Very few technologies have had as much impact on the trajectory of evolution of wireless communication systems as Multiple Input Multiple Output (MIMO) systems. MIMO systems have already been employed in the existing 802.11n and 802.16e standards resulting in a huge leap in their achievable rates. A relatively recent idea of extending the benefits of MIMO systems to multi-user scenarios seems promising in the context of achieving high data rates envisioned for future cellular standards after 3G. Although substantial research has been done on the theoretical front, recent focus is on making multi-user MIMO (MU-MIMO) practically realizable. It offers an enormous scope for further research in the coming years. As in the case of any evolving technology in communication systems, the literature concerning MU-MIMO systems involves complex mathematical analysis, making it difficult for an ordinary reader to comprehend. This article aims at giving an insight into MU-MIMO systems—the concept, fundamentals, and trends including an overview of important research results. It is intended at giving a good start to amateurs interested in being part of the community that shapes the future of wireless systems.

Spatial diversity and multiplexing

Multiple antennas at the transmitter and receiver side have been traditionally used to realize diversity in order to improve reliability of the channel. Spatial diversity is achieved by transmitting the data signal over multiple channel links created due to the use of multiple transmit and receive antennas. Diverse multipath fading at the receiver antennas offers multiple "views" of the transmitted data at the receiver thus increasing its robustness. Figure 1 explains spatial diversity using a common scenario. The two cameras viewing the object from different locations offer the observer additional information about the object as compared to a single camera. In a multipath scenario where each antenna would experience a different interference environment, there is high probability that, if one antenna is suffering a deep fade, another antenna has a sufficient signal level. Spatial diversity improves channel reliability without using any additional spectrum or transmission time but at the expense of increase in the transmitted power.

In 1996, A.J. Paulraj and T. Kailath showed that multiple transmit and receive antennas can improve the channel capacity by transmitting independent data streams over multiple channel links created by multiple transmit and receive antennas. The gain in channel capacity is proportional to the available number of antennas at the transmitter or receiver side, whichever is less. The antenna arrays at the transmitter and the receiver side create a multilink channel that can be expressed in the form of a channel matrix $H_{\text{NRxNT}}$ (shown in Fig. 2) where $N_R$ and $N_T$ are the number of transmit and receive antennas. If $x$ is a vector containing transmitted symbols, then a simplistic way to express the symbol at the receiver can be

$$y = Hx + w$$

Digital Object Identifier 10.1109/MPOT.2009.000000
where \( w \) is additive white Gaussian noise. The number of independent data streams that can be transmitted is less than or equal to \( \min(N_T, N_R) \). It was found out that decoding at the receiver (i.e., separating data streams from the received symbols at each receiver antenna) is possible if the receiver knows the channel matrix.

The equalization of a MIMO channel is done at the receiver using zero forcing (linear transformation of the received symbol for removing inter-channel interference by multiplying it with the inverse of channel matrix) or by maximum likelihood detection (mapping the received data vector to its closest match), both of which require the receiver to have channel matrix information that can be derived from suitable training. The format of typical 802.11n packet is shown in Fig. 3. A training field is a fixed symbol specified in the standard and is known both at the transmitter and receiver. The training fields help the receiver in frequency and phase correction, detecting symbol boundaries, and tone equalization.

A look at MU-MIMO

The use of MIMO systems was traditionally intended for point to point communication. The natural extension of this thought would be to consider MIMO systems in a multiuser scenario. The vision for next generation cellular networks includes data rates approaching 100Mbps for highly mobile users and up to 1Gbps for low mobile or stationary users. This calls for efficient use of the existing spectrum. MU-MIMO technology is expected to play a key role in this context.

There are two challenges in a multiuser MIMO scenario: uplink (where multiple users transmit simultaneously to single base station) and downlink (where the base station transmits to multiple independent users). The uplink challenge is addressed using array processing and multi-user detection techniques by the base station in order to separate the signals transmitted by the users. The downlink challenge is somewhat different. MU-MIMO downlink channel is similar to that of single user MIMO (SU-MIMO) except that the receiver antennas are distributed among different independent users as shown in Fig. 4. This creates a challenge in decoding the received symbols since joint decoding requires each user to have the symbol received from all the receiver antennas of all the users. It is practically impossible to achieve this level of coordination between all users.

