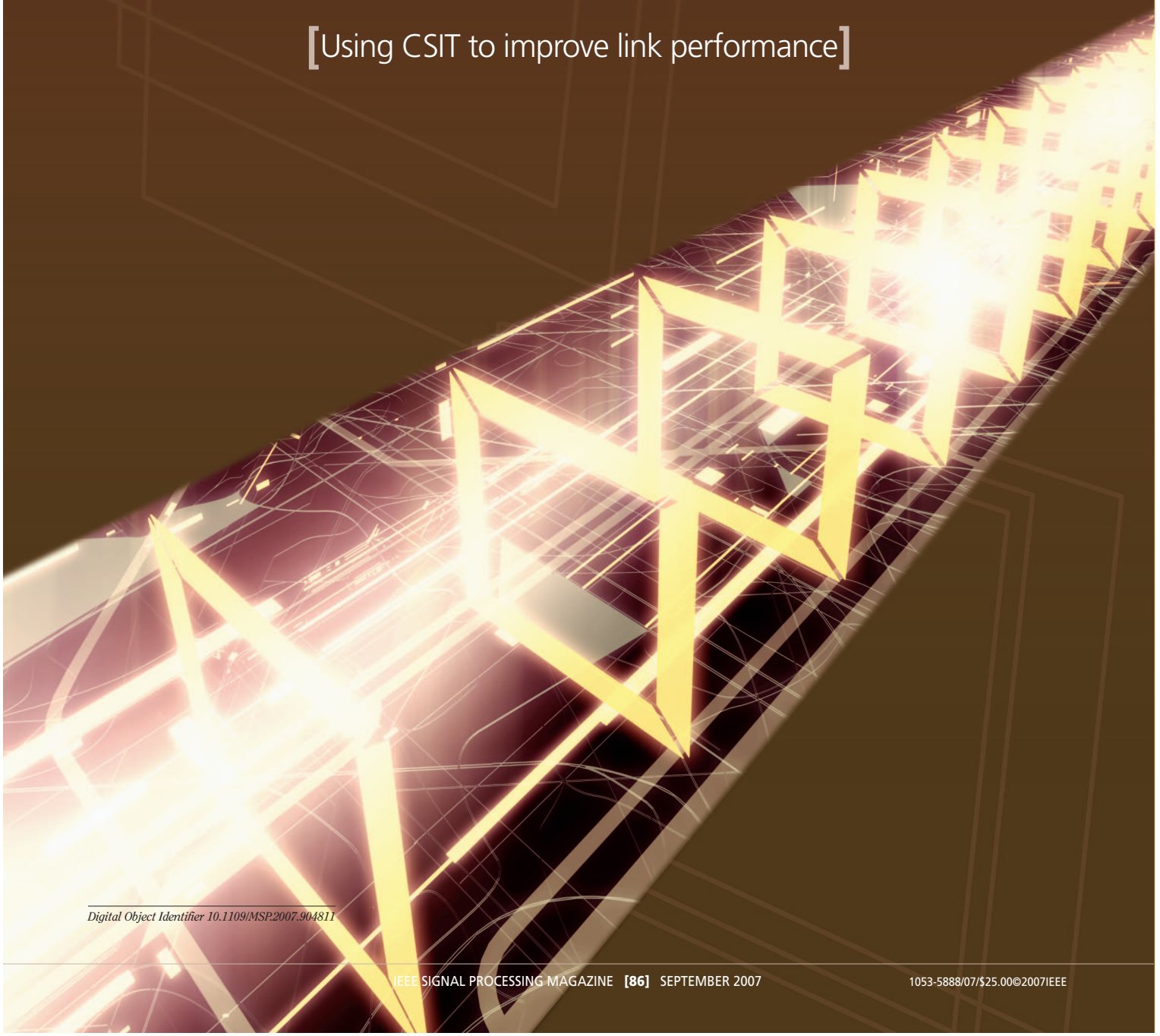
[Mai Vu and Arogyaswami Paulraj]

MIMO Wireless Linear Precoding

[Using CSIT to improve link performance]



Digital Object Identifier 10.1109/MSP.2007.904811

1053-5888/07/$25.00©2007IEEE

The benefits of using multiple antennas at both the transmitter and the receiver in a wireless system are well established. Multiple-input multiple-output (MIMO) systems enable a growth in transmission rate linear in the minimum number of antennas at either end [1], [2]. MIMO techniques also enhance link reliability and improve coverage [3]. MIMO is now entering next generation cellular and wireless LAN products with the promise of widespread adoption in the near future.

While the benefits of MIMO are realizable when the receiver alone knows the communication channel, these are further enhanced when the transmitter also knows the channel. The value of transmit channel knowledge can be significant. For example, in a four-transmit two-receive antenna system with independent identically distributed (i.i.d.) Rayleigh flat-fading, transmit channel knowledge can more than double the capacity at −5dB signal-to-noise ratio (SNR) and add 1.5 b/s/Hz additional capacity at 5 dB SNR. Such SNR ranges are common in practical systems such as WiFi and WiMax applications. In a non-i.i.d. channel (such as correlated Rician fading), channel knowledge at the transmitter offers even greater leverage in performance. Therefore, exploiting transmit channel side information is of great practical interest in MIMO wireless. In this article, we assume full channel knowledge at the receiver and study how channel-side information at the transmitter (CSIT) can be used to improve link performance. While the origins of using CSIT at the transmitter or precoding dates back to Shannon [4], MIMO precoding has been an active research area during the last decade, fueled by applications in commercial wireless technology.

Precoding is a processing technique that exploits CSIT by operating on the signal before transmission. For many common forms of partial CSIT, a linear precoder is optimal from an information theoretic view point [4]–[6]. A linear precoder essentially functions as a multimode beamformer, optimally matching the input signal on one side to the channel on the other side. It does so by splitting the transmit signal into orthogonal spatial eigenbeams and assigns higher power along the beams where the channel is strong but lower or no power along the weak. Precoding design varies depending on the types of CSIT and the performance criterion.

TYPES OF CSIT

The random time-varying wireless medium makes it difficult and often expensive to obtain CSIT. In closed-loop methods, the limited feedback resources, associated feedback delays, and scheduling lags degrade CSIT for mobile users with small channel coherence time. In open-loop methods, antenna calibration errors and turn-around time lags again limit CSIT accuracy. Therefore, we often only have imperfect instantaneous channel state information. We may, however, decide to exploit only certain parameters of the channel such as the Rician K factor or channel condition number to reduce the amount of information to be tracked instantaneously. In some cases such as fast fading channels or systems with long delay, we may give up tracking real-time information and provide CSIT in terms of the channel statistics, such as the channel mean and covariance or antenna correlations. Statistical CSIT is obtained from channel observations over multiple channel coherence times. In this article, we use CSIT to mean channel side information at the transmitter, which includes not only instantaneous channel state information, but also the channel parameters and statistics.

To understand different types of CSIT in wireless, it is necessary to know how the CSIT is obtained. There are two principles for obtaining CSIT: reciprocity and feedback. Reciprocity involves using the reverse channel information (open-loop), while feedback requires sending the forward channel information back to the transmitter (closed-loop). These techniques are discussed in detail in the following sections. In both cases, there exists a delay, such as a scheduling or a feedback delay, between when the channel information is obtained and when it is used by the transmitter. The information accuracy will depend on this delay and on the channel estimation technique. Channel estimation either at the receiver or transmitter is the starting point for deriving CSIT, and its accuracy will depend on the estimation technique and SNR. Since for most applications, channel estimation is also required for receiver processing, it is usually sufficiently accurate for precoding purposes. Depending on the type of information and how fast the channel changes with time, however, the delay in CSIT acquisition can significantly affect the CSIT accuracy.

Error-free instantaneous channel state information or perfect CSIT, therefore, is usually difficult to obtain in wireless; more often, only incomplete or partial channel information is available to the transmitter. Instantaneous CSIT can be characterized by a channel estimate and an associated error covariance [7], [8]. Both quantities are dependent on the delay in acquiring CSIT. As this delay increases, the CSIT approaches the channel statistics [8]. Thus, both instantaneous and statistical CSIT can be expressed in the same form: a channel estimate or mean, and an error or channel covariance.

APPROACHES TO PRECODING DESIGN

Although the term *precoding* is sometimes used in the literature to represent any transmit processing besides channel coding, we clarify its use here to strictly mean the transmit signal processing that involves CSIT. MIMO techniques without CSIT are clarified as space-time (ST) coding. Since the work of Shannon [4], more recent results show that, for a flat-fading wireless channel, provided a mild condition that the current channel state is independent of the previous CSIT when given the current CSIT, the capacity can be achieved by CSIT-independent coding together with CSIT-dependent linear precoding [5], [6]. The linear precoder directs signal spatially and allocate power in a water-filling fashion over both space and time. Power allocation over time can slightly increase the capacity of a fading channel at low SNRs, but has diminishing impact as the SNR increases beyond roughly 15 dB [9]. Depending on the antenna configuration, allocation over space, on the other hand, can significantly increase the capacity at all SNRs. This motivates precoding designs to exploit spatial CSIT.

In designing the precoder, various performance criteria have been used. To achieve the ergodic capacity, the precoder shapes the covariance matrix of the optimal transmit signal to match the CSIT [7], [10]–[17]. Precoders can also be designed according to more practical measures, such as the mean-square error (MSE), an error probability [pair-wise error probability (PEP), symbol error rate (SER), bit error rate (BER)], or the received SNR [7], [18]–[31]. These different precoder designs can be analyzed using the common linear precoding structure.

