to decouple the data layers from the received signals as shown in Fig. 1.

A. Linear MMSE Receiver

We let the number of data layer be M where $1 \le M \le \min\{M_{\mathrm{R}}, M_{\mathrm{T}}\}$. For simplicity we assume for

© 2009 ACADEMY PUBLISHER

now $M = M_{\mathrm{T}} = M_{\mathrm{R}}$. Denote

$$\mathbf{H}(k) = \left[\mathbf{h}_1(k) \quad \mathbf{h}_2(k) \quad \cdots \quad \mathbf{h}_{M_{\mathrm{T}}}(k)\right] \in \mathbb{C}^{M_{\mathrm{R}} \times M_{\mathrm{T}}} \tag{1}$$

the $M_{\mathrm{R}} \times M_{\mathrm{T}}$ channel matrix, k the OFDM subcarrier index, $\mathbf{s}(k) \in \mathbb{C}^{M}$ the normalized symbol vector modulated onto the kth subcarrier. The received signal $\mathbf{y}(k) \in \mathbb{C}^{M_{\mathrm{R}}}$ for each OFDM subcarrier k can be written as

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{U}\mathbf{s}(k) + \mathbf{n}(k) = \mathbf{G}(k)\mathbf{s}(k) + \mathbf{n}(k) \tag{2}$$

where $\mathbf{U}(k) \in \mathbb{C}^{M_{\mathrm{T}} \times M_{\mathrm{T}}}$ is the $M_{\mathrm{T}} \times M_{\mathrm{T}}$ spatial multiplexing (or precoding) matrix, $\mathbf{G}(k) \in \mathbb{C}^{M_{\mathrm{R}} \times M}$ is the effective channel matrix

$$\mathbf{G}(k) \triangleq \mathbf{H}(k)\mathbf{U} = \left[\mathbf{g}_1(k) \quad \mathbf{g}_2(k) \quad \cdots \quad \mathbf{g}_M(k)\right], \tag{3}$$

and $\mathbf{n}(k) \in \mathbb{C}^{M_R}$ is AWGN with $E\left\{\mathbf{n}(k)\mathbf{n}^H(k)\right\} = \sigma^2(k)\mathbf{I}$. For spatial multiplexing, $\mathbf{U}(k)$ can be a random unitary matric or simply an $M_{\mathrm{T}} \times M_{\mathrm{T}}$ diagonal matrix, *i.e.*, M_{T} independent layers are transmitted from each of the transmit antennas.

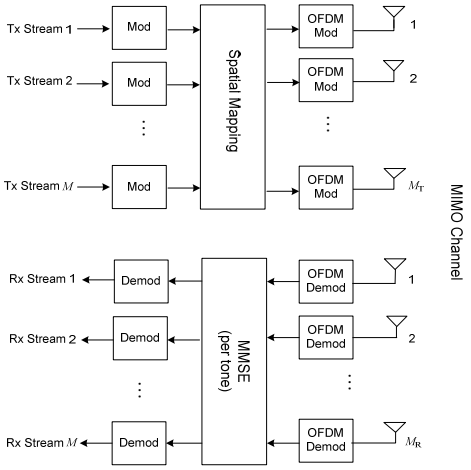

Fig. 1. MIMO spatial multiplexing diagram.

The MMSE equalizer can be shown to be

$$\mathbf{W}(k) = \left(\hat{\mathbf{G}}^H(k)\hat{\mathbf{G}}(k) + \sigma^2(k)\mathbf{I}\right)^{-1} \hat{\mathbf{G}}_k^H(k) \tag{4}$$

where $\mathbf{W}(k) = \left[\mathbf{w}_1(k) \quad \mathbf{w}_2(k) \quad \cdots \quad \mathbf{w}_M(k)\right] \in \mathbb{C}^{M_R \times M}$. The recovered M layer modulation symbol vector is

$$\hat{\mathbf{s}}(k) = \mathbf{W}^H(k)\mathbf{y}(k) \tag{5}$$

with the SNR, conditional on the channel matrix, for the mth layer [18]

$$\hat{\gamma}_m(k) = \frac{1}{\sigma^2(k)\left[\left(\hat{\mathbf{G}}^H(k)\hat{\mathbf{G}}(k) + \sigma^2(k)\mathbf{I}\right)^{-1}\right]_{m,m}} - 1 \tag{6}$$

B. Multi Code Word and Single Code Word

There are typically two main MIMO modes, namely, single code word (SCW) transmission mode and multiple code word (MCW) transmission mode. In a MCW transmission mode, the data stream on each transmit spatial layer is *independently* encoded with different spectral efficiencies. In a SCW transmission mode, the encoded packet is de-multiplexed and transmitted across all spatial layers with *identical* spectral efficiency.

In the MCW transmission mode (Fig. 2), M ($1 \le M \le \min\{M_{\mathrm{R}}, M_{\mathrm{T}}\}$) data packets are transmitted in parallel. Each of the M packets, denoted by m ($1 \le m \le M$), is individually turbo/LDPC encoded and QAM modulated using the m^{th} selected packet format via rate predication. The M layers are then mapped to the physical antennas and transmitted.

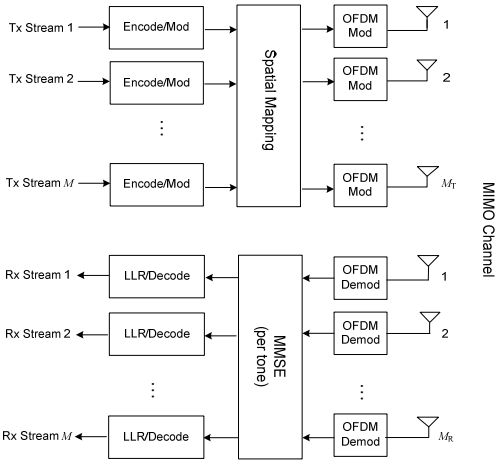

Fig. 2. Illustration of multiple codeword MIMO.

