On the implementation of MIMO for LTE systems

Abheek Saha

November 23, 2012
1 Introduction and terminology

In recent years, wireless technology has been driven by the use of OFDM as the technology of choice for 4th generation wireless systems. By converting a broadband channel to a series of narrow band channels, OFDM makes it possible to combat frequency selective fading without requiring complex multi-tap adaptive equalizers. This allows OFDM wireless systems to use larger and larger bandwidths in a single channel and more complex modulations techniques; in the latest wireless LAN standards, a single channel of up to 160Mhz bandwidth is proposed.

Along with the introduction of OFDM, there has also been a resurgence in multiple antenna techniques - these build on the top of the success that adaptive coding and modulation . There are two major categories of multiple antenna technologies. The first is called beam forming or adaptive antenna systems; these use antennae placed closed together and a deliberately created pattern of constructive and destructive interference in the coverage region to reinforce signals and eliminate noise. For the rest of this paper, we call them Adaptive Antenna Systems. The second technology is frequently referred to as MIMO and refers to the use of ‘widely spaced’ antenna elements using independent paths to create multiple spatially separate, independent channels over a single physical resource. The two technologies are thus, in a sense, diametrically opposed to each other; one depends on the independence of spatially separated paths and the other on the coherence of spatially proximate paths.

MIMO techniques can be applied blind or can be made to adapt to the channel state information. In the latter case, the transmitter will estimate the current channel and use this to transform (or precode) the transmitted signal in such a way so as to take maximum advantage of the channel structure. The receiver then uses an inverse transform (postcode) on the received signal to recover the original data. The act of adapting a signal to match the path characteristics is often referred to as beamforming\(^1\) or closed loop transmit diversity/spatial multiplexing \(^2\). It is equally applicable in MIMO as well as Adaptive Antenna Systems (where it is also referred to as beamsteering).

In terms of usage, there are two main categories of MIMO operations. The first is spatial multiplexing, where separate independent streams are transmitted, one per transmit antenna...this helps raise aggregate throughput linearly and only works when all the paths are relatively independent of each other and have good path characteristics. Allocation of power to the different paths can take into account the path characteristics, subject to the overall power constraint. The other is transmit diversity, where the same data stream is transmitted simultaneously on multiple paths. The latter is a way of improving the quality of low SINR channels. The technique is distinct from standard multipath receivers used in earlier technologies due to the active precoding and the existence of multiple antennae on the transmitter, though many of the reception techniques are adaptation of the multipath techniques; specifically, LMMSE.

\(^1\)This terminology is usually used in IEEE standards
\(^2\)Preferred terminology in the 3gpp world
A newer technology is \textit{Multi user MIMO}, where a single transmitter with multiple transmit antennae uses spatial multiplexing to transmit simultaneously to separate receivers, each with a single or small number of antennae. Multi-user MIMO has a unique set of challenges on top of the standard MIMO, since it must find a way to separate the transmissions of different users.

In this document, we focus on the realization and implementation of MIMO in the LTE air interface and a set of practical methods for setting up a MIMO channel within the constraints imposed by LTE. First, we give an overview of the basics of how MIMO works.

\section{Fundamental MIMO theory}

\subsection{How MIMO works}

In a simplified MIMO system, we have a transmitter with $N_t$ transmit antennae and a single receiver with $N_r$ receive antennae. Each transmit antenna is used to transmit a stream of symbols $S_i(t)$, which may be derived from the same stream of data or from different streams of data; however, necessarily the signals will be time synchronized. The $j^{th}$ receive antenna receives a signal which combines the $N_t$ different transmissions, plus any additional noise $R_j(t) = \sum_i h_{ij}S_i(t) + n_j(t)$.