SCOPE

This article provides a tutorial of linear precoding for a frequency-flat, single-user MIMO wireless system, examining both theoretical foundations and practical issues. The article first discusses principles for CSIT acquisition and develops a dynamic CSIT model, which spans perfectly to statistical CSIT, taking into account channel temporal variation. It then presents the capacity benefits of CSIT and information theoretic arguments for exploiting the CSIT by linear precoding. A precoded system structure is then described, involving an encoder and a linear precoder. Criteria for designing the precoder are then discussed, followed by specific designs for different CSIT scenarios. These designs are analyzed in terms of the linear precoding structure, and their performance is illustrated by numerical examples. A brief survey of application follows, involving practical channel acquisition techniques and precoding deployment in current wireless standards. Finally, the article concludes with a discussion of other partial CSIT types and the continuing role of precoding. The aim is to build intuition and insight into this important field of MIMO linear precoding while leaving the details to references.

CSIT ACQUISITION AND MODELING

CSIT ACQUISITION TECHNIQUES

In a communication system, since the signal enters the channel after leaving the transmitter, the transmitter can only acquire channel information indirectly. The receiver, however, can estimate the channel directly from the channel-modified received signal. Pilots are usually inserted in the transmitted signal to facilitate channel estimation by the receiver. Fortunately, modern communication systems are usually full-duplex with a transceiver at each end. The transmitter thus can acquire CSIT based on the channel estimates at a receiver, by either invoking reciprocity or using feedback.

OPEN-LOOP CHANNEL ACQUISITION

The reciprocity principle in wireless communication states that the channel from an antenna A to another antenna B is identical to the transpose of the channel from antenna B to antenna A, provided the two channels are measured at the same time, the same frequency, and the same location. This principle suggests that the transmitter can obtain information of the forward (A to B) channel from the reverse (B to A) channel measurements, which the receiver at A can measure. This information can involve the instantaneous channel or other channel parameters, including the channel statistics. In real full-duplex communications, however, the forward and reverse links cannot use all identical frequency, time, and spatial instances. The reciprocity principle may still hold approximately if the difference in any of these dimensions is relatively small, compared to the channel variation across the referenced dimension.

Consider a base node for example. The node measures the reverse channel during reception and uses this measurement for the CSIT of the next transmission. In voice applications, the forward and reverse links to all the users operate in back-to-back time slots. Therefore, reverse channel measurements can be made regularly using embedded pilots. These measurements periodically refresh the CSIT. In data communications, the forward and reverse links may not operate back-to-back; hence, specially scheduled reverse-link transmissions for channel measurements known as channel sounding are used. A subset of the users, for whom CSIT is required, is scheduled to send a sounding signal. The sounding signals are orthogonal among simultaneously scheduled users, using orthogonal subcarriers as in orthogonal frequency division modulation (OFDM) or orthogonal codes as in code division multiple access (CDMA). Channel sounding is efficient for systems with many antennas at the base node.

One complication in using reciprocity methods is that the principle only applies to the radio channel between the antennas, while in practice, the channel is measured and used at the baseband processor. Different transmit and receive RF hardware chains therefore become part of the forward and reverse channels. Since these chains have different frequency transfer characteristics, reciprocity requires transmit-receive chain calibration to equalize the two chains (see [32] for example). Calibration is expensive and has made open-loop methods less attractive in practice.

CLOSED-LOOP CHANNEL ACQUISITION

Another method of obtaining CSIT is using feedback from the receiver of the forward link. The channel information is measured at the receiver at B during the forward link (A to B) transmission, then sent to the transmitter at A on the reverse link. In practice, the forward-link transmission from a base node includes pilot signals, received by all active users. These users can thus measure their respective receive channels. The required users then send their channel information on a reverse link back to the base node for use as their CSIT. The feedback communication can either be scheduled separately or piggybacked on on-going transmissions. In data communications, CSIT may be needed for only a subset of users, who are then scheduled to transmit their channel information.

Feedback is not limited by the reciprocity requirements. However, it imposes additional system overhead by using up transmission resources. Techniques to reduce the amount of feedback have been a subject of intense study, for example, designing vector codebooks, quantizing channel information, or selecting only the important information. See the conclusion for further discussion on this topic.

Furthermore, feedback information is susceptible to channel variation due to the delay in the feedback loop. The usefulness of

feedback depends on this delay and the channel Doppler spread. For a fast time-varying channel in mobile communications, feedback techniques are usually effective up to a certain mobile speed, depending on the carrier frequency, the transmission frame length, and the turn-around time. The effects of feedback delay and error have been analyzed for various precoding techniques in 3GPP [33], revealing potentially severe performance degradation. Therefore, the optimal use of feedback must account for the information quality.

APPLICATION AND OVERHEADS IN MIMO CSIT ACQUISITION

Both reciprocity and feedback methods are used in practical wireless systems, including time-division-duplex (TDD) and frequency-division-duplex (FDD). TDD systems may use reciprocity techniques. While the forward and reverse links in a TDD system often have identical frequency bands and antennas, there is a time lag between these two links. In voice systems, this lag is the ping-pong period; in asynchronous data systems, the lag is the scheduling delay between the reception of the signal from a user and the next transmission to that user. Such time lags must be negligible compared to the channel coherence time for reciprocity techniques to be applicable. FDD systems, on the other hand, usually have identical temporal and spatial dimensions on the forward and reverse links, but the link frequency offset (normally at 5% of the carrier frequency) is often much larger than the channel coherence bandwidth, making reciprocity techniques infeasible. FDD systems therefore commonly use feedback techniques.

An important practical issue is the pilot related overhead when using multiple antennas. While there is no penalty for multiple receive antennas, with the exception of transmit beam forming, multiple transmit antennas require additional pilot overhead proportional to the number of transmit antennas, if the receiver needs to learn the complete MIMO channel. In the case of transmit beam forming, this overhead can be avoided if the pilots are also beam-formed along with the signal (data associated pilots). In an open-loop system, the overhead is the product of the number of training pilots on the reverse link and the number of users participating in reverse channel sounding. In a closed-loop system, the overhead consists of both the training pilots and the feedback. The training overhead is independent of the number of users. The feedback overhead is proportional to the number of designated users on the reverse link multiplied with the size of their feedback information. For OFDM systems, the amount of feedback is further increased due to the multiple subcarriers. Exploiting frequency continuity by tone sampling can help reduce this overhead, making it sublinear in the number of OFDM subcarriers. The overhead comparison in open- vs. closed-loop systems typically favors open-loop. However, when the number of receive antennas on the forward link is much larger than the number of transmit antennas, closed-loop systems may be more efficient.

THE MIMO CHANNEL AND CSIT MODELING

A wireless channel exhibits time, frequency, and space selective variations, known as fading. This fading arises due to Doppler, delay, and angle spreads in the scattering environment [3], [34]. The channel spreading can be observed by sending a single impulse in frequency or time (CW signal) or angle (point source) through the channel and receiving a signal spread along the spectral, temporal, or spatial dimension, respectively. In this article, we focus on a time-selective channel, assuming frequency-flat and negligible angle-spread. A frequency-flat solution, however, can be applied to a frequency-selective channel by decomposing the transmission band into multiple narrow, frequency-flat subbands. Specifically, we can apply the solution per subcarrier in systems employing OFDM.

In a rich scattering environment, a frequency-flat MIMO wireless channel can be modeled as a complex Gaussian random process, represented as a time-varying matrix. The channel at a time instance is a Gaussian random variable, specified by the mean and its covariance. A nonzero channel mean signifies the presence of a direct line-of-sight or a cluster of strong paths, and the channel envelop has the Rician statistics, while zero mean corresponds to the Rayleigh statistics. The channel covariance, on the other hand, captures the correlation among the antennas at both the transmitter and the receiver. Assuming the channel is stationary, the channel temporal variation can be captured by the channel auto-covariance, measuring the correlation between two channel instances separated by a delay. At zero delay, the channel auto-covariance coincides with the channel covariance.

This article considers CSIT at the transmit time in the form of a channel estimate and the estimation error covariance, derived from a channel measurement at an initial time and the channel statistics [8]. Since the main source of irreducible error in channel estimation is the random time-variation of the channel between the initial measurement and its use by the transmitter, we assume that the initial channel measurement is error-free. The error in the channel estimate therefore depends only on the time delay and the channel time selectivity, or the Doppler spread.

Let $H(M \times N)$ denote the channel matrix in a system with N transmit and M receive antennas. The channel has mean $\bar{H}$ and covariance R_0, defined as

$$\begin{aligned} \bar{H} &= E[H] \\ R_0 &= E[hh^*] - \bar{h}\bar{h}^*, \end{aligned} \tag{1}$$

where the lower-case letter denotes the vectorized version of the upper-case matrix variable, and $(.)^*$ denotes conjugate transpose. Assume that we have an initial, accurate channel measurement H_0. The channel auto-covariance R_s at time delay s then indicates the correlation between this initial measurement H_0 and the current channel H_s, defined as

$$R_s = E[h_s h_0^*] - \bar{h}\bar{h}^*. \tag{2}$$

Intuitively, when this correlation is strong (R_s is large when measured in a suitable norm) then H_0 is useful for estimating H_s. The strongest correlation is when the delay is 0; that is, if $s \to 0$, then $R_s \to R_0$. In a scalar system, R_s and R_0 reduce to scalars r_s and r_0, respectively. They are related as $r_s = \rho(s) r_0$, where $|\rho(s)| \leq 1$ is the temporal correlation coefficient.