To improve the receiver decoding performance, instead of decoding each layer independently, a successive interference cancellation receiver [17] with linear MMSE equalizer (MMSE-SIC) can be used to decouple the incoming layers. The receiver attempts to decode the first layer. The linear MMSE equalizer generates the soft estimates of the modulation symbols corresponding to the first layer

$$\hat{\mathbf{s}}_1(k) = \mathbf{w}_1^H \mathbf{y}(k), \; k = 1, 2, \cdots, K \tag{7}$$

and all the K subcarriers in the user's assignment. The soft estimates are used to compute the log-likelihood ratios (LLRs) that are fed to the turbo/LDPC decoder. If the first layer is decoded successfully (*i.e.*, it passes the cyclic redundancy check (CRC)), the receiver regenerates a clean version of the modulation symbols corresponding to the first layer, multiples each modulation symbol by the corresponding channel coefficient, and subtracts the contribution of the first layer from the received signal. The receiver then applies the MMSE equalizer on the receive data with the first layer interference removed

$$\hat{\mathbf{s}}_2(k) = \mathbf{w}_2^H \left(\mathbf{y}(k) - \hat{\mathbf{g}}_1 \mathbf{s}_1(k)\right), \; k = 1, 2, \cdots, K \tag{8}$$

and attempts to decode the second layer and repeats the same process

$$\hat{\mathbf{s}}_m(k) = \mathbf{w}_m^H \left(\mathbf{y}(k) - \sum_{i=1}^{m-1} \hat{\mathbf{g}}_i \mathbf{s}_i(k)\right), \; k = 1, 2, \cdots, K \, . \tag{9}$$

© 2009 ACADEMY PUBLISHER

If one of the layers, m', fails to decode, the receiver stops decoding and sends an ACK indicating the layers that have been successfully decoded. As a result, the access network transmit redundancy information only on the un-decoded layers, $m', \cdots, M_{\mathrm{T}}$. For synchronous HARQ (which is the case for UMB), within the maximum number of HARQ transmissions, no new packets are transmitted on the successfully decoded layers. Total power is reallocated only to the outstanding layers.

In the MCW transmission mode, the channel quality (CQI), or differential CQI, is fed back to the transmitter for each layer. That is, rate prediction is done per layer.

In the SCW transmission (Fig. 3), a single packet of data is encoded and sent over different antennae. The input data stream is turbo/LDPC encoded and QAM modulated to produce the transmit symbols according to the packet format specified by the rate prediction algorithm. The stream of modulation symbols is then de-multiplexed to M layers (parallel sub-streams). The layers or sub-streams are spatially mapped to the M transmit antennas. An MMSE receiver is used to decouple the M layers and multiplex them back to one data stream and demodulated and decoded.

The SCW scheme has simple HARQ identical to SISO, simple ACK/NACK messaging identical to SISO, and low complexity MMSE receiver. The drawback of the SCW (with a linear MMSE receiver) is that it is not capacity achieving and suffers from a throughput loss in low rank channels, such as, spatially correlated channels or LOS channels with high Rician K-factor.

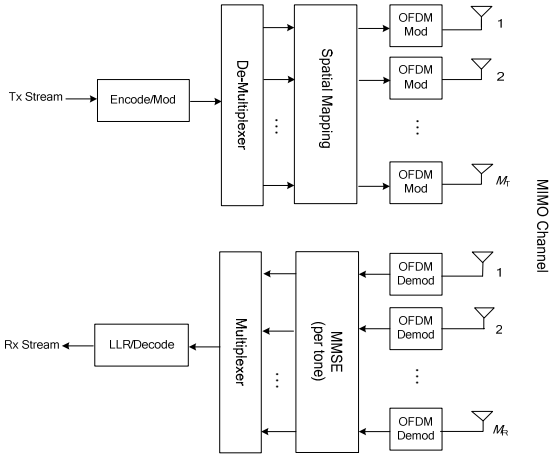

Fig. 3. Illustration of single code word MIMO.

However, the MCW transmission with a successive interference cancellation (SIC) receiver is capacity achieving and hence optimal in performance. A successive interference cancellation (SIC) receiver is used to decouple the M layers providing higher throughput and more tolerance to spatial correlation. However, MCW with SIC comes at the cost of increased signaling overhead, receiver complexity and memory requirements. The signaling overhead for the reverse link and forward link control channels are larger for the MCW as compared to an SCW transmission since the channel quality CQI, coding and modulation, and acknowledgements have to be signaled for each MIMO layer. Furthermore, the SIC receiver memory requirements are high since the MIMO channel and received signals have to be stored for all HARQ transmissions. The receiver processing is more complicated and bursty since lower layers can not be decoded until the upper layers are decoded.

C. Rank Prediction

To reduce the inter-layer interference, especially in SCW transmission, the spatial layers M can not be higher than the channel rank, *i.e.*, trading off inter-layer interference with spatial multiplexing gains. This can be achieved via rank prediction, where the receiver feeds back the optimal rank, in the sense of maximum capacity or spectral efficiency in addition to the channel quality information (CQI).

For each subcarrier, *k*, the effective MIMO channel matrices, corresponding to different ranks $\mathbf{G}_M(k) = \hat{\mathbf{H}}(k)\mathbf{U}_M(k)$ are computed, where $\mathbf{U}_M \in \mathbb{C}^{M_{\mathrm{T}} \times M}$ is the first M columns of the spatial multiplexing matrix $\mathbf{U}$ and $1 \le M \le M_{\mathrm{T}}$.

Assuming a linear MMSE receiver, the SNR for each rank and layer is calculated according to (6) as

$$\gamma_{M,m}(k) = \frac{1}{\sigma^2 \left[\left(\mathbf{G}_M^H(k)\mathbf{G}_M(k) + \sigma^2(k)\mathbf{I} \right)^{-1} \right]_{m,m}} - 1 \quad (10)$$

where $1 \le m \le M,\ 1 \le M \le M_{\mathrm{T}}$. A modulation-constrained capacity mapping ϕ is then applied to the above SNR to generate the corresponding capacity and averaged across subcarriers

$$C_{M,m} = \frac{1}{N_{\mathrm{P}}} \sum_{k=1}^{N_{\mathrm{P}}} \phi\left(\gamma_{M,m}(k)\right) \quad (11)$$

to generate the average capacity of layer $m = 1, 2 \cdots, M$ for MCW mode where N_{P} is the total number of channel estimates from the common MIMO pilots (Section V).