If we consider the block of antennae at the receiver operating together, then the
vector of received signals is given as
\[
\begin{bmatrix}
    r_0(t) \\
    r_1(t) \\
    \vdots \\
    r_{N_r-1}(t)
\end{bmatrix} = \begin{bmatrix}
    \sum_{j=0}^{N_t-1} h_{0,j} s_j(t) \\
    \sum_{j=0}^{N_t-1} h_{1,j} s_j(t) \\
    \vdots \\
    \sum_{j=0}^{N_t-1} h_{N_r-1,j} s_j(t)
\end{bmatrix} + \begin{bmatrix}
    n_0(t) \\
    n_1(t) \\
    \vdots \\
    n_{N_r-1}(t)
\end{bmatrix}
\]
(1)

\[
\begin{bmatrix}
    \sum_{j=0}^{N_t-1} h_{0,j} s_j(t) \\
    \sum_{j=0}^{N_t-1} h_{1,j} s_j(t) \\
    \vdots \\
    \sum_{j=0}^{N_t-1} h_{N_r-1,j} s_j(t)
\end{bmatrix} + \begin{bmatrix}
    n_0(t) \\
    n_1(t) \\
    \vdots \\
    n_{N_r-1}(t)
\end{bmatrix}
\]
(2)

In standard MIMO practice, this is written as the vector equation \( \vec{r}(t) = H(t) \vec{s}(t) + \vec{n}(t) \), where \( H \) is the CSI or channel state information matrix.

### 2.1.1 The properties of \( H \)

Note that \( H \) is a complex matrix, with its entries chosen at random; such a matrix is known as a Wishart matrix and its properties were extensively studied by mathematicians such as [Ede88] and others. If we take a Wishart matrix \( H \) and multiply it by its hermitian transpose \( H^* \) we get a positive semi-definite matrix \( O = HH^* \) with interesting properties; namely that its eigenvalues are distributed as per a Laguerre polynomial. It has been shown by [Tel99] and others, that the channel capacity of a MIMO channel is given by the modification of Shannon’s classic theorem for a matrix form

\[
C = \log \det (I + HQH^*)
\]
(3)

Here \( Q \) is the covariance matrix for the original signal; \( Q = E( ss^* ) \). Maximization of the RHS in (3) in the line of [Tel99] yields an optimum when \( Q = \text{diag} (\lambda_{i}^{-1}) \) where \( \lambda_{i} \) is the \( i \)th eigenvalue of \( HH^* \), the so called ‘waterfilling’ solution.

Thus, it is the eigenvalues of the CSI matrix which drives the performance of a MIMO system. The task of precoding the transmission vector is to introduce the appropriate covariance structure so as to maximize the SNR.

### 2.2 MIMO transceivers - precoding and postcoding

By the spectral decomposition theorem, any positive semidefinite matrix can be decomposed as:

\[
O = U \Sigma V^T
\]
(4)

where \( U \) and \( V \) are orthonormal matrices i.e. \( UU^T = I \) and \( \Sigma \) is the diagonal matrix of the eigenvalues of \( O \). This then, suggests the way to transmit a signal in the face of a known CSI matrix \( H \). The transmitter precodes the signal by
multiplying the signal vector with $\mathbf{U}^T \mathbf{H}$, the received signal is then given by

$$\mathbf{r} = \mathbf{U}^T \mathbf{H}^* \mathbf{s} + \mathbf{n} = \mathbf{U}^T (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^T) \mathbf{s} + \mathbf{n} = \mathbf{\Sigma} \mathbf{V}^T \mathbf{s} + \mathbf{n} \tag{5}$$

The receiver then post-codes this signal as:

$$\mathbf{r}^* = \mathbf{r} \mathbf{V} = \left[ \mathbf{\Sigma} \mathbf{V}^T \mathbf{s} + \mathbf{n} \right] \mathbf{V} = \mathbf{\Sigma} \mathbf{s} + \mathbf{n} \mathbf{V} \tag{6}$$

The entire operation is tantamount to a pseudo-inversion of the CSI matrix, since a direct inversion of a complex matrix is difficult and may cause undesirable side behaviours (clipping or amplification of the noise vector). Note that $\mathbf{V}$ being an orthonormal matrix, the postcoded noise vector $\mathbf{n} \mathbf{V}$ has the same vector norm as the original noise vector, thus the SNR is unchanged. The received signal $\mathbf{r}^*$ has no phase error, it simply has a linear scaling equivalent to the eigen values of the Wishart matrix, whose properties we have already discussed above.