We now make an important assumption about channel temporal homogeneity. We assume that the temporal correlation coefficient $\rho(s)$ between any pair of transmit and receive antennas is identical. This assumption is based on the premise that the channel temporal statistics can be expected to be the same for all antenna pairs. It is now possible to separate the temporal correlation from the spatial correlation in the channel auto-covariance as

$$R_s = \rho(s) R_0. \tag{3}$$

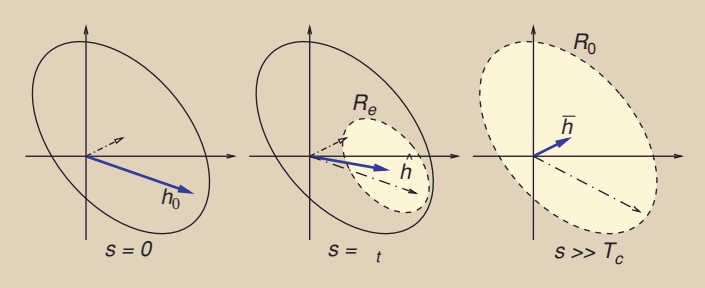
The temporal correlation ρ is a function of the time delay s and the channel Doppler spread. In Jake's model for example, $\rho(s) = J_0(2\pi s f_d)$, where f_d is the channel Doppler spread and $J_0(.)$ is the zero$^{\text{th}}$-order Bessel function of the first kind [35].

An estimate of the channel at time s together with the estimation error covariance then follow from the minimum mean squared error (MMSE) estimation theory [36] as

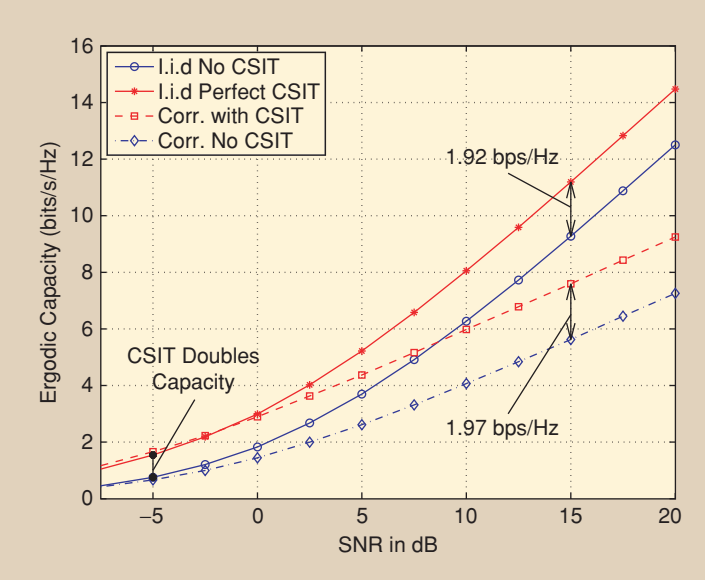
$$\begin{aligned} \hat{H} &= \rho H_0 + (1-\rho)\bar{H} \\ R_e &= (1-\rho^2) R_0. \end{aligned} \tag{4}$$

The two quantities $\hat{H}$ and R_e function effectively as a new channel mean and a new channel covariance, and thus are referred to as the effective mean and the effective covariance, respectively. Together, they constitute the CSIT. This CSIT ranges from perfect channel knowledge when $\rho = 1$ to pure statistics when $\rho = 0$. Since the CSIT depends on ρ which captures the channel time-variation, it is called dynamic CSIT. Here, ρ functions as a measure of CSIT quality. When $\rho = 1$, the channel estimate coincides with H_0 and is error-free. As ρ decreases to 0, the influence of the initial channel measurement diminishes, and the estimate moves toward the channel mean $\bar{H}$. In parallel, the estimation error covariance R_e is zero when $\rho = 1$, and grows to R_0 as ρ decreases to 0. Figure 1 illustrates this CSIT evolution as a function of the time delay s.

Several special cases of dynamic CSIT are of interest. First is perfect CSIT, in which the effective covariance is zero, and the effective mean is the instantaneous channel. Second is mean CSIT, in which the effective mean is nonzero and arbitrary, but the effective covariance is the identity matrix, corresponding to uncorrelated antennas. Third is covariance CSIT, in which the effective covariance matrix is nonidentity and arbitrary, but the effective mean is zero, corresponding to Rayleigh fading. The general case in which both the mean and covariance matrices are arbitrary is referred to as statistical CSIT (at a given ρ).



[FIG1] Dynamic CSIT model.



[FIG2] Capacity of 4 × 2 Rayleigh fading channels without and with perfect CSIT.

BENEFITS AND OPTIMAL USE OF CSIT

In a frequency-flat MIMO channel, CSIT can be exploited in both the spatial and temporal dimensions, in contrast to the scalar case, in which only temporal CSIT is relevant. It is well known that temporal CSIT—channel information across multiple time instances—provides little capacity gain, which becomes negligible at medium-to-high SNRs (approximately above 15 dB) [9]. Spatial CSIT, on the other hand, can offer a significant increase in channel capacity at all SNRs.

Figure 2 provides an example of the capacity increase based on spatial CSIT for two 4×2 Rayleigh fading (zero-mean) channels. For the i.i.d channel, capacities with perfect CSIT and without are plotted. For the correlated channel with a rank-one transmit covariance matrix (and uncorrelated receive antennas), capacities with the covariance knowledge and without are shown. The capacity gain from CSIT at high SNRs here is significant, reaching almost 2 b/s/Hz at 15 dB SNR. At lower SNRs, although the absolute gain is not as high, the relative gain is much more pronounced. For both channels, CSIT helps to double the capacity at -5 dB SNR. Subsequently, exploiting spatial CSIT, particularly in the form of an effective channel mean and covariance (4), will be the focus of this article.

BENEFITS OF CSIT

The capacity gain from CSIT is different at low and high SNRs [8]. At low SNR, CSIT can help increase the ergodic capacity multiplicatively. The transmitter relies on the CSIT to focus transmit power only on strong channel modes, whereas without CSIT, the optimal strategy for ergodic capacity is to transmit with equal power in every

direction. For example, with perfect CSIT at low SNRs, only the strongest eigen-mode of the channel is used. The low-SNR capacity ratio r between perfect CSIT and no CSIT is given by

$$r = \frac{C_{\text{perfect CSIT}}}{C_{\text{no CSIT}}} = \frac{NE[\lambda_{\max}(HH^*)]}{\text{tr}(E[HH^*])}, \tag{5}$$

where N is the number of transmit antennas and tr(.) is the trace of a matrix. For an i.i.d. Rayleigh fading channel, as the number of antennas increases to infinity, provided the transmit to receive antenna ratio N/M stays constant, this ratio approaches a fixed value as

$$r \to \left(1 + \sqrt{\frac{N}{M}}\right)^2. \tag{6}$$

The ratio r is always larger than one and can be significant in systems with more transmit than receive antennas ($N > M$). Examples of the capacity ratio versus the SNR for several systems with twice the number of transmit as receive antennas are given in Figure 3. This ratio increases at lower SNRs and at larger numbers of antennas. For these systems, it asymptotically approaches 5.83.

With statistical CSIT, similarly, the CSIT helps to increase the low-SNR capacity multiplicatively. The capacity ratio between statistical CSIT and no CSIT is given by

$$r = \frac{C_{\text{statistical CSIT}}}{C_{\text{no CSIT}}} = \frac{N\lambda_{\max}(G)}{\text{tr}(G)}, \tag{7}$$

where $G = E[H^*H]$. Again, the statistical CSIT helps the transmitter to focus its energy along the dominant eigen-mode of G at low SNRs.

At high SNRs, the capacity gain from CSIT is incremental and dependent on the relative antenna configuration. For systems with equal or fewer transmit than receive antennas, the capacity gain from perfect CSIT diminishes at high SNRs, since the optimal input signal with CSIT then also approaches equipower. For systems with more transmit than receive antennas ($N > M$), however, CSIT helps increase the capacity even at high SNRs. Since the channel rank here is smaller than the number of transmit antennas, CSIT helps the transmitter direct the signal to avoid the channel null-space and achieve an incremental capacity gain at high SNRs as

$$\Delta C = M \log\left(\frac{N}{M}\right). \tag{8}$$

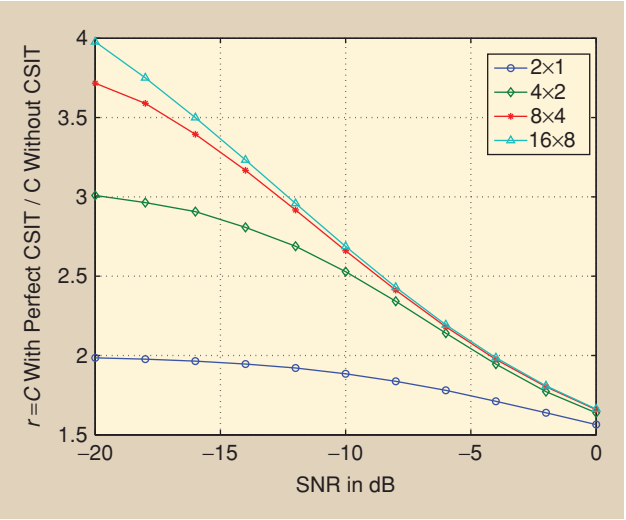
This gain is proportional to the number of receive antennas M and depends on the ratio of the number of transmit to receive antennas N/M. For example, for systems with twice the number of transmit as receive antennas, the capacity incremental gain approaches the number of receive antennas in bits per second per hertz and can be achieved at an SNR as low as 20 dB, as illustrated in Figure 4.