For SCW, the layer capacity is further averaged across all layers to obtain the overall average capacity

$$C_M = \frac{1}{M} \sum_{m=1}^{M} C_{M,m} \,. \quad (12)$$

The optimal rank is chosen so as to maximize the overall capacity

$$\hat{M} = \underset{1 \le M \le \min\{M_{\mathrm{T}}, M_{\mathrm{R}}\}}{\arg\max} \left\{ M C_M \right\} \quad (13)$$

For MCW, the capacity mapping is then used to generate the *effective* SNR for an AWGN channel per subcarrier of the *m*th layer $m = 1, 2 \cdots, M$

$$\gamma_{\hat{M},m} = \phi^{-1}\left(C_{\hat{M},m}\right), \quad m = 1, 2 \cdots, M \quad (14)$$

which is used as the CQI for layer *m*. The *M* CQI values (each quantized to 4 bits in UMB) are together fed back to

© 2009 ACADEMY PUBLISHER

the transmitter. For SCW, the effective AWGN SNR (per layer and per subcarrier) corresponding to the maximum capacity $C_{\hat{M}}$ is

$$\gamma_{\hat{M}} = \phi^{-1}\left(C_{\hat{M}}\right) \tag{15}$$

which is communicated back to the transmitter as the SCW CQI. The rank and CQIs are made available at the transmitter via the reverse link control channel [22] feedback every 5 msec.

In the case of $M < M_{\mathrm{T}}$, the extra transmit antennas can be used for spatial diversity gain.

For both SCW and MCW transmission, both the packet format and the rank of the M layers are adapted to the MIMO channel. The receiver runs a rank prediction algorithm by which it determines the value of M to be used. The receiver also computes the M CQI values, one for each layer, in MCW mode, but only one averaged CQI over the layers in SCW mode, and feeds them back to the transmitter. The transmitter adjusts the transmission power on each layer, based on the power control loop and rank, and runs a rate prediction algorithm by which it chooses the packet format for each layer. Rank predication brings SCW performance closer to that of MCW.

Fig. 4 shows the typical rank distribution where less than 1% of the users are able to use rank 4. Most users have either rank 1 or 2.

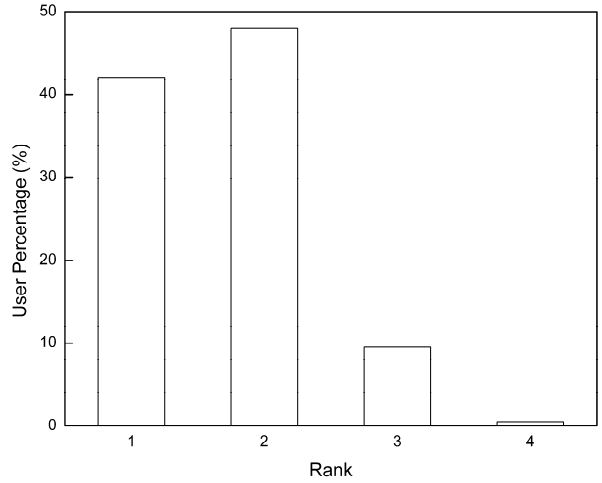

Fig. 4. Rank distribution in a 4×4 MIMO system.

III. PRECODING

As we know, MIMO spatial multiplexing provides substantial gain at high SNR but performs poorly at low SNR. Precoding [14][15][16] provides an effective closed-loop MIMO for array (direction) gain. Array gain is the increase in receive SNR that results from a coherent combining effect of the wireless signals at a receiver. Precoding is especially beneficial for unbalanced MIMO system, *i.e.*, $M_{\mathrm{R}} < M_{\mathrm{T}}$.

Consider a linear precoding structure $\mathbf{x} = \mathbf{U}_M\mathbf{s}$ where $\mathbf{U}_M \in \mathbb{C}^{M_{\mathrm{T}}\times M}$ is an $M_{\mathrm{T}}\times M$ unitary matrix derived from a pre-specified precoding codebook, that is, $\mathbf{\Omega} \triangleq \left\{\mathbf{U}_j, j = 1,2,\cdots,2^J\right\}$, between the access network and the access terminal. Typically, $\mathbf{\Omega}$ is a codebook for precoding matrix of size $M_{\mathrm{T}}\times M_{\mathrm{T}}$. A precoding matrix $\mathbf{U}_{j,M}$ of size $M_{\mathrm{T}}\times M$ is obtained by taking the first M columns of the full size $M_{\mathrm{T}}\times M_{\mathrm{T}}$ matrix $\mathbf{U}_j \in \mathbf{\Omega}$. For the beamforming case, $\mathbf{U}_{j,M}$ is simply an $M_{\mathrm{T}}\times1$ vector. In UMB, precoding codebooks are configurable and can be adapted to specific antenna topology and/or propagation conditions.

The receiver observes a channel realization and selects the best precoding matrix in the codebook. The precoding matrix index (J bits), together with the rank and CQI, is fed back to the transmitter. The size of the precoding codebook is limited by the bit width J of the precoding matrix index communicated to the access network.

In precoding design, a key issue is the trade-off between precoding gains and the feedback overhead. The overhead is primarily due to the need to capture frequency as well as space selectivity. Fig. 5 shows the trade-off between performance and feedback overhead. In UMB, J is selected to be 6 and, therefore, the largest size of the precoding codebook can be 64. Due to the frequency selectivity, the optimal precoding matrix can vary from frequency to frequency. Ideally, therefore, an optimal precoding matrix index should be fed back to the access network for each frequency. However, in UMB, the optimal precoding matrix is selected over the entire transmission bandwidth or a sub-band to avoid increased feedback overhead, increased sensitivity to channel time variations as well as increased estimation errors.

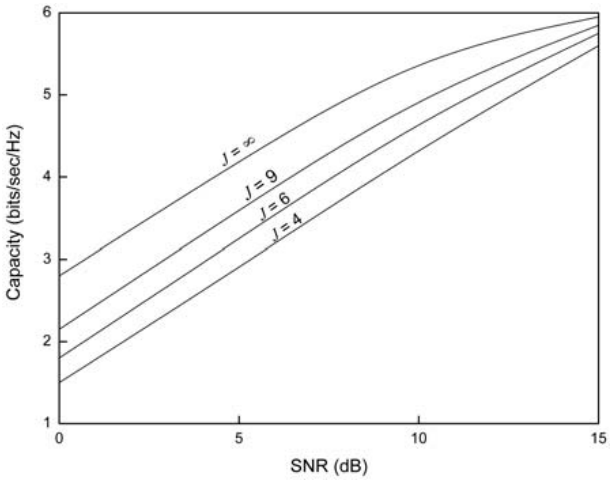

Fig. 5. Effect of number of precoding matrix index bits on performance (64 QAM constellation, 8×1, flat fading).