Naturally, this is a simplification - for example, we assume that the channel matrix has full rank - in reality, matrices with small condition numbers i.e. the ratio of the smallest and the largest eigenvalues can cause the precoding and postcoding matrices to be unmanageable. Further, there may be constraints on the transmit power, which limits the choices of pre-coding vectors. In a more generic setting, we can define the problem as:

$$\text{Select } \mathbf{A}, \mathbf{B} \text{ so as to maximize}$$

$$\mathcal{I} = \log_2 \det \left( \mathbf{I}_r + \mathbf{H} \| \mathbf{B} \mathbf{A} (\mathbf{BA})^* \| \mathbf{H}^* \right) \tag{7}$$

Note that $\mathbf{r}, \mathbf{s}$ and $\mathbf{n}$ are the received signal vector, transmitted signal vector and noise vectors respectively, which are related by the equation

$$\mathbf{r} = \mathbf{B} \mathbf{H} \mathbf{s} + \mathbf{B} \mathbf{n} \tag{8}$$

$\mathbf{B}$ and $\mathbf{A}$ are the post and pre-coding matrices respectively and $\mathbf{I}_r = \mathcal{E} (\mathbf{m} \mathbf{s}^*$), representing the noise term.

An obvious constraint that comes is the power constraint i.e. $\| \mathbf{A} \mathbf{s} \| \leq P_t$, where $P_t$ is the maximum transmit power available.

### 2.3 The challenge of feedback and codebook design

#### 2.3.1 Feedback constrained MIMO

In an ideal system, the transmitter and receiver both know the value of $\mathbf{H}$ at a rate fast enough to keep up with channel variations, either due to doppler or natural environmental conditions. This allows both the transmitter and
the receiver to jointly compute the optimal pair of $A$, $B$ - the transmitter then uses $A$ for transmission and the receiver uses $B$ for reception. This also requires that the transmitter and receiver both use the same objective functions and constraints for solving the optimization problem in (7). However, this is not practical due to many reasons, one of the most significant ones being the unavailability of feedback bandwidth.

The feedback bandwidth is the amount of resources available for the receiver to give sufficient information to the sender regarding the state of the channel. Feedback is of two types, explicit and implicit. In the former, the receiver simply signals the channel state information to the transmitter and lets the transmitter compute the optimal pre-coding vector. Note that the transmitter has to somehow signal the chosen matrix to the receiver, so as to allow the receiver to compute the corresponding receive matrix; this makes explicit feedback of lower utility and somewhat rare in implementation. However, much research has been done in this area, including joint transmit receive beamforming, where an optimal pair of precoding and postcoding is simultaneously computed based on the estimated $H$ [PCL03]. Another option suggested is for a hybrid approach, where a fixed codebook is iteratively refined.

In the implicit feedback approach, the receiver tells the transmitter of the correct transmit configuration to be used; this takes the computing load off the transmitter and, in case of a fixed codebook transmitter, drastically reduces the amount of bandwidth to be used, since the receiver only has to signal a codebook entry. There has been a tremendous amount of research devoted to both options, including hybrid options i.e. having a fixed codebook at the transmitter which is then iteratively refined etc. Most modern systems, including LTE prefer to use the implicit feedback approach exclusively, or as the default option i.e. WiMAX.

### 2.3.2 Codebook design

The principal challenge of implicit feedback is to design a sufficiently generic codebook which can be applied to all channel matrices. The problem was initially studied in [LH05, LHS03]. The authors expressed the problem in terms of Grassmanian line packing and identified relevant metrics for codebook which maps to a specific optimization criterion. Specifically, if we consider a codebook to be a family of complex matrices, we need to know that they cover the domain space of possible channel matrices $H = (H_1, H_2, \ldots, H_{\infty})$. In [LH05], the following metrics have been identified:

- The chordal distance between any two entries in the codebook $F_1$ and $F_2$ is given by

$$d_{\text{chord}}(F_1, F_2) = \sqrt{M - \sum_{1}^{M} \lambda_i^2 (F_1^*F_2)}$$

---

3The problem of Grassmanian line packing is to orient $N$ lines in a $\mathbb{R}^k$ space through the origin, so that the sine of the minimum angle between any two lines is maximized [CHS96]
where $\lambda_i(K)$ is the $i$th eigenvalue of the matrix $K$

- The projection two norm distance is given by
  $$d_{\text{proj}}(F_1, F_2) = \sqrt{1 - \lambda_{\text{min}}^2(F_1^* F_2)}$$

- Finally, the Fubini Study distance is computed by
  $$d_{\text{FS}}(F_1, F_2) = \arccos\left| \det((F_1^* F_2)) \right|$$

The above three metrics are slightly different takes on the same concept of distance between linear operators. In [LH05], the authors have shown that

$$d_{\text{chord}}(F_1, F_2) \leq \sqrt{M}d_{\text{proj}}(F_1, F_2) \leq \sqrt{M} \sin(d_{\text{FS}}(F_1, F_2))$$

Further, the maximizing the projection norm $d_{\text{proj}}$ is optimal for multiple categories of receivers, including both the mean square error and maximum likelihood receivers described in section 4.