OPTIMAL USE OF CSIT

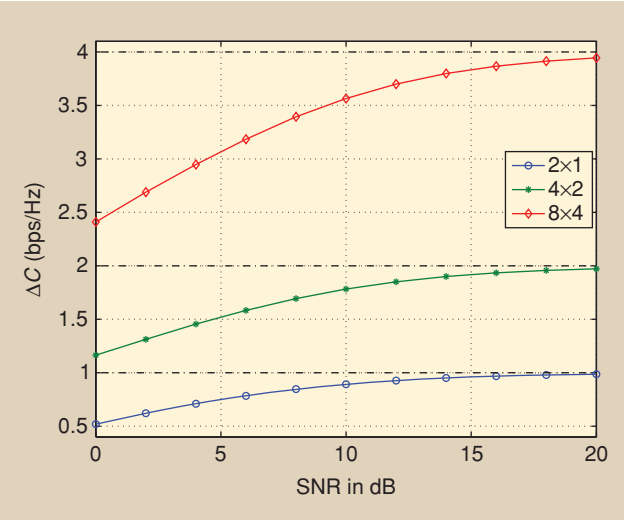
The optimal use of CSIT for achieving the capacity of a frequency-flat fading channel can be established by first examining the scalar channel [5]. Assume that the transmitter has causal channel state information $\text{U}_1^s = \{U_1, \ldots U_s\}$, provided that the channel is independent of the past CSIT given current CSIT

$$\Pr\left(h_s | U_1^s\right) = \Pr(h_s | U_s). \tag{9}$$

The channel capacity is then a stationary function of the current CSIT, but not dependent on the entire CSIT history. This condition covers the dynamic CSIT model (4). The receiver knows the channel perfectly, it also knows how the CSIT is used at the transmitter. Such assumptions are practically reasonable since the receiver can obtain channel information more readily than



[FIG3] Capacity ratio gain from perfect CSIT for i.i.d. channels. Asymptotically as the number of antennas increases, the ratio approaches 5.83.



[FIG4] Incremental capacity gain from perfect CSIT for i.i.d. channels. The dashed lines are the respective limits at high SNRs.

the transmitter, and they can both agree on a precoding algorithm. The capacity of the channel with CSIT (now denoted by U) can then be achieved by a single Gaussian codebook designed for a channel without CSIT, provided that the code symbols are dynamically scaled by a power-allocation function determined by the CSIT

$$C = \max_{f} E\left[\frac{1}{2}\log(1 + hf(U))\right], \tag{10}$$

where the expectation is taken over the joint distribution of h and U. In other words, the combination of this power-allocation function $f(U)$ and the channel creates an effective channel, outside of which coding can be applied as if the transmitter had no CSIT. This insight, in fact, can be traced back to Shannon in [4]. For a scalar fading channel, therefore, the optimal use of CSIT is for temporal power allocation.

This result has been subsequently extended to the MIMO fading channel [6]. Under similar assumptions, the capacity-optimal input signal with CSIT can be decomposed as the product of a codeword optimal for a channel without CSIT and a weighting matrix dependent on the CSIT. The optimal use of CSIT is now linear precoding, which allocates power in both spatial and temporal dimensions. In other words, the capacity-optimal signal is zero-mean Gaussian distributed with the covariance determined by means of the precoding matrix. This optimal configuration is shown in Figure 5.

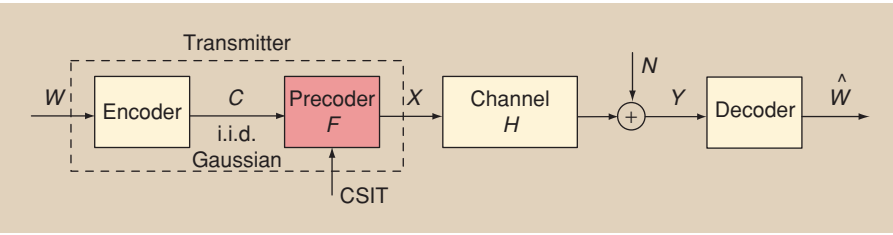
These results establish important properties of capacity-optimal signaling for a fading channel with CSIT. First, it is optimal to separate the function that exploits CSIT and the channel code, which is designed for a channel without CSIT. Second, a linear precoder is optimal for exploiting the CSIT. These separation and linearity properties are the guiding principles for MIMO frequency-flat precoder designs. In particular, this article focuses on designing a precoder based on the CSIT, given predetermined channel coding and detection technique. Before discussing about specific designs, however, the structure of a system with precoding is analyzed next.

PRECODING SYSTEM STRUCTURE

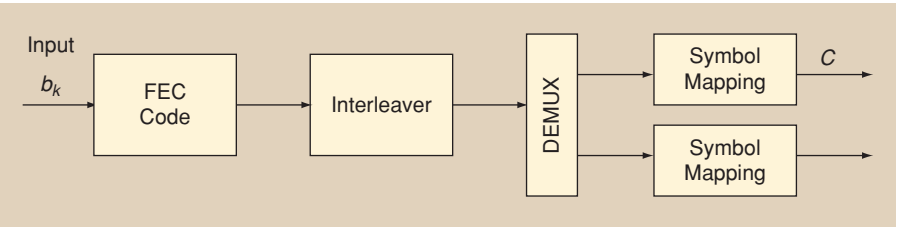
The transmitter in a system with precoding consists of an encoder and a precoder, as depicted in Figure 5. The encoder intakes data bits and performs necessary coding for error correction by adding redundancy, then maps the coded bits into vector symbols. The precoder processes these symbols before transmission from the antennas. At the other side, the receiver decodes the noise-corrupted received signal to recover the data bits, treating the combination of the precoder and the channel as an effective channel. The structures of these processing blocks are discussed in detail next.

ENCODING STRUCTURE

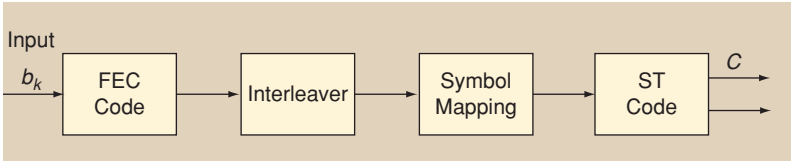
An encoder contains a channel coding and interleaving block and a symbol-mapping block, delivering vector symbols to the precoder. We classify two broad structures for the encoder: spatial multiplexing and ST coding, based on the symbol mapping block. The spatial multiplexing structure de-multiplexes the output bits of the channel coding and interleaving block to generate independent bit streams. These bit streams are then mapped into vector symbols and fed directly into the precoder, as shown in Figure 6. Since the streams are independent with individual SNR, per-stream rate adaptation can be used.



[FIG5] An optimal configuration for exploiting CSIT.



[FIG6] A multiplexing encoding structure.



[FIG7] A space-time (ST) encoding structure.

In ST coding structure, on the other hand, the output bits of the channel coding and interleaving block are first mapped directly into symbols. These symbols are then processed by a ST encoder (such as in [38], [39]), producing vector symbols as input to the precoder, shown in Figure 7. If the ST code is capacity lossless for a channel with no CSIT (for example, the Alamouti code for a 2×1 channel [38]), then this structure is also capacity optimal for the channel with CSIT. The ST coding structure contains a single data stream; hence, only a single rate adaptation is necessary. The rate is controlled by the FEC-code rate and the constellation design.

The difference between these two encoding structures therefore lies in the temporal dimension of the symbol-level code. Spatial multiplexing spreads symbols over the spatial dimension alone,

resulting in a one-symbol-long input block, while ST coding usually spreads symbols over both the spatial and the temporal dimensions. While these two structures have different implications on rate adaptation, this issue is not discussed in this article. Therefore, for precoding analysis and design, we will treat spatial multiplexing as a special case of ST coding with the block length of one. Assuming a Gaussian-distributed codeword C of size $N \times T$ with a zero mean, we define the codeword covariance matrix as

$$Q = \frac{1}{TP} E\left[CC^*\right], \tag{11}$$

where P is the transmit power (here we assume that the codeword has been scaled by the transmit power, so this definition provides the normalized covariance), and the expectation is taken over the codeword distribution. When C is spatial multiplexing, $Q = I$.

Of particular interest is ST block code (STBC), usually designed to capture the spatial diversity in the channel, assuming no CSIT. Diversity determines the slope of the error probability versus the SNR and is related to the number of spatial links that are not fully correlated [42]. High diversity is useful in a fading link since it reduces the fade margin, which is needed to meet required link reliability. A STBC can be characterized by its diversity order; a full-diversity code achieves the maximum diversity MN in a channel with N transmit and M receive antennas. There is, however, a fundamental trade-off between the diversity and the multiplexing orders in ST coding [43]. The multiplexing order relates to rate-adaptation; it is the scale at which the transmission rate asymptotically increases with the SNR. A fixed-rate system therefore has a zero multiplexing order. (Recently there has been new development of the diversity-multiplexing trade-off at finite [low] SNRs with a modified definition of multiplexing order [46].) Without CSIT, STBC design achieving the optimal diversity-multiplexing trade-off is an active research area (see [44], [45] for some examples). With CSIT, on the other hand, precoding focuses on extracting a coding gain (an SNR advantage) from the CSIT; hence it is independent of, and complementary to, the diversity-multiplexing trade-offs for ST codes.

LINEAR PRECODING STRUCTURE

The precoder is a separate transmit processing block from channel and ST coding. It depends on the CSIT, but a linear precoder has a general structure. A linear precoder functions as a combination of an input shaper and a multimode beamformer with per-beam power allocation. Consider the singular value decomposition (SVD) of the precoder matrix

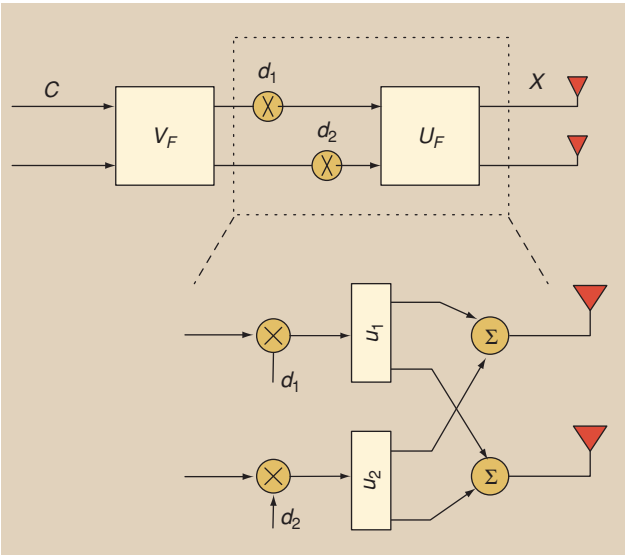
$$F = U_F D V_F. \tag{12}$$

The orthogonal beam directions are the left singular vectors U_F, of which each column represents a beam direction (pattern). Note that U_F is also the eigenvectors of the product FF^*, thus the structure is often referred to as eigen-beamforming. The beam power loadings are the squared singular values D^2. The right singular vectors V_F mix the precoder input symbols to feed into each beam and hence is referred to as the input shaping matrix. This structure is illustrated in Figure 8. To conserve the total transmit power, the precoder must satisfy

$$\mathrm{tr}(FF^*) = 1. \tag{13}$$

In other words, the sum of power over all beams must be a constant. The individual beam power, however, can differ according to the SNR, the CSIT, and the design criterion.