For each precoding matrix, $\mathbf{U}_j \in \mathbf{\Omega}$, at each subcarrier, k, the effective MIMO channel matrices corresponding to different ranks $1 \le M \le M_{\mathrm{T}}$ are computed at the receiver

$$\mathbf{G}_{j,M}(k) = \mathbf{H}(k)\mathbf{U}_{j,M}. \tag{16}$$

The SNR for each rank and layer are calculated as

© 2009 ACADEMY PUBLISHER

$$\gamma_{j,M,m}(k)=\frac{1}{\sigma^2(k)\left[\left(\mathbf{G}_{j,M}^H(k)\mathbf{G}_{j,M}(k)+\sigma^2(k)\mathbf{I}\right)^{-1}\right]_{m,m}}-1 \quad (17)$$

The average capacity over layer and subcarrier is computed as

$$C_{j,M}=\frac{1}{MN_{\mathrm{P}}}\sum_{m=1}^{M}\sum_{k=1}^{N_{\mathrm{P}}}\phi\left(\gamma_{j,M,m}(k)\right) \quad (18)$$

The best rank and its associated precoding matrix is chosen so as to maximize the overall spectral efficiency

$$\hat{M},\hat{j}=\underset{1\le j\le 2^J,\ 1\le M\le M_{\mathrm{T}}}{\arg\max}\left\{MC_{j,M}\right\} \quad (19)$$

The effective SNR per subcarrier per layer, or overall CQI, corresponding to the maximum spectral efficiency $C_{\hat{M},\hat{j}}$ is

$$\gamma_{\hat{j},\hat{M}}=\phi^{-1}\left(C_{\hat{j},\hat{M}}\right). \quad (20)$$

The effective SNR per subcarrier, or CQI, of each layer is

$$\gamma_{\hat{j},\hat{M},m}=\phi^{-1}\left(\frac{1}{P}\sum_{k=1}^{P}\phi\left(\gamma_{\hat{j},\hat{M},m}(k)\right)\right),\ 1\le m\le\hat{M} \quad (21)$$

The preferred precoding matrix index $\hat{j}$, together with rank $\hat{M}$ and CQI, $\gamma_{\hat{j},\hat{M}}$ for SCW and $\gamma_{\hat{j},\hat{M},m}$, $m=1,2,\cdots,\hat{M}$ for MCW, are fed back to the Access Network.

Fig. 6 shows the precoding gains for SCW and MCW modes for 4×2 and 4×4 MIMO transmissions. We see that precoding provides larger gain for SCW than MCW and larger gain for unbalanced MIMO than balanced MIMO. For balanced MIMO MCW, the precoding gain is limited.

Fig. 7 shows the beamforming gain for cell edge (*i.e.*, geometry = 0 dB) users via precoding in an FDD system.

IV. Space Division Multiple Access (SDMA)

In multi-user communication systems, the system throughput can be increased by simultaneous transmission to/from several users in the same time-frequency slot by means of spatial division multiple access (SDMA), or multi-user MIMO, via the use of multiple antennas [19][20]. This section describes the SDMA techniques for both forward and reverse links.

A. Forward Link SDMA

SDMA on the forward link is a multi-antenna transmission technique where multiple users are signaled on the same time-frequency resources. SDMA opens up a new dimension at the expense of reduced signal to interference ratio. When multiple transmit antennas are available at the access network, they can be used to mitigate intra-sector interference in SDMA. This is achieved by transmitting simultaneously to the overlapping users using properly defined beams for each user. Clearly these beams are dependent on the spatial channels of the overlapping users. The intra-sector interference in turn is dependent on the beam of the overlapping users. If the user channels or spatial signatures are similar, the beams used to transmit to the users are likely to be similar, resulting in high intra-cell interference. User grouping or clustering plays an important role in mitigating intra-sector interference and realizing SDMA gain, *i.e.*, overlapping users should have sufficiently different spatial signatures.

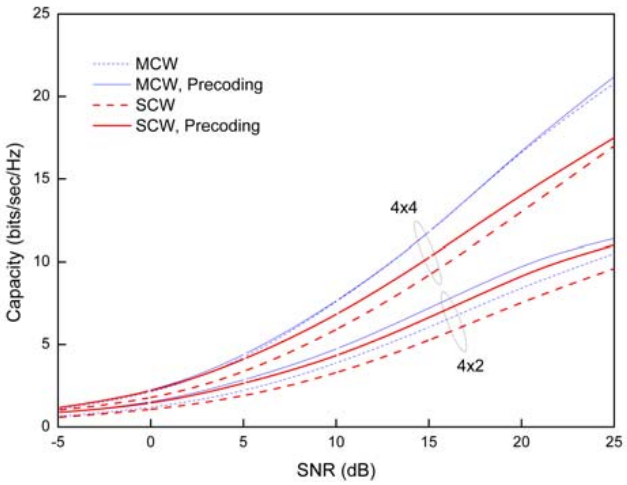

Fig. 6. Precoding gain (64 QAM constellation, PedB channel model).

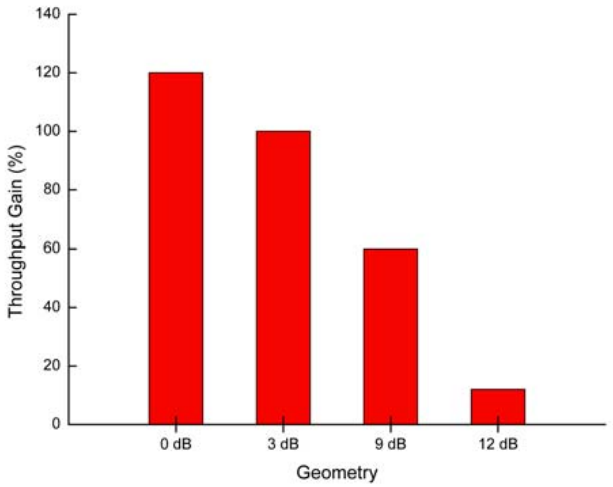

Fig. 7. Beamforming via precoding for cell edge performance improvement.

Since the access terminal or the user clustering is based on the forward link channel, for FDD, the clustering has to be performed based on information provided by a reverse link feedback mechanism. The same mechanism defined for precoding can be used for SDMA.

The precoding matrices in the codebook for SDMA operation are grouped into clusters. In this case, matrices of a single cluster, $\boldsymbol{\omega}\subset\boldsymbol{\Omega}$, typically span only part of the space. The columns of the matrices in different clusters are used to form spatial beams covering spatially distinct groups of users. If the access terminal feeds back a beam index within a cluster, the access network treats this as an indication that it may schedule other access terminals on different clusters, *i.e.*, allowing for SDMA.

© 2009 ACADEMY PUBLISHER

In SDMA, the access terminal finds the preferred precoding matrix $\mathbf{U}_{\hat{j},\hat{M}}$ that is contained within a particular SDMA cluster $\boldsymbol{\omega}_{\hat{j}}$ as well as the CQI $\gamma_{\hat{j},\hat{M}}$ associated with the beam, similar to precoding. In addition, the access terminal calculates the intra-sector interference, the interference from the potential SDMA users. The intra-sector interference can be estimated as

$$\begin{aligned}\sigma^2_{\text{SDMA},q}(k) &= \frac{1}{\hat{M}}\text{trace}\left\{\sum_{\mathbf{U}_q\notin\boldsymbol{\omega}_{\hat{j}}}\mathbf{G}^H_{\hat{M},q}(k)\mathbf{G}_{\hat{M},q}(k)\right\}\\ &= \frac{1}{\hat{M}}\text{trace}\left\{\sum_{\mathbf{U}_q\notin\boldsymbol{\omega}_{\hat{j}}}\mathbf{U}^H_{\hat{M},q}\mathbf{H}^H(k)\mathbf{H}(k)\mathbf{U}_{\hat{M},q}\right\}\end{aligned} \tag{22}$$

where $\mathbf{U}_{\hat{j}}\in\boldsymbol{\omega}_{\hat{j}}$. The precoding matrix (index) associated with the largest interference in each cluster $\boldsymbol{\omega}_i\subset\boldsymbol{\Omega}$ is

$$q_i = \arg\max_{q\in\boldsymbol{\omega}_i}\left\{\sum_{k=1}^{K}\sigma^2_{\text{SDMA},q}(k)\right\}, \tag{23}$$

for all $\boldsymbol{\omega}_i\neq\boldsymbol{\omega}_{\hat{j}}$. These largest interferers from different clusters of the precoding codebook are sorted and the top $Q-1$ largest interferers are used for estimating the SDMA SNR, where Q is the number of SDMA dimensions

$$\sigma^2_{\text{SDMA}}(k) = \sum_{q_i=1}^{Q-1}\sigma^2_{\text{SDMA},q_i}(k). \tag{24}$$

The SNR with SDMA becomes

$$\gamma^{\text{SDMA}}_{\hat{j},\hat{M},m}(k) = \frac{1}{\left(\sigma^2(k)+\sigma^2_{\text{SDMA}}(k)\right)\cdot\dfrac{1}{\left[\left(\mathbf{G}^H_{\hat{j},\hat{M}}(k)\mathbf{G}_{\hat{j},\hat{M}}(k)+\left(\sigma^2(k)+\sigma^2_{\text{SDMA}}(k)\right)\mathbf{I}\right)^{-1}\right]_{m,m}}-1}$$

The effective SNR under SDMA is

$$\gamma^{\text{SDMA}}_{\hat{j},\hat{M}} = \phi^{-1}\left(\frac{1}{\hat{M}K}\sum_{m=1}^{\hat{M}}\sum_{k=1}^{K}\phi\left(\gamma^{\text{SDMA}}_{\hat{j},\hat{M},m}(k)\right)\right). \tag{25}$$

The SDMA channel CQI (the differential CQI, $\gamma_{\hat{M},\hat{j}}-\gamma^{\text{SDMA}}_{\hat{M},\hat{j}}$), in addition to the precoding index and the regular CQI, are reported back to the access network.

All users corresponding to the same SDMA cluster are placed into the same group. Users within a group are scheduled so that they are always orthogonal to each other, *i.e*. they are not allowed to overlap, since the beams within the same SDMA cluster have similar spatial characteristics; therefore, users preferring these beams are also likely to have similar spatial characteristics and should not be overlapped. This is achieved by scheduling users via the SDMA resource tree (Fig. 8).

In UMB, the fundamental resource unit for scheduling/ assignment is the *logical subcarrier*, or, a *hop port* [21]. A logical subcarrier is a static resource that maps to a unique physical subcarrier, where the mapping changes over time (hopping). Sets of logical subcarriers specified using *nodes* on the resource tree. Each *base node* (a node that does not have child nodes) addresses one channel element, a *tile*. A tile consists of hop-port block (16 subcarriers) over a physical frame (8 OFDM symbols). A tile is the minimum resource allocation unit. Taking Fig. 8 for an example, an access terminal assigned node 7 occupies base nodes 15 and 16 or the first two tiles. There are Q SDMA subtrees. The mapping of logical subcarriers to physical subcarriers is the same across all SDMA subtrees. Therefore, logical subcarriers from different SDMA subtrees overlap with each other in the physical subcarrier domain, *i.e*., if the logical subcarrier assigned to a user on one tree are the same as the one assigned to a different user on another tree, they essentially are scheduled on the same frequency resource. The logical subcarriers on the same subtree are orthogonal to each other.

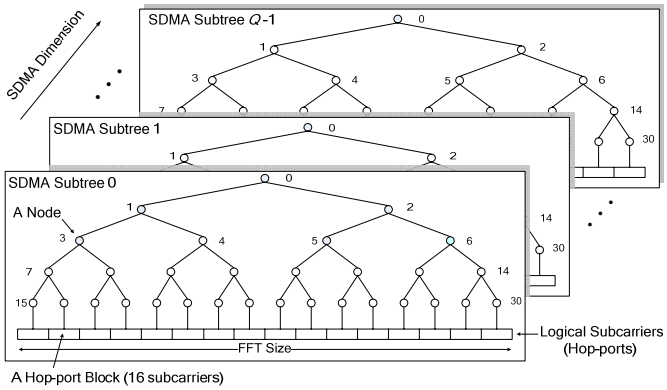

Fig. 8. SDMA channel resource trees (FFT size=256 as shown).

Based on the user group and the differential channel quality reports from different access terminals, the access network selects Q SDMA user groups and a non-SDMA user group. An SDMA subtree is assigned per SDMA group. The SDMA access terminals from different groups thus scheduled are, therefore, superposed to the same channel resources. The logical subcarriers on the primary tree (Subtree 0) allocated to the non-SDMA users are not used in any of the other trees.

Fig. 9 shows the forward link sector throughput gain through SDMA (dimension Q = 2) for $M_{\text{R}}=2$ and 4. More receive antennas clearly provide more gain as a result of a more effective MMSE receiver for interference suppression.