### 3 Implementation of MIMO in LTE

LTE offers comprehensive support for MIMO, with up to 7 different modes supported (though some of these come closer to adaptive antenna beamsteering). Among the modes supported are spatial diversity (a single stream being transmitted in multiple spatial modes), open loop and closed loop MIMO on up to 4 layers, multi-user MIMO (where the same broadcast transmission goes to multiple UEs) and two additional beamforming modes. The original LTE specs supported 4x4 mimo; with LTE advanced up to 8 separate layers are possible.

For a given air interface to be MIMO compliant, it has to support

- A method for the sender and receiver to share information about the channel state
- A method for the sender and receiver to signal each other and negotiate the mode of MIMO transmission
- A method to measure the channel state for a given number of channel paths, based on the MIMO mode

#### 3.1 Channel estimation and MIMO feedback

In LTE FDD, channel estimation is carried out by the receiver and sent to the transmitter. Channel estimation can be done using embedded pilots in the signal (DL) or by using dedicated sounding channels (UL). The methods for channel estimation have been studied extensively - a small survey can be found in [wik]. MMSE and Maximum Likelihood estimation techniques are useful for getting accurate channel estimation.

Once the channel is estimated, we have a channel state matrix which represents the receiver best guess about the channel state.
3.2 Transmit precoding

LTE defines the operation of transmit precoding in two steps. In the first step, there is a layer mapping. Each layer represents an individual stream and has a code-word (the output of the FEC coding) mapped to it. The mapping can range from pure transmit diversity i.e. the same code-word to all layers, to pure SM i.e. each layer has a different code word and a mix of the two (for example, 2 code words split over 4 layers). Then the layers are combined into a vector and multiplied by a precoding matrix indicated from a pre-defined code book; the output is then mapped to antenna ports which are the actual streams being transmitted.

3.3 Selection of transmit beamforming matrices

The LTE specification has designed a codebook of precoding matrices so as to simplify the selection procedure at the receiver. While the 2x2 code-book is a simple set of orthogonal matrices with determinant 1, the 4x4 and 4x2 matrices are defined in terms of Householder matrices. A householder matrix is an orthogonal, symmetric matrix which preserves norm i.e. if $P$ is a householder matrix and $Y = PX$, then $Y$ and $X$ will have the same norm.

Householder matrices are constructed as $P = I - \frac{2\langle w, x \rangle}{w^Tw}$. Consequently, a householder matrix corresponding to the vector $[1 \ -1 \ -1 \ -1]$ becomes

$$
I - \frac{2}{4} \begin{pmatrix}
1 & -1 & -1 & -1 \\
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 \\
1 & -1 & -1 & -1
\end{pmatrix}
$$

An alternate way of thinking of a Householder matrix is as a reflection matrix; when derived from the vector $\vec{w}$ it reflects the vector $\vec{X}$ around the hyperplane which is normal to the vector $\vec{w}$.

In the case of LTE, 4x4 and 4x2 code books are defined as selected columns from the Householder matrix $P(\vec{w})$, where the vectors $\vec{w}$ correspond to 16 mutually orthogonal hyperplanes passing through the origin of a 4 dimensional space. Thanks to this construction, the algorithm to choose the optimal precoding matrix becomes simplified, as discussed below.

3.4 Optimality of the LTE codebook

We can apply the criterion given in 2.3.2 to the codebook defined in LTE, specifically householder matrices which are generated using the vectors specified in LTE. A simple computation of chordal distance gives values in the range of 3.4 to 3.7. A study in [VODC08] shows that the Householder matrices underperform optimal Grassman matrices by about 2 to 6 dB.
4 Optimal decoding

We now come to the topic of optimal decoding at the receiver, given knowledge of the transmitters configuration. While this is an enormous subject, we will focus on the top few methods.