Essentially, a precoder has two effects: decoupling the input signal into orthogonal spatial modes, in the form of eigen-beams, and allocating power over these beams, based on the CSIT. If the precoded, orthogonal spatial-beams match the channel eigen-directions (the eigenvectors of H^*H), there will be no interference among signals sent on different modes, thus creating parallel channels and allowing transmission of independent signal streams. This effect, however, requires the full channel knowledge at the transmitter. With partial CSIT, the precoder tries to approximately match its eigen-beams to the channel eigen-directions and therefore reduces the interference among signals sent on these beams. This is the decoupling effect. Moreover, the precoder allocates power on the beams. For orthogonal eigen-beams, if all the beams have equal power, the total radiation pattern of the transmit antenna array is isotropic. Figure 9(a) shows an example of this pattern using a uniform linear antenna array. If the beam powers are different, however, the overall transmit radiation pattern will have a specific, noncircular shape, as shown in Figure 9(b). By allocating power, the precoder effectively creates a radiation shape to match to the channel based on the CSIT, so that higher power is sent in the directions where the channel is strong and reduced or no power in the weak. More transmit antennas will increase the ability to finely shape the radiation pattern and therefore will likely to deliver more precoding gain.



[FIG8] A linear precoder structure as a multimode beamformer.

RECEIVER STRUCTURE

Consider a system with an encoder producing a codeword C, and a precoder F at the transmitter, as shown in Figure 5. The codeword C is normalized according to the transmit power, which is constant over time, with zero mean and covariance as defined in (9). This codeword may contain channel coding, it may also represent only a ST codeword. An analysis for a system without a channel code is referred to as uncoded, otherwise it is coded. A system with ST coding alone thus qualifies for uncoded analysis. In this system, we assume that C is predetermined and hence is not a design parameter. In other word, the input codeword covariance Q (11) is given and fixed.

At the receiver, the received signal then is

$$Y = HFC + N, \tag{14}$$

where N is a vector of additive white Gaussian noise. The receiver knows à prior the precoding matrix F and treats the combination HF as an effective channel. It detects and decodes the received signal to obtain an estimate of the transmitted codeword C. The receiver can use one of several detection methods, depending on the performance and complexity requirements. Here we discuss two representative methods, maximum-likelihood (ML) and linear MMSE. ML detection is optimal, in which the receiver obtains the codeword estimate $\hat{C}$ as

$$\hat{C} = \arg\min_{C} ||Y - HFC||_F^2. \tag{15}$$

ML requires the receiver to consider all possible codewords before making the decision and hence can be computationally expensive. A simpler, although suboptimal, receiver is the linear MMSE. In this case, the receiver contains a weighting matrix W, which is designed according to

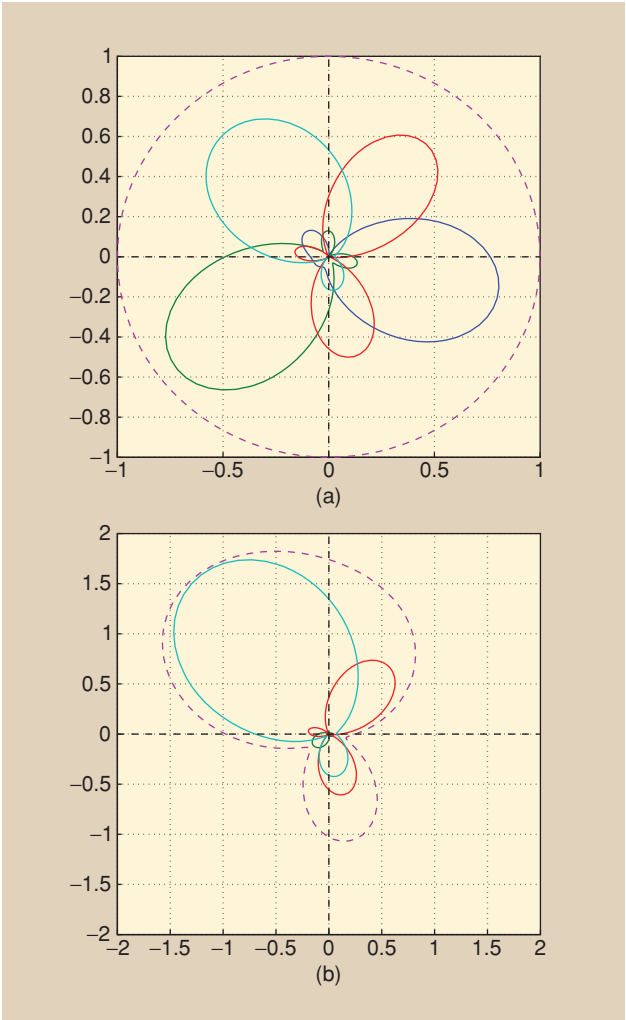
$$\min_{W} E||\hat{C} - C||_F^2 = E||(WHF - I)C + WN||_F^2, \tag{16}$$

where the expectation is taken over the input signal and noise distributions. For zero-mean signals with covariance in (11), the optimum MMSE receiver is given as

$$W = \gamma QF^*H^*(\gamma HFQF^*H^* + I)^{-1}, \tag{17}$$

where γ is the SNR. Due to its attractive simplicity, the linear MMSE receiver has often been used in designing a precoder [26]–[28]. A weighted MSE design, giving different weights to different received signal streams, can yield different criteria, such as maximum rate and target SNRs [26]. Other structures that are less computationally demanding than ML include the sphere decoder, successive cancellation receiver, and, if a channel code is present, iterative receiver iterating between the channel decoder and a simple symbol level detector (such as the MMSE).

In this article, however, to emphasize precoding at the transmitter and its potential gains, we assume the optimal ML receiver in the following analysis.



[FIG9] Orthogonal eigen-beam patterns of a uniform linear array with 4 transmit antennas and unit distance between them. (a) Equal beam power. (b) Unequal beam power. The purple dotted line is the total radiated pattern (of the four eigen-beams) from the antenna array.

PRECODING DESIGNS

The precoder connects between the encoder and the channel. Depending on the code used, the encoder produces codewords with a certain covariance Q. We assume that this encoder, and hence Q, is predetermined and is not a design target here. Such a configuration is supported by the optimal principle of separating the channel coding (assuming no CSIT) and precoding (exploiting the CSIT), discussed previously in the "Optimal use of CSIT" section. It includes the case $Q = I$, in which the input code can be capacity-optimal without CSIT and the precoder then represents a linear transmitter. Further motivation comes from the practical consideration of keeping the same channel and ST coding in an existing system and adapting the precoder alone to available CSIT. In all cases, the precoder transforms the codeword covariance into the transmit signal covariance. A precoder design essentially aims at producing the optimal signal covariance according to the CSIT and a performance criterion.

DESIGN CRITERIA

There are alternate precoding design criteria based on both fundamental and practical measures. The fundamental measures include the capacity and the error exponent, while the practical measures contain, for example, the PEP, detection MSE, SER, BER, and the received SNR. Fundamental measures usually assume ideal channel coding; the ergodic capacity implies that the channel evolves through all possible realizations over arbitrarily long codewords, while the error exponent applies for finite codeword-lengths. Analyses using practical measures, on the other hand, usually apply to uncoded systems and assume a quasistatic block fading channel. The choice of the design criterion depends on the system setup, operating parameters, and the channel (fast or slow fading). For example, systems with strong channel coding, such as turbo or low-density parity check codes with long codeword lengths, may operate at close to the capacity limit and thus are qualified to use a coded fundamental criterion. Those with weaker channel codes, such as convolutional codes with small free distances, are more suitable using a practical measure with uncoded analysis. The operating SNR is also important in deciding the criterion. As the SNR increases, the shortest-distance input pairs increasingly dominate the error rate, requiring coding for better average performance. Thus, a high SNR usually favors coded criteria for designing precoders, while at low SNRs, uncoded criteria can yield better performance.

Precoding design maximizing the channel ergodic capacity has been studied extensively for various scenarios: perfect CSIT [37], mean CSIT [7], [10]–[12], transmit covariance CSIT [7], [14], [16], both transmit and receive covariance CSIT [15], [17], and both mean and transmit covariance CSIT [8]. For more practical measures, many of the earlier designs focused on perfect CSIT, often jointly optimizing both a linear precoder and a linear decoder based on the MSE, the SNR, or the bit-error-rate (BER) (see [26]–[29] and references therein). More recent work considered partial CSIT. Precoding with mean CSIT was designed to maximize the received SNR [7], or minimize the SER [19], the MSE [20], or the PEP [18], [21]. Precoding with transmit covariance CSIT was similarly developed to minimize the PEP [22], the SER [23], or the MSE [24]. Precoding for both mean and transmit covariance CSIT has been developed to minimize the PEP [25]. In this article, we focus on two example criteria, one from each measure: the ergodic capacity and the PEP.

MAXIMIZING THE SYSTEM ERGODIC CAPACITY

The system ergodic capacity criterion aims at maximizing the average transmission rate with a vanishing error probability, assuming asymptotically long codewords and an ideal ML receiver. With perfect channel knowledge at the receiver, the capacity-optimal input signal is zero-mean Gaussian distributed with an optimal covariance [37]. For the system under study in Figure 5, the input codeword covariance Q is predetermined, hence we can only design the precoder F to produce a signal covariance that achieves the maximum system transmission rate, called the system capacity. This system capacity depends on Q. When Q is the capacity-optimal covariance for the channel without CSIT, then the system capacity coincides with the channel capacity; otherwise, it is strictly smaller.

With a given Q (11), the signal covariance for system in Figure 5 is $S = FQF^*$. The capacity-optimal precoder F then is the solution of the optimization problem

$$\begin{array}{ll} \max & E_H\left[\log\det(I + \gamma HFQF^*H^*)\right] \\ \text{subject to} & \operatorname{tr}(FF^*) = 1, \end{array} \tag{18}$$

where γ is the SNR. This formulation maximizes the mutual information, averaged over the channel distribution, subject to a transmit power constraint. Here the codeword covariance Q is predetermined and is not part of the design, and the constraint is over the precoder F alone. This constraint is based on the optimal separation between channel coding (assuming no CSIT) and precoding (exploiting the CSIT) as discussed in [5] and later generalized to MIMO in [6]. When $Q = I$, this constraint is the same as total transmit power constraint and the system capacity coincides with the channel ergodic capacity, such as the formulation in [13]. (When Q is a nonidentity, the two constraints on tr(FF*) and tr(FQF*) lead to a precoder with the same optimal beam directions; only the power loadings are different. However, we shall focus only on the tr(FF*) constraint in this article.) Note that in (18), the objective function usually cannot be simplified any further with partial CSIT and the optimization problem is stochastic.

MINIMIZING THE PAIR-WISE ERROR PROBABILITIES

The pair-wise error criterion, on the other hand, concerns the probability of a codeword $\hat{C}$ having a better detection metric at the receiver than the transmitted codeword C. In this case, a parameter of interest is the distance product between the two codewords

$$A = \frac{1}{P}(C - \hat{C})(C - \hat{C})^*, \tag{19}$$

which is related to the codeword covariance. With ML detection, the PEP can be upper-bounded by the well-known Chernoff bound (similar to [39])

$$P(C \to \hat{C}) \le \exp\left(-\frac{\gamma}{4} tr\,(HFAF^*H^*)\right), \tag{20}$$

which provides an analytical framework for precoding design. We consider two choices in minimizing the Chernoff bound on the PEP: minimizing for a chosen codeword distance A, and minimizing the average over the codeword distribution. The corresponding criterion is referred to as the PEP per-distance and the average PEP, respectively. In both cases, the performance averaged over channel fading is of interest.

For the PEP per-distance criterion, with a chosen A matrix, the precoder F is designed to minimize the Chernoff bound, averaged over the channel distribution as

$$\begin{array}{ll} \min & E\left[\exp\left(-\frac{\gamma}{4} tr(HFAF^*H^*)\right)\right] \\ \text{subject to} & \text{tr}\,(FF^*) = 1. \end{array} \tag{21}$$

For a fading channel with Gaussian distribution, the above objective function can be explicitly evaluated as a function of the channel mean and covariance [18]. In particular, for a channel with mean H_m and transmit antenna correlation R_t, but no receive correlation (i.e., $R_r = I$), the above problem is equivalent to [25]

$$\begin{array}{ll} \min & \text{tr}(H_m W^{-1} H_m^*) - M \log\det(W) \\ \text{subject to} & W = \frac{\gamma}{4} R_t F A F^* R_t + R_t \\ & \text{tr}(FF^*) = 1. \end{array} \tag{22}$$

In this case, the objective function becomes deterministic. The convexity of this problem, which helps in providing analytical solutions, depends on the distance matrix A (19). An often used A is the minimum codeword distance, which corresponds to the maximum PEP. For some codes, the minimum A is well-defined and can be a scaled-identity matrix, for which the problem has closed-form solution. Other choices of A include, for example, the average codeword distance. Depending on the code, the choice of A can significantly affect the performance of the resulting precoder.

For the average PEP criterion, the Chernoff bound is averaged over both the codeword distribution and the fading statistics. This average PEP criterion is independent of the specific codeword distance A (19). Noting that E[A] = 2Q (11), the precoder optimization problem in this case becomes

$$\begin{array}{ll} \min & E_H\left[\det\left(I + \frac{\gamma}{2} HFQF^*H^*\right)^{-M}\right] \\ \text{subject to} & \text{tr}(FF^*) = 1. \end{array} \tag{23}$$

Note the similarity between this formulation and the capacity formulation (18), both involve the expectation of functions of similar forms without a closed-form expression. Again, this formulation includes a predetermined code with covariance Q, and the constraint therefore is imposed over the precoder F alone (see [18]–[25]). When $Q = I$, the formulation becomes similar to those in [27]–[29] in the sense that F then represents the whole linear transmitter. Thus it can be thought of as a generalization of such setups to include a predetermined code with covariance Q.

CRITERIA GROUPING

In general, the precoder design problems can be divided into two categories, stochastic or deterministic. Stochastic optimization problem usually involves as the objective the expected value of a function over the channel distribution, in which the expectation has no closed-form expression [57]. Often, the function is convex in a matrix variable, for example, $\text{logdet}(.)^{-1}$, $\det(.)^{-1}$, or $\text{tr}(.)^{-1}$. While the statistical properties of the underlying channel distribution sometimes allow partial closed-form solution (such as the beam directions), the full solution usually requires numerical methods, in which the objective function is approximated by, for example, sampling or bounding. Deterministic problems, on the other hand, involves a deterministic objective function, obtained in closed-form from the problem formulation, with parameters given by the CSIT. Examples of stochastic problems include the capacity, the error exponent, the average PEP and the MSE criteria; while the deterministic includes the PEP per-distance, the SER, and the SNR criteria. (The connection among the mutual information, the sum MSE, and the Chernoff bound for STC, is recently analyzed in [58].) In both categories, some formulations lead to closed-form analytical precoder solutions, while others may require numerical optimal solutions (often the stochastic ones). Next, we will discuss typical precoder solutions for these problems with different CSIT scenarios.

OPTIMAL PRECODER DESIGNS

A linear precoder composes of an input shaping matrix, a beamforming matrix, and the power allocation over these beams, as discussed previously (12). For both criteria mentioned in the "Design Criteria" section, the capacity and the PEP, together with other criteria such as the error exponent, MSE, and SNR [56], the optimal input shaping matrix is determined by the input code alone, the beamforming matrix by the CSIT alone, and the power allocation by both. We first discuss the optimal input shaping matrix solution, which is independent of CSIT; then discuss the optimal beam directions and power allocation for different CSIT scenarios: perfect CSIT, covariance CSIT, mean CSIT, and statistical CSIT consisting of both mean and covariance information.

THE INPUT-SHAPING MATRIX

The encoder shapes the covariance (or the product distance matrix) of the codeword input to the precoder; the precoder in response chooses its input-shaping matrix to match this covariance. Suppose the input codeword covariance matrix Q (9) has the eigenvalue decomposition $Q = U_Q \Lambda_Q U_Q$, the optimal input-shaping matrix is then given by [55]

$$V_F = U_Q. \tag{24}$$

This optimal input-shaping matrix results directly from the predetermined input code covariance Q, which is not an optimization variable nor involved in the power constraint (13). The covariance Q characterizes the code chosen for the system. By matching the input codeword covariance, the precoder spatially de-correlates the input signal and optimally collects the input energy. In the special case of isotropic input ($Q = I$), such as with spatial multiplexing, the optimal V_F depends on the optimization criterion. For all aforementioned criteria, including the capacity, error exponent, MSE, PEP per-distance, average PEP, and SNR, V_F becomes an arbitrary unitary matrix and can usually be omitted. For some other criteria (which can be characterized using Schur convexity [54]), such as minimizing the maximum MSE among the received streams or minimizing the average BER, however, the optimal input-shaping matrix with $Q = I$ must be a specific rotational matrix [28], [29]. When channel coding such as a turbo-code is considered with a practical constellation, a rotational matrix can also improve performance [31].

THE BEAMFORMING MATRIX

Unlike the input-shaping matrix, which is independent from the CSIT, the beamforming matrix is a function of the CSIT. We now present the optimal beamforming solutions for the CSIT models developed previously: perfect CSIT, mean CSIT, covariance CSIT, and statistical CSIT.

With Perfect CSIT

Given perfect CSIT, the MIMO channel can be decomposed into independent and parallel additive-white-noise channels [37]. The number of parallel channels equals the minimum between the numbers of transmit and receive antennas. These parallel channels are established by first performing the SVD of the channel matrix as

$$H = U_H \Sigma_H V_H^*, \tag{25}$$

then multiplying the signal at the transmitter with V_H and at the receiver with U_H. The parallel channels can be processed independently, each with independent modulation and coding, allowing per-mode rate control and simplifying receiver processing.

The optimal beam directions with perfect CSIT for all aforementioned criteria are matched to the channel right singular vectors as

$$U_F = V_H. \tag{26}$$

The optimality can be established using matrix inequalities that show function extrema obtained when the matrix variables have the same eigenvectors [54]. Therefore, the optimal beam directions are given by the eigenvectors of H^*H, or the channel eigen-directions. For multiple-input single-output (MISO) systems, the solution reduces to the well-known scheme: transmit maximum-ratio-combined (MRC) single-mode beamforming [35]. These optimal beam directions are independent of the SNR.

Consequently, the optimal precoder matrix for perfect CSIT, under all criteria and at all SNRs, has the left and right singular vectors determined separately by the eigenvectors of the channel gain H^*H and the input codeword covariance Q, respectively. Therefore, the precoder spatially matches both sides. It effectively re-maps the spatial directions of the input code into those optimally matched to the channel given the CSIT, as shown in Figure 10.

With Mean CSIT

Mean CSIT composes of an arbitrary effective mean matrix H_m and an identity effective covariance. This model can correspond to an uncorrelated Rician channel or to a channel estimate with uncorrelated errors. Let the SVD of H_m be $H_m = U_m \Sigma_m V_m^*$, then the optimal precoding beam directions for all criteria are given by the right singular vectors of this effective mean

$$U_F = V_m. \tag{27}$$

The proof for the capacity criterion can be found in [10]–[12] and can be extended to other stochastic formulations. The proof for the PEP criterion, which has a deterministic formulation, is first established in [18].

Note that these directions are also the eigenvectors of $H_m^* H_m$. In effect, because the identity channel covariance is isotropic, the channel mean eigen-directions become the statistically preferred directions. They are the channel eigen-directions on average, and signaling along these directions is optimal.

With Covariance CSIT

Covariance CSIT composes of a zero effective-mean and an arbitrary effective-covariance. From the model developed, the effective covariance is a linear function of the antenna correlation matrix. This matrix captures the correlations among all the transmit antennas, among all the receive antennas, and between the transmit and receive antennas. A common, simplified correlation model assumes that the transmit and the receive antenna arrays are uncorrelated, often occurred when these arrays are sufficiently far apart with enough random scattering between them [47]. The transmit antenna correlation R_t and the receive antenna correlation R_r can then be separated according to a Kronecker structure as

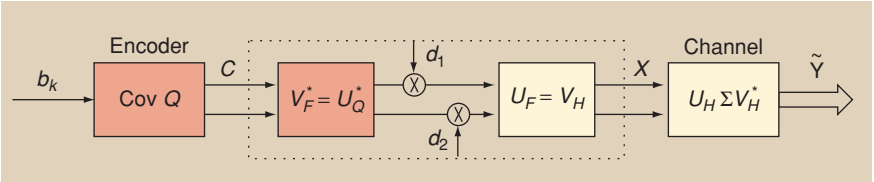
$$R_0 = R_t^T \otimes R_r. \tag{28}$$

This Kronecker correlation model has been experimentally verified for indoor channels of up to 3×3 antennas [48], [49], and for outdoor of up to 8×8 [50]. More general antenna correlation models have also been proposed in [51], [52], in which the transmit covariances (R_t) corresponding to different reference receive antennas are assumed to have the same eigenvectors, but not necessarily the same eigenvalues; similarly for R_r.

The optimal beamforming matrix has been established for covariance CSIT assuming the Kronecker correlation model (19). Furthermore, since precoding is primarily affected by transmit correlation, we assume uncorrelated receive antennas ($R_r = I$) in most cases, unless otherwise specified. Let the eigenvalue decomposition of R_t be $R_t = U_t \Lambda_t U_t^*$, then the optimal beamforming matrix for all criteria is given by the transmit correlation eigenvectors

$$U_F = U_t. \tag{29}$$

The proof for the capacity criterion can be found in [10] and [14] and for the PEP criterion, in [22]. The techniques in these proofs can be applied to other criteria.



[FIG10] The precoder matches both the input code structure and the channel.

Thus for a zero-mean channel, the correlation between the transmit antennas dictates the beam directions: its eigenvectors are the statistically preferred directions. When the antennas are uncorrelated, the beamforming matrix becomes an arbitrary unitary matrix and can be omitted. For the channel capacity criterion, even if the receive antenna correlation exists ($R_r \neq I$), it has no effect on the optimal beamforming directions [15]. The optimal beam directions for a non-Kronecker correlation structure, however, is still an open problem.

With Statistical CSIT

For statistical CSIT involving an arbitrary effective-mean and an arbitrary effective-covariance matrix, the optimal beamforming matrix has been established for only a few criteria, the PEP per-distance and the SNR. For the PEP per-distance criterion (21), assuming transmit antenna correlation alone, if the input codeword is isotropic such that $Q = \mu_0 I$, the optimal beamforming matrix can be obtained as part of the optimal precoder as

$$FF^* = \frac{4}{\gamma \mu_0} \left(\Phi - R_t^{-1} \right) \tag{30}$$

where Φ is given by

$$\Phi = \frac{1}{2\nu} \left[MI + \left(M^2 I + 4\nu R_t^{-1} H_m^* H_m R_t^{-1} \right)^{1/2} \right] \tag{31}$$

in which ν is the Lagrange multiplier associated with the power equality constraint in (21). Solving for ν is carried out using a dynamic water-filling process [25]. This process is similar to the standard water-filling, in that at each iteration, the weakest eigen-mode of FF^* may be dropped to ensure its positive semi-definiteness, and the total transmit power is re-allocated among the remaining modes. There is, however, a significant difference in that the mode directions here also evolve at each iteration. Details of the algorithm solving for ν can be found in [25].

The optimal beam directions of (30) depend on both the channel mean and covariance and are complicated functions of the channel K factor and the SNR. At high K, the channel mean H_m tends to dominate the beam directions; but as K drops, the channel covariance $\mathrm{R_t^{-1}}$ has more dominant effect. The SNR also influences the beam directions here, in contrast to the previous special CSIT cases. At low SNR, the PEP-optimal beam directions depend on both the mean and the covariance, but as the SNR increases, they asymptotically depend on the covariance alone. This effect shows that at high SNRs, the channel variation becomes more dominant in affecting the precoder.

For the SNR criterion, on the other hand, the precoder aims to maximized the received SNR by single-mode beamforming at all SNRs, with the beam as the dominant eigenvector of the average channel gain $E[H^*H]$.

For other criteria such as the capacity, a suboptimal solution for the beamforming matrix with statistical CSIT can be obtained by using the eigenvectors of the average channel gain $E[H^*H]$. At low SNRs, this solution is asymptotically capacity-optimal, while also being optimal for the PEP and SNR criteria. At high SNRs, if the number of transmit antennas is no more than the receive, it is also asymptotically capacity-optimal since the optimal input then becomes isotropic with an arbitrary set of beams. (With more transmit than receive antennas, however, the capacity-optimal solution may still require specific beamforming with unequal power among the beams at all SNRs, for example, when there is a strong antenna correlation or a strong channel mean [8].) Note that when applied to the special cases, mean CSIT and covariance CSIT, these beamforming directions become optimal.

THE POWER ALLOCATION

In contrast to the beam directions, the optimal power allocation across the beams varies for each design criterion and is a function of the SNR. With perfect CSIT, for example, it varies from water-filling for capacity to single-mode for the PEP criterion. The difference reflects the selectivity in power allocation, in which the more selective scheme allocates power to fewer modes at the same SNR. The power allocation tends to become more selective when the criterion shifts from fundamental (coded) towards practical (uncoded). In other words, this selectivity depends on the strength of the channel code. Systems with strong codes tend to allocate power to more channel eigenmodes, while those with weak codes tend to activate fewer, only strong modes, and drop the rest at the same SNR.

CSIT also affects the optimal power allocation. With perfect CSIT, the optimal power allocation is known in analytical closed-form for all criteria; while for partial CSIT, the solution may require numerical methods, depending on the criteria. However, the optimal power allocation often follows the water-filling principle, in which higher power is allocated to the beams corresponding to known strong channel directions, and reduced or no power to the weak. Next, we discuss the power solution each CSIT scenario, perfect CSIT, mean CSIT, covariance CSIT, and statistical CSIT.

With Perfect CSIT

As established in the previous two sections, the precoder with perfect CSIT matches to the input codeword covariance Q on the one side and to the channel H on the other. Because of this direction matching, the optimal power allocation depends only on the eigenvalues of both the input codeword covariance and the channel, but not their eigenvectors. Denote the eigenvalue product of these two matrices as

$$\lambda_i = \lambda_i(H^*H)\lambda_i(Q), \tag{32}$$

where the eigenvalues of each matrix are sorted in the same order. The power p_i allocated to beam number i, which is the square of the precoder singular value number i, is a function of these λ_i and the SNR.

For the capacity criterion (18), the optimal power allocation is obtained through water-filling on the composite eigenvalues λ_i as [37]

$$p_i = \left(\mu - \frac{N_0}{\lambda_i} \right)_+ , \tag{33}$$

where N_0 is the noise power per spatial dimension, and μ is chosen such that the sum of all p_i equals the total transmit power. Notation $(.)_+$ represents the value inside the parenthesis if this value is positive, and zero otherwise.

Similarly, for the average PEP criterion (23), the optimal power allocation is water-filling as for the capacity, but with the noise scaled-up by a factor of two. This solution thus is a more selective power allocation scheme. At low SNRs, weak modes tend to have a high error rate; therefore, dropping these modes and allocating power to stronger modes leads to better overall system error performance. As the SNR increases, power is allocated across more modes, but again, at a slower rate than is the case for the capacity solution.

For the PEP per-distance criterion (21), the optimal solution is to allocate all power to the strongest eigen-mode of the channel,

$$p_1 = 1, \quad \text{and} \quad p_i = 0 \quad \text{for} \quad i \neq 0\,, \tag{34}$$

thus effectively reducing the precoder to single-mode beamforming. This scheme is an extreme case of selective power allocation; it also maximizes the received SNR. Furthermore, it achieves the full transmit-diversity (see proof in [3], Section 5.4.4).

Perfect CSIT usually simplifies the power allocation problem significantly and allows for closed-form solution for most criteria (for other examples, see [24], [26]–[29]). With partial CSIT, however, the power allocation often requires numerical solutions, especially with the stochastic problems.

With Mean CSIT

With mean CSIT, the power allocation depends only on the singular values of the effective channel mean, but not its singular vectors. The capacity criterion (18) requires numerical, convex search for the optimal power. For the PEP per-distance criterion (21), the power allocation has a semi-analytical solution, obtained as a form of water-filling [55]

$$p_i = \left[\frac{1}{2\nu}\left(M + \sqrt{M^2 + 16\nu\frac{\lambda_i\left(H_m^* H_m\right)}{\gamma\,\lambda_i(A)}}\right) - \frac{4}{\gamma}\right]^+ \tag{35}$$

where $\lambda_i(.)$ are the eigenvalues of the corresponding matrix, sorted in the same order, and ν is the Lagrange multiplier associated with the equality power constraint. Simple binary search algorithm for finding ν can be found in [55], [56]. The solution with $A = I$ can also be found in [18].

For all criteria, the channel K factor and the rank of the mean matrix can have a strong influence here. A larger K factor causes the power allocation to depend strongly on the channel mean; for example, a rank-one mean then is likely to result in single-mode beamforming. Specifically for the channel capacity criterion [(18) with Q = I], if the K factor is above a certain threshold increasing with the SNR, single-mode beamforming is optimal for MISO systems [7]. When K approaches infinity, mean CSIT becomes equivalent to perfect CSIT. As K decreases, however, the impact of the channel mean diminishes. If K reduces to zero, the optimal allocation approaches equipower, hence the precoder becomes an arbitrary unitary matrix and can be omitted.

With Covariance CSIT

With covariance CSIT, in which the antenna correlation has a Kronecker structure, the optimal power allocation depends only on the eigenvalues of the correlations, but not their eigenvectors. Both transmit and receive correlation eigenvalues affect the optimal power allocation for the capacity criterion [15], which requires convex numerical solving techniques. For the PEP per-distance criterion (21) without receive correlation, the optimal power allocation can be obtained analytically by water-filling over the transmit correlation eigenvalues [22]

$$p_i = \left(\mu - \frac{4}{\gamma}\lambda_i^{-1}(A)\lambda_i^{-1}(R_t)\right)_+ \tag{36}$$

where $\lambda_i(.)$ are the (nonzero) eigenvalues of the corresponding matrix, and μ is chosen such that the sum of all p_i equals the total transmit power.

For all criteria, the stronger the antenna correlation, measured by a larger condition number for example, the more selective the optimal power allocation becomes. An extreme case of selectivity is single-mode beamforming. Thresholds for its optimality are observed for covariance CSIT in MIMO systems [14], [15], in which two largest eigenvalues of R_t must satisfy an inequality related to the dominance of the largest eigenvalue. Intuitively, if this largest mode is sufficiently dominant, then water-filling will drop all other modes. At higher SNRs, the required eigenvalue dominance must increase, implying a more correlated channel. A similar trend is observed for an increasing number of receive antennas. If R_t is full-rank, however, the capacity-optimal power allocation for systems with equal or fewer transmit than receive antennas always asymptotically approaches equipower as the SNR increases.

Furthermore, the impact of correlation, particularly the eigenvalues of R_t, on the average mutual information under different CSIT conditions—perfect, covariance, and no CSIT—can be described using majorization theory [16]. Transmit antenna correlation generally reduces the channel ergodic capacity at high SNRs, compared to an i.i.d channel, but the loss is bounded as the number of transmit antennas increases. At low SNRs, on the other hand, transmit correlation can help increase the capacity (see [8] and references therein).

With Statistical CSIT

For the ergodic capacity (18), as with most stochastic criteria, in contrast to the beamforming matrix, the optimal power allocation is so far unavailable in closed-form for statistical CSIT, including both mean and covariance CSIT as special cases. It often requires a numerical solution, which can usually be efficiently implemented because of the convexity of the problem [53]. The power solution now depends on both the eigenvalues and the eigenvectors of the mean and covariance

matrices. At low SNRs, the optimal power allocation concentrates all power in a single beam, often the dominant eigenvector of $E[H^*H]$. As the SNR increases, transmit power is spread to an increasing number of beams to a maximum that depends on the antenna configuration and the CSIT parameters. With statistical CSIT, a N-transmit antenna system can activate up to N orthogonal beams. When there are equal or fewer transmit than receive antennas ($N \leq M$), all N beams will be activated and the allocation approaches equipower at high SNRs, at which the precoder can usually be omitted. When there are more transmit than receive antennas ($N > M$), however, CSIT parameters strongly influence the optimal power allocation. Channels with a strong mean or a strong transmit antenna correlation may activate only a fraction of the beams (fewer than N) even at high SNRs. Simple thresholds on the channel K factor and the transmit covariance condition number for mode-dropping at all SNRs can be derived [8]. For a transmit covariance matrix with two levels of eigenvalues, for example, mode-dropping always occurs when its condition number satisfies

$$\kappa \geq \frac{L}{L-M} \tag{37}$$

where L is the number of stronger eigenmodes, provided that $N > L > M$. Using the inverted noncentral complex Wishart distribution, a threshold on the channel K factor can also be established, independent of the number of receive antennas [8].

For the PEP criterion (21), the optimal power, as part of the optimal precoder, has a semi-analytical solution given in (30), obtained using a dynamic water-filling algorithm [25], in which both the beam power and the beam direction evolve with the water-filling iterations. The asymptotic behavior of this precoder when the channel K factor or the SNR increases is worth noting. When K increases, the precoder converges to a solution dependent on the channel mean alone; furthermore, it becomes a single-mode beamformer aligned to the dominant right-singular-vector of H_m, hence maximizing the received signal power. As the SNR increases, however, the precoder approaches a solution dependent on the transmit correlation alone, and the power allocation approaches equipower. If both the K factor and the SNR increase, then there exists a K factor threshold increasing with the SNR, above which the optimal power allocation results in a single-beam precoder. In "Precoding with Dynamic CSIT," we show an example of this single-beam threshold.

DISCUSSION

The presented precoding designs lead to several observations. First, the optimal input shaping matrix, composing of the precoder right-singular-vectors, is the same for all CSIT scenarios at all SNR. It is matched to the covariance of the precoder input signal. This input shaping matrix ix optimal for most criteria, including the ergodic capacity, the PEP, and others such as the SNR, MSE, or error exponent [56]. When the precoder input covariance is the identity matrix, the optimal input shaping matrix becomes an arbitrary unitary matrix and can be omitted for these criteria; but for some others, a fixed-rotation matrix is required [28]. Second, the beamforming matrix, composing of the precoder left-singular-vectors, is independent of the design criteria and the SNR for most CSIT scenarios, except the general statistical CSIT case. These optimal beam directions are matched to the channel according to the CSIT, often as the eigenvectors of the channel mean or transmit covariance matrix. When both the mean and transmit covariance are present, however, the beam directions becomes dependent on the criterion and the SNR. Third, the main difference among the precoding solutions under different criteria is the power allocation. For both the ergodic capacity and the PEP criteria, the optimal power allocation follows the water-filling principle, in which higher power is allocated to stronger modes and reduced or none to weaker ones as a function of the SNR. This power selectivity, however, depends on the criterion. More selective schemes tend to drop more modes at low SNR. For examples, the selectivity increases going from the capacity to the PEP criterion. As the SNR increases, most power allocation schemes approach equipower, but at different rates. A more selective scheme approaches equipower more slowly. Schemes that do not approach equipower at high SNR occur under the PEP criterion with perfect CSIT, or generally with statistical CSIT involving a strong mean or a strong antenna correlation in channels with more transmit than receive antennas. Power allocation according to the water-filling principle also applies to other criteria such as the error exponent [56], the SER [19], [23], but by no means to all criteria. The MSE criterion, for example, tends to allocate more power to weaker channel modes [24]. The beam power allocation depends strongly on the performance criterion and the SNR and can be the main factor in differentiating the performance of different precoders.

A linear precoder therefore has two main effects: decoupling the signals into orthogonal spatial directions to reduce the interference between them, and allocating power to these directions according to the channel strength. In short, the precoder optimally collects the input signal power and spatially redistributes this power into the channel according to the design criterion and the CSIT.

The water-filling type power allocation leads to mode-dropping at low SNR. For a practical constellation, care should be taken in system design to ensure that the employed encoder functions in such a situation, especially for high rate codes. In most cases, the input-shaping matrix combines input codeword symbols such that all symbols are transmitted even with mode-dropping. With an identity codeword covariance ($Q = I$), even though the input shaping matrix can theoretically be omitted, some rotation matrix may still be necessary for practical constellations to ensure the transmission of all distinct symbols. An initial study of this rotation effect for spatial multiplexing can be found in [31]. The precoder input shaping matrix, thus, helps to prevent the adverse effect of mode-dropping with practical constellations on the system performance.

PRECODING NUMERICAL PERFORMANCE

SIMULATION SETUP

The simulation system has four transmit and two receive antennas and employs the quasiorthogonal STBC [40], [41]. Although a 4×2 system can support up to a spatial rate of two, this STBC has only the spatial rate one. With this STBC, the precoder input shaping matrix (24) is the identity matrix and is omitted. The system employs the [133, 171] convolutional code with rate one-half, used in the IEEE 802.11a wireless LAN standard, a block interleaver, and QPSK modulation. The receiver uses ML detection and a soft-input soft-output Viterbi decoder.

System performance is measured for two representative CSIT scenarios: perfect CSIT, and dynamic CSIT (4) involving both channel mean and transmit covariance information. Assume quasistatic block-fading channels, the block-length for the perfect CSIT is 96 bits, and for dynamic CSIT is 48 bits. Performance without and with different precoders, based on the PEP per minimum-distance (21), average PEP (23), and system capacity (18) criteria, are studied.

PRECODING WITH PERFECT CSIT