B. Reverse Link SDMA

In an interference-limited multi-access scheme, such as CDMA (*e.g*., the reverse link of 1xEVDO [2]), the SNR gain as a result of the use of multiple receive antennas can be converted to the addition of more users (more CDMA codes). It can be shown that the CDMA sum-rate spectral efficiency is

$$R_{\text{CDMA}} \le \frac{1-\delta^{-1}}{1+\lambda}\frac{M_{\text{R}}}{\Gamma^2}\log_2 e \tag{26}$$

where δ is the rise over thermal, $\Gamma^2\ge 1$ is the CDMA gap to capacity (mainly defined by the codec used), and λ is the ratio of the total inter-sector interference to the total intra-sector interference. It can be seen that the capacity of

© 2009 ACADEMY PUBLISHER

a CDMA system scales linearly with the number of receive antennas.

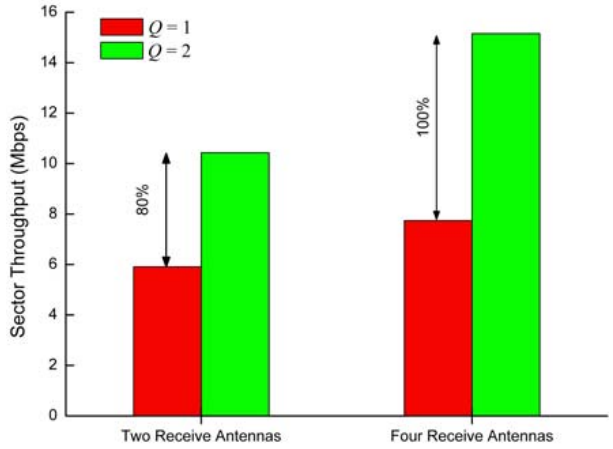

Fig. 9. Forward link sector throughput/5 MHz ($M_T = 4$, $M_R = 2$ and 4, 3km/h Pedestrian B channel, 16 access terminals per sector).

However, like any orthogonal multi-access scheme, OFDMA reverse link suffers from dimension limitation. Its capacity scales logarithmically with the number of receive antennas:

$$R_{\text{OFDMA}} \le \log_2\left(1+\frac{1-\zeta^{-1}}{\beta}\frac{M_R}{\Gamma^2}\right) \qquad (27)$$

where ζ is the interference over thermal (ratio of the average power of the total interference over thermal) and β is the ratio of the average inter-sector interference to the average user power spectral density. This limitation can be improved by allowing spatial multiplexing of users. Reverse link SDMA can be employed to overcome the inherent dimension limitation of OFDM systems.

Unlike the forward link SDMA where implementing SDMA requires to group users on spatial properties of their channel state, one of the reverse link SDMA approach is to collect intra-sector interference diversity through random hopping. That is, Q access terminals are statistically separated with multiple receive antennas M_R at the access point, where $Q \le M_R$, allowing reverse link capacity to scale linearly with number of receive antennas. Similar to forward link SDMA, Q access terminals are assigned overlapped resources, *i.e.*, same time-frequency allocation (but still maintaining orthogonality between pilots (Section V.B)), via SDMA subtrees (Fig. 8), with the exception that the mappings of logical subcarrier to the physical subcarriers for different SDMA subtrees are different. Specifically, access network assignment to each access terminal consists of a set of time-frequency blocks (tiles). Each access terminal overlaps with a set of access terminals on each tile. The set of access terminals are different for different tiles and vary over time, hence, providing co-channel interference diversity. An optimal MMSE receiver at the access network can be employed to recover Q colliding user signals.

Table 1 lists the sector throughput gains through SDMA and Fig. 10 shows the access terminal throughput for Q = 2 against Q = 1. In the simulation, same multiplexing factor is used for all users including those in power limited regime. It is seen that strong users experience significant gains through SDMA. It is not surprising that weak users suffer on the other hand since these users are mostly at the linear capacity region. This problem can be overcome by relying on scheduling policy not to schedule low SNR users for SDMA.

Table 1. Reverse link Sector throughput/5MHz with Q = 1 and Q = 2 ($M_T = 1$, $M_R = 4$).

Channel Model	Sector Throughput (kpbs)		
	$Q = 1$	$Q = 2$	Gain
Pedestrian B (3km/h)	5716	7251	27 %
Vehicular A (30km/h)	5646	6990	24 %

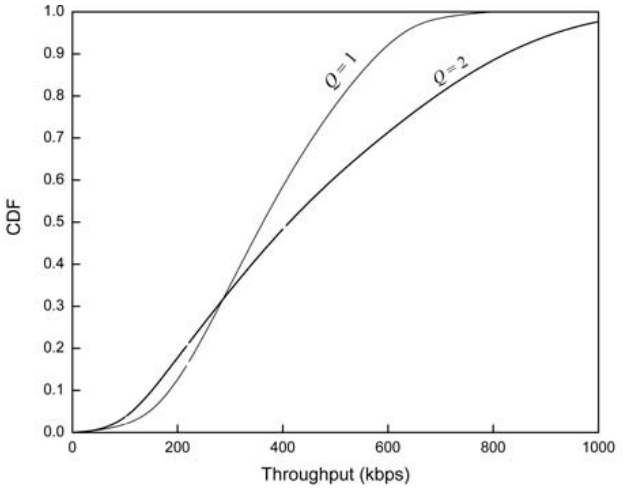

Fig. 10. Reverse link access terminal throughput/5MHz CDFs for Q = 1 and Q = 2 (3km/h Pedestrian B channel, $M_T = 1$, $M_R = 4$, 16 access terminals per sector).

V. MIMO PILOTS

Special pilots are needed for MIMO channel estimation at the receiver. This section is devoted to the MIMO pilot design for both forward and reverse links. For the ease of discussion, we will use UMB MIMO pilots as an example.

A. Forward Link Pilots

The UMB forward link supports Block Resource Channel (BRCH) and Distributed Resource Channel (DRCH) structures [21]. In the BRCH structure, the subcarriers of a tile are mapped to contiguous physical subcarriers and the mapping is constant during a physical frame (8 OFDM symbols). With the DRCH structure, the logical subcarriers of a tile are mapped to a set of regularly spaced subcarriers that are scattered across the entire available bandwidth.

In the DRCH structure, only the common pilot is present (F-CPICH [21]) which provides a wideband reference of the channel across the whole band. As depicted in Fig. 11, F-CPICH consists of a disjoint sets of subcarriers for each of transmit antenna, *i.e.*, the pilot transmitted from one antenna is orthogonal to that from

© 2009 ACADEMY PUBLISHER

the other antennas, providing a means for estimating the channel matrix, $\mathbf{H}$, at the receiver. In this mode, the precoding matrix index, $\hat{j}$, is also transmitted to the receiver for obtaining the effective channel matrix, $\mathbf{G}_{\hat{j}} = \mathbf{H}\mathbf{U}_{\hat{j}}$.

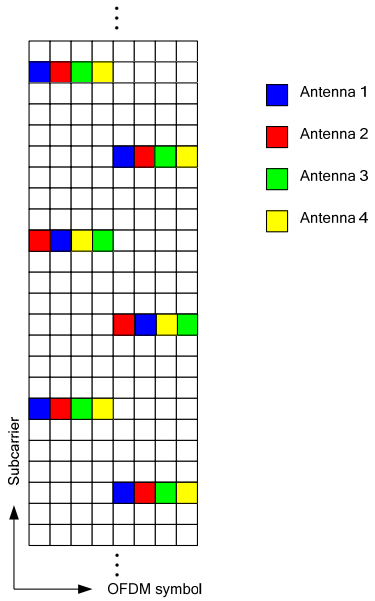

Fig. 11. Illustration of forward common pilot placement ($M_{\mathrm{T}} = 4$).

In the BRCH structure, the dedicated pilot (F-DPICH [21]) is present in a BRCH tile to enable the access terminal to estimate the effective channel matrix, $\mathbf{G}_j$, and interference variance, σ^2.

As shown in Fig. 12, to provide the orthogonality among layer pilots, the dedicated pilot modulation symbols are grouped into clusters, each of which has M_{c} symbols occupying a contiguous region in time. The pilot symbol t in each cluster of the tile associated with the layer m is modulated by an orthonormal sequence

$$s_m(k) = \exp\left(j\left(2\pi/M_c\right)mt\right), \quad 0 \le t \le M_c \qquad (28)$$

where $0 \le m \le M-1$ and $M \le M_c$. In the case of $M < M_c$, the $M_c - M$ dimensions of each cluster can be used for noise/interference estimation. The different pilot formats are designed to accommodate different channel conditions and transmission layers. The pilot format used for transmission is communicated to the access terminal via the assignment message.

The common pilot is also present in the BRCH tile with frame index i satisfying $i \bmod 8 = 4$, (*i.e.*, transmitted in one out of the eight physical frames) as illustrated in Fig. 13, for the purpose of estimation of the channel matrix, $\mathbf{H}$, for the best precoding matrix determination and CQI calculation.

B. Reverse Link Pilots

The dedicated pilots for the reverse link (R-DPICH [22]) have two formats, *i.e.*, Format 0 and Format 1 as depicted in Fig. 12. In SDMA mode, the pilot clusters are used to orthogonally multiplex pilots of different ATs. This multiplexing can be achieved by assigning the orthogonal sequences in (28) to different SDMA channel subtrees, *i.e.*,

$$s_q(k) = \exp\left(j\left(2\pi/M_c\right)qt\right), \quad 0 \le t \le M_c \qquad (29)$$

where $0 \le q \le Q-1$ and $Q \le M_c$ is the reverse link SDMA dimension.

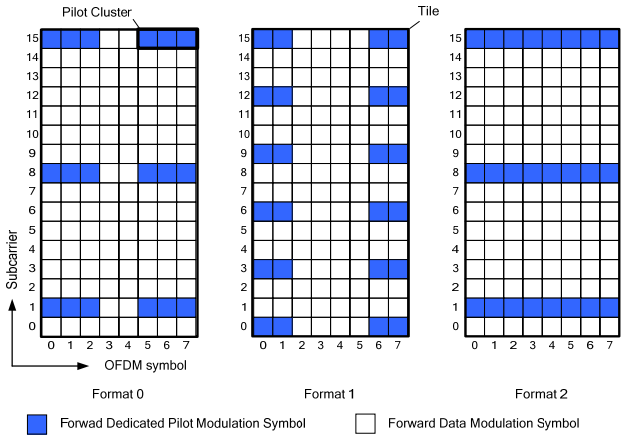

Fig. 12. Illustration of forward dedicated pilot placement within a tile where $M_{\mathrm{c}} = 3$, 2, and 8 for Formats 0, 1, and 2, respectively.

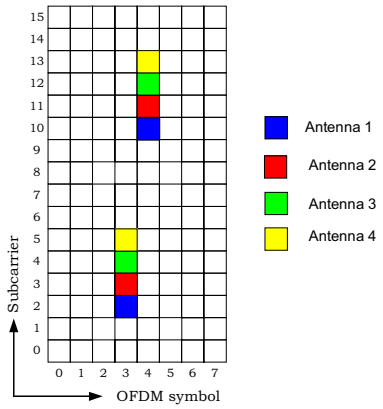

Fig. 13. BRCH mode common pilot placement.

VI. Conclusion

This paper gives an overview of the advanced multiple antenna techniques in evolved 3G wireless communication systems using UMB (FDD) as a design paradigm. The UMB forward link employs multiple antenna techniques such as MIMO and SDMA. For MIMO, both single (SCW) and multiple code word (MCW) are supported. The MCW scheme with MMSE-SIC receivers is capacity achieving in the presence of accurate rate prediction. However, the MCW-SIC design comes at the cost of increased signaling overhead, and complexity and memory requirements at the receiver. On the other hand, the conventional SCW scheme has simple Hybrid-ARQ and simple ACK/NACK messaging identical to that in a SISO system, as well as a

© 2009 ACADEMY PUBLISHER

lower complexity MMSE receiver. However, SCW with a linear MMSE receiver performs poorly compared to MCW for spatially correlated channels as a result of increased inter-layer interference in low rank channels. One way to improve the robustness of spatial multiplexing to rank deficiencies of the channel is to use rank predication that adapts the number of spatial layer to the channel rank to balance the cross-layer interference with MIMO multiplexing gains, resulting in SCW performance approaching MCW with limited increase in overhead. Spatial multiplexing gain is limited at low SNR (*e.g.*, cell-edge users). Precoding is used to improve the performance at low SNR *via* array gain. Precoding is also useful for unbalanced MIMO system, *i.e.*, $M_R < M_T$.

SDMA is another powerful antenna technique used by UMB for both forward and reverse links for increasing system capacity by exploiting another dimension, *i.e.*, the spatial separation of the access terminals. Note that although simultaneously exploiting all the benefits provided by multi-antennas (*e.g.*, spatial multiplexing gain, array gain, spatial diversity gain, and spatial separation) for a particular user may not be possible due to the conflicting demands on the spatial degrees of freedom, the use of some of the combination of the multi-antenna benefits throughout the network results in improved capacity, coverage and reliability [23].

REFERENCES

[1] H Bolcskei, "MIMO-OFDM wireless systems: basics, perspectives, and challenges," *IEEE Wireless Communications*, vol.12, pp31-37, Aug. 2006.

[2] TIA/EIA-IS-856-A, "*cdma2000 High Rate Packet Data Air Interface Specification*", Apr. 2004.

[3] 3GPP2 C.S0084-001 v2.0, "*Ultra Mobile Broadband Air Interface Specification*", Sep. 2007.

[4] IEEE Standard for Local and metropolitan area networks Part 20: *Air Interface for Mobile Broadband Wireless Access Systems Supporting Vehicular Mobility - Physical and Media Access Control Layer Specification*, Aug. 2008.

[5] M. Wang, T. Ji, J. Borran, and T. Richardson, "Interference management and handoff techniques in Ultra Mobile Broadband communication systems**",** *The 10th International Symposium on Spread Spectrum Techniques and Applications*, pp. 166-172, August. 2008.

[6] M. Wang, A. Khandekar, A. Gorokhov, N. Bhushan, and A. Agrawal, "Preamble design in Ultra Mobile Broadband communication systems", *The Third International Workshop on Signal Design and Its Applications in Communications*, pp. 328-333, Sept. 2007.

[7] M. Wang, S. Aedudodla, A. Khandekar, R. Palanki, and A. Agrawal, "Preamble design and system acquisition in Ultra Mobile Broadband communication systems", *IEEE* 68th *Vehicular Technology Conference*, pp.1-8, Sept. 2008.

[8] 3GPP TS36.201, *LTE Physical Layer – General Description*, Aug. 2007.

[9] IEEE P802.16e/D12, "*Air interface for fixed and mobile broadband wireless access systems*", Oct. 2005.

[10] S. Nanda, R. Walton, J. Ketchum, M. Wallace, and S. Howard, "A high-performance MIMO OFDM wireless LAN", *IEEE Communications Magazine*, vol. 43, pp 101-109, Feb. 2005.

[11] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multiple antennas," *Bell Labs Tech. J.* vol. 1 , no. 2, pp. 41-59, 1996.

[12] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun*., vol. 6, pp.311-335, Mar. 1998.

[13] I. E. Telatar, "Capacity of multi-antenna Caussian channels," *Europ. Trans. Telecommun*., vol.10, pp585-595, Nov.-Dec., 1999.

[14] H. Sampath, P. Stoica, and A. Paulraj, "Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion," *IEEE Trans. Commun*., vol.49, pp.2198-2206, Dec., 2001.

[15] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inf. Theory*, vol.47, pp.2632-2639, Sep. 2001.

[16] D. Love, R. Heath, Jr., W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?" *IEEE Commun. Mag*., vol. 42, pp.54-59, Oct. 2004.

[17] D. Tse and P. Vishwanathan, *Fundamentals of Wireless Communications*, Cambridge University Press, 2005.

[18] P. Li, D. Paul, R. Narasimhan, and J. Cioffi, "On the distribution of SINR for the MMSE MIMO receiver and performance analysis, *IEEE Trans. Inf. Theory*, vol. 52, pp. 271-286, Jan. 2006.

[19] I. Kovalyov, *SDMA for Multipath Wireless Channels: Limiting Characteristics and Statistical Models*, Springer, 2004.

[20] A. Osseiran, P. Skillermark, and M. Olsson, "Multi-antenna SDMA in OFDM Radio network systems: Modeling and evaluations," *The 18th Annual IEEE Internal Symposium on Personal, Indoor and Mobile Radio Commun*., pp.1-5, Sept. 2007.

[21] M. Wang, and M. Dong, "Channelization in Ultra Mobile Broadband communication systems: The forward link," *International Wireless Communications and Mobile Computing Conference*, pp.614-620, Aug. 2008.

[22] M. Wang, and M. Dong, "Channelization in Ultra Mobile Broadband communication systems: The reverse link," *International Wireless Communications and Mobile Computing Conference*, pp.367-372, Aug. 2008.

[23] E. Biglierl, R. Calderbank, A. Goldsmith, A. Paulraj, and H. Vincent Poor, *MIMO Wireless Communications*, Cambridge University Press, 2007.

Michael Mao Wang received the B.S. and M.S. degrees in electrical engineering from Southeastern University in 1984 and 1987, respectively, and the M.S. degree in biomedical engineering in 1992 and the Ph.D. degree in electrical engineering in 1995 from the University of Kentucky. From 1995 and 2003, he was a Distinguished Member of Technical Staff at Advanced Radio Technology, Motorola Cellular Infrastructure Group, Arlington Heights, Illinois. He is currently with Qualcomm Corporate Research and Development, San Diego, CA. His current research interests are in the area of signal processing and wireless communication networks.

Dr. Wang was a recipient of Motorola Distinguished Innovator Award.

Alexei Gorokhov received his PHD in electrical engineering from Ecole Nationale Superieure des Telecommunications (Telecom Paris) in 1997. From 1997 till 2000 he was with French National Research Center (CNRS). From 2000 to 2003 he was with Philips Research Labs focusing on MIMO WLAN systems design and performance analysis. During the same period, he was visiting scientist in Information Systems Laboratory at Stanford University. He joined Qualcomm Inc in fall 2003 where he is currently Principal Engineer with Corporate Research and Development. His focus at Qualcomm is in design, evaluation and implementation of cellular broadband systems.

Min Dong received the B.Eng. degree from Tsinghua University, Beijing, China, in 1998, and the Ph.D. degree in Electrical and Computer Engineering with minor in Applied Mathematics from Cornell University, Ithaca, New York, in 2004. She is currently an Assistant Professor in the Faculty of Engineering and Applied Science, University of Ontario Institute of Technology (UOIT), Oshawa, Ontario, Canada. Prior to joining UOIT, she was with the Corporate Research and Development, Qualcomm Inc., San Diego, CA.

Dr. Dong received the 2004 IEEE Signal Processing Society Best Paper Award. Her research interests are in the general areas of communications, signal processing, and mobile networks.

© 2009 ACADEMY PUBLISHER