Almost all of the proposed techniques ideated for addressing the MU-MIMO downlink challenge employ processing of data symbols at the transmitter itself known as precoding. Although precoding is not a new concept and has been used in SU-MIMO systems as well, it was optional and used only to improve SNR at the receiver. However, in MU-MIMO systems precoding is essential to eliminate or minimize inter-user interference. Precoding is performed with the help of downlink channel state information or CSI. This requires the transmitter to know the downlink CSI of each user in order to model the precoding transformation variables for each user. A number of different techniques to address the issue of MIMO downlink transmission and reception have been proposed. A few of them have been described in the sections below.

Dirty paper coding

Dirty paper coding is an information theoretic concept proposed by M. Costa in 1983. It states that if the transmitter knows the interference that will affect the signal during its propagation, then the capacity of such a channel is same as that of a channel without interference.
This means that preemptive interference cancellation at the transmitter is possible without an increase in signal power. Although Costa’s approach is theoretical and practically difficult to implement, it serves as a yardstick for practical MU-MIMO techniques to characterize the sum capacity of multi-antenna, multi-user channels as each proposed technique aims to reach close to this theoretical limit. The sum capacity in a multi-user MIMO broadcast channel is defined as the maximum aggregation of all users’ data rates. In spite of its complexity, dirty paper coding has inspired a few nonlinear precoding techniques that are described later in this article.

**Quality of service paradigm**

One issue inherent to MU-MIMO systems is choosing the measure of quality of service (QoS). The dirty paper coding-based approach uses sum capacity as a measure of QoS. This approach aims at maximizing sum capacity with a constraint on transmitted power. The problem with this approach is that it may result in an uneven QoS to individual users by creating “strong users” who take a dominant share of the available power leaving the “weak users” with little or no throughput. The second approach is each user-centric. It aims at minimizing the transmitted power subject to achieving a desired arbitrary rate or threshold for Signal to Interference Noise Ratio (SINR) for each user. For single user systems with the absence of SINR and assuming Gaussian noise, these two approaches are equivalent.

**Downlink precoding techniques**

A typical configuration of a multi-user MIMO downlink system is shown in Fig. 4. In this section, some approaches to solving MU-MIMO downlink challenge are described in brief.

### Decomposition of MU-MIMO channel to SU-MIMO channel

This approach is aimed at perfect cancellation of interuser interference at each user at the transmitter itself. Figure 5 shows the linear precoding applied to symbols to achieve this cancellation. Assume that the base station has M transmit antennas and there are K users, each with one receiver antenna. Let \( b(k) \) be the symbol meant for \( k \)th user. Each user is assigned a precoding matrix \( M(k) \) that will transform the symbol for \( k \)th user before it is mixed with symbols from other users at the transmit antennas. This is a simplistic model assuming flat fading channel. The symbol seen by the \( k \)th user is expressed as

\[
r(k) = H(k) \sum_{i=1}^{K} M(i)b(i) + n(k).
\]

Here, \( H(k) \) is the downlink channel for \( k \)th user and \( n(k) \) represents the additive white Gaussian noise (AWGN). Note that the received symbol consists of interference from other users. This component of interuser interference is given by

\[
H(k) \sum_{i=1}^{K} M(i)b(i) =: V(k)A(k).
\]

In order to cancel interuser interference, the task is to design the precoding matrix \( M(\cdot) \) for each user such that

\[
H(k) \sum_{i=1}^{K} M(i)b(i) = 0.
\]

It implies that the precoding matrix \( M(k) \) should lie in the subspace, which is the intersection of the kernels of the channel matrices \( H(i) \) of all the users except that of \( k \)th user. \( M(k) \) can be expressed as \( M(k) = V(k)A(k) \) where \( V(k) \) represents the orthonormal basis of the above subspace, which achieves diagonalization of the effective channel matrix after considering the precoding as a part of the transformation. \( A(k) \) is a nonzero matrix that can be designed alone by some criteria or jointly designed with the structure of the receiver or based on the QoS targeted. The precoding vector in this case decomposes the effective channel matrix into parallel single user MIMO channels as seen by the receiver so that the receiver can use

---

**Fig. 4**

**Fig. 5**
SU-MIMO detection techniques without being bothered with interuser interference.

This technique can be extended for use in scenarios where each user is equipped with more than one receive antenna and is capable of receiving more than one space multiplexed symbols. In this scenario we need not achieve diagonalization of effective channel matrix for each user antenna since joint decoding is possible for antennas of the same user. If the network channel matrix is expressed as $H_k = [H_1^T H_2^T \cdots H_K^T]$ and precoding matrix for all users as $M = [M_1 M_2 \cdots M_K]$ then the optimal transmit precoding vectors $M_s$ are found such that the resulting product $H_k M_s$ is block diagonal. Note that the single antenna per user case described previously is a subset of this method when the product $H_k M_s$ will be completely diagonalized.

**Signal to leakage ratio maximization algorithm**

The above method of perfect cancellation of interuser interference puts forth a constraint that the number of transmit antennas at the base station should be greater than or equal to the sum of all the receive antennas of all the users. An alternative to this approach is the maximization of signal to leakage ratio for each user given a constraint on total transmitted power. As seen from Eq. 1, each user sees some amount of interference from other users' signal (i.e., there is some power leakage from the effective channel of one user to that of others). It can be proved that the signal to leakage ratio for $k$th user is given by

$$\text{SLR}(k) = \frac{|H(k)M(k)|^2}{|H_k^T H_k|^2}$$

where $H_k^*$ is $H$ without $H(k)$ and $M_k$ is $M_s$ without $M(k)$. This approach can support as many subchannels as the user's antennas, which is not achievable in perfect interference cancellation schemes. This algorithm further employs an iterative process called successive optimization to calculate the precoding vectors for each user. After determining the precoding vector for the first user, it calculates the leakage from the first user to all the other users in order to calculate SLR and the precoding vector for next successive users and so on.

**Vector perturbation, lattice precoders, and regularized precoding**

The channel inversion technique described above is an attractive solution due to its simplicity. However, the data transmitted after precoding using channel inversion has a large norm, which means that there is increase in the transmitted signal power. Due to this, the sum rate for channel inversion in its plain form is poor. The sum rate does not grow linearly with minimum of transmit or user antennas as expected due to the large spread in the singular values of the channel matrix. Nonlinear precoding techniques, although more complex than zero forcing techniques, perform closer to sum capacity.

Vector perturbation technique modifies the data to be transmitted $b$ to $b'$ before multiplying it with inverse of channel matrix by perturbing it with an integer offset so that the data vector after zero forcing has a smaller norm and reduced power compared to the plain channel inversion method.

$$b'(k) = b(k) + \tau l$$

where $l$ is a positive real number and $l$ is a complex vector $a + ib$ with dimensions equal to the number of user antennas where $a$ and $b$ are integers. The receiver uses modulo operation to recover the signal (see Fig. 6). It has been proved that choosing the optimum value of $l$ makes the power of the transmitted signal independent of the number of transmitter or receiver antennas, whichever is less and also achieves its lower bound. It does so by placing the largest signal components along the smallest singular values of the inverse channel and the smallest signal components along the largest singular values of the inverse channel.

Finding the optimum value of $l$ is a $K$-dimensional integer-lattice least squares problem where $K$ is the number of user antennas. Recent efforts are aimed at discovering a solution to this problem of finding optimum integers from a lattice.

One solution could be the use of a sphere decoder developed by Fincke and Pohst. It avoids an exhaustive search over all possible integers in a lattice by limiting the search space to a sphere of some given radius centered on a starting point. The scalar $\tau$ is chosen to provide symmetric decoding region around every signal constellation point. A large $\tau$ increases the decoding region at the upper and lower extremes of the constellation but reduces the effect of vector perturbation by increasing the norm of the transmitted data matrix. The plain channel inversion technique decomposes the effective channel matrix using singular value decomposition. In order to cancel multi-user interference, it assumes that each receiver has the knowledge of the left singular matrix of the SVD. This requires global channel state information (CSI) at the receivers or additional training. It has been shown that lattice-based techniques used in conjunction with zero forcing techniques eliminate this requirement. In this case, the receiver requires a simple decoding filter.

Another method to improve the performance of channel inversion technique in low SNRs and high values of $K$ is regularization of the channel inverse by adding a multiple of identity matrix before inverting, i.e., using $s = H^T (H H^T + \alpha I)^{-1} b(k)$ instead of plain inversion $s = H^T (H H^T)^{-1} b(k)$. Using channel inverse regularization along with vector
perturbation has been shown to achieve near-capacity performance at all SNRs.

**Multi user diversity and scheduling**

A practical situation that has not been considered in the discussion until now is multi-user diversity. Multi-user diversity is a form of selection diversity among users; the base station can schedule its transmission to those users with favorable channel fading conditions to improve the system throughput. In MU-MIMO schemes, this feature is particularly important due to the possibility for the base station to group users for spatial multiplexing that improve the sum rate. It has been proved that the suboptimal zero forcing techniques approach optimal performance by scheduling due to the large available sample space when the number of users is large. An optimum scheduling algorithm performs an exhaustive search over all possible combination of all active users so as to find a group that will help achieve the maximum sum rate. A user is considered to be active if there is incoming data for intended for it at the access point. However, such a technique is computationally onerous. Moreover, it neglects the fairness criterion, which expects time constrained service delay for each user irrespective of its downlink channel quality. A “greedy” scheduling scheme reduces computational complexity by choosing the user with the highest channel capacity and then choosing successive users of the group who will maximize the sum capacity. To address the fairness criteria the greedy scheduling scheme can be combined with round-robin scheduling. It means that after the first group of users is scheduled for transmission, a second group of users is selected from the remaining users in the same “greedy” fashion. This procedure is iterated until each active user is accommodated in some group.

**Converging toward a realizable solution**

The recent focus of research work in MU-MIMO systems has been toward addressing practical challenges in implementing MU-MIMO systems as a tangible technology for future cellular standards. Imperfect CSI, quantization of CSI feedback by the users to the transmitter, feedback overhead, receiver training, and application to frequency selective fading channels are some of the practical difficulties.

The necessity of CSI at the transmitter will introduce substantial overhead on the system throughput. A number of limited feedback schemes have been proposed to minimize this overhead. One scheme proposes to feedback only the shape of the channel (i.e., normalized channel vector). This is because the shape vector mainly captures the directional knowledge of paths, which usually varies much more slowly than the amplitude of the channel. In other schemes it has been proposed to quantize the precoder for a MIMO channel and not simply the channel coefficients. The challenge of extending this work to the multi-user channel is that the transmit precoder depends on the channels of the other users in the system.

A codebook based quantized feedback is another promising scheme that incurs the least feedback overhead. The base station and the users have access to a CSI codebook, which is designed offline. Each user sends the binary index of the best code vector from the codebook through a zero-error, zero-delay feedback channel to the base station. It has also been shown that the channel shape alone does not achieve the full multiplexing and multi-user diversity gain simultaneously. To achieve both gains, channel quality information feedback (CQI) is necessary, and that CQI should be the SINR rather than just the channel magnitude, since SINR captures both the channel magnitude and the quantization error.

The optimal gain due to multi-user diversity is achieved when there is a large number of users. However, as the users increase, the CSI feedback overhead also increases. One scheme to address this issue proposes the use of threshold-based feedback before scheduling the users (i.e., the users only send the information that whether the SINR at their end is greater than a required threshold value). This allows the base station to schedule users for transmission. Only then are the scheduled users requested for a full CSI feedback.

Grassmannian line packing has also been proposed to compute precoding vectors that generate nearly orthogonal beams. This method allows the base station to generate number of beams larger than the number of available base station antennas, which have minimum correlation. The number of beams generated depends on the minimum rate-per-user requirement. In the training phase each user sequentielly calculates the SINR for each beam and feeds back the index of the beam with minimum SINR along with the SINR value.

Figure 7 shows a MAC framing structure proposed in “Opportunistic Feedback for Multiuser MIMO Systems with Linear Receivers” by Tang, Heath, Cho and Yun. The broadcast message is followed by $M_t$ contention channel fields each with $K$ mini slots for feedback from the users. Each user with its channel quality above certain level vies for one of the channel slots selected randomly for transmitting its id and channel quality information to the base station. A user can successfully send its information to the base station only if
there is no contention for that mini slot. The number of feedback slots \( K \) is chosen in order to minimize contention at the same time controlling feedback overhead. It was shown that this strategy requires lower feedback overhead and at the same time exploits multi-user diversity.

**Conclusion**

The consummation of all the research work in MU-MIMO systems will be in the form of a standard. It will require collating all the research findings into a deployable standard with well defined transmit and receive processing, scheduling options, receiver training fields, CSI feedback rates, and other specifications essential for compatibility. MU-MIMO technology offers aspiring researchers an exciting challenge and an opportunity to work on a solution that vies for a spot that will make it key technology in future cellular systems.

**Read more about it**

- S. Shim, C. B. Chae, and R. W. Heath, Jr., “A lattice-based MIMO broadcast precoder for multi-stream transmission.”<AU: Is this a journal type reference? If yes, kindly provide the journal title, volume number, issue number, complete page range, and year of publication.>

**About the author**

Aditya Kurve (adikurve@gmail.com) is a member of the IEEE and received his bachelor of engineering degree in electronics engineering from the University of Pune, India, in 2006. He has worked on the digital part of the 802.11n baseband PHY processor at Conexant Systems. Currently, he is working as a design engineer at Sasken Communication Technologies, Pune.