4.1 Linear MMSE

MMSE or minimum mean squared error estimator attempts to find an estimation function which minimizes the average error for each given input. Consider, as in the case above, we have received a signal $\vec{y}$, which could have been generated by the vector $\vec{x} \in \chi$ which can take values from some constrained space $\chi$. The MMSE estimator tries to find an estimation function $g()$ so as to minimize the mean square error $E_{\vec{x} \in \chi} \|g(\vec{y}) - \vec{x}\|$. If we restrict ourselves to linear operations, we can replace the function $g()$ by a matrix multiplication $G_{MMSE}$.

Given that we know the relationship between $\vec{y}$ and $\vec{x}$ from (8), it can be shown that the optimal MMSE solution is given by (9)

$$G_{MMSE} = \left((AH)^H (AH) + \sigma^2 I\right)^{-1} (AH)^H$$

Linear MMSE techniques in estimation and decoding have been shown to have extremely good performance, matching techniques such as RAKE in high SNR conditions and outperforming them substantially in the face of multi-user interference i.e. when the transmitter is transmitting multiple signals simultaneously for different users.

SIC or Successive Interference Cancellation is a technique on top of MMSE proposed in the BLAST system to further improve performance in the face of multiple access interference.

4.2 ML and Lattice Sphere Decoding

Maximum likelihood estimation, on the other hand, does not look for an estimator; it tries to directly find the 'most likely' value of the transmit vector i.e. $\vec{x}^* = \arg \min_{x \in \Omega} \|\vec{y} - AH\vec{x}\|$. ML solutions were, at one time, considered to be intractable and in general, can be shown to be NP-complete. However, a related class of algorithms can be found, which can find solutions which are within a C sphere of the optimal. These are known as Sphere algorithms; examples include the Schnorr Euchner algorithm [GN06]. These algorithms try to solve the inequality $\|y - Hx\|^2 \leq C_0$. By doing a QR decomposition of $H$, we can get a modified version of this as $\|Qy - Rh\|^2 \leq C_0$. Since $R$ is a right triangular matrix, the system can be converted into the set of linear equations:

$$\sum_j |\tilde{y}_j| - \sum_{l=1}^{N} r_{j,l} |x_l| \leq C_0$$

This can then be solved using a multiplicity of techniques, with fairly good average running complexities (though the worst case running complexities are exponential).

Regardless of the specific receiver algorithm used, one critical factor is the ability to compute $HAA^H H^H$ for different possible precoding matrices. The choice of householder matrices for the precoding matrix $A$ makes the computation of simpler by up to 50% [WWZ10]. A matrix of the form of $W^H H^H WH$ where $W$ is the householder matrix generated by the vector $\vec{w}$ can be simplified to be expressed in terms of $\vec{w}^H H^H \vec{h}_i$, where $\vec{h}_i$ is the $i^{th}$ column of $H$.

4.3 Multi-user MIMO

If multiple user terminals select the same precoding matrix (for downlink transmission) or are assigned to use the same pre-coding matrix (for uplink transmission), then they can be scheduled on the same physical resource simultaneously, by simultaneously transmitting individual codewords to individual user terminals from different antennae. There are further challenges (for example, maintaining the power level to the user terminal, allocating Ue specific reference signals, etc.) which are also handled in the LTE specifications.

5 Conclusion

Support for MIMO is a fundamental aspect of the LTE air interface; the consequent technical choices are a tradeoff between implementation considerations and performance. As compared to other comparative air interfaces, e.g. IEEE 802.16e, the LTE air interface is relatively simpler. For example, the IEEE 802.16e supports dynamic generation of codebooks and upload of the same from the receiver to the transmitter over-the-air. Many of these features have been skipped in LTE simply because they were not considered to be practically realizable. Subsequently, LTE-Advanced has added and will further add additional features, including 8x8 antenna support and more advanced forms of uplink and downlink multi-user MIMO [LHZ09]. As more and more MIMO capable terminals are commercially available, further challenges in the actual implementation will be discovered. Already, significant research is being carried out in areas related to estimation error (of the CSI) and the consequences, thereof.

References


[wik] *Frequency domain estimation for lte ofdm systems - a tutorial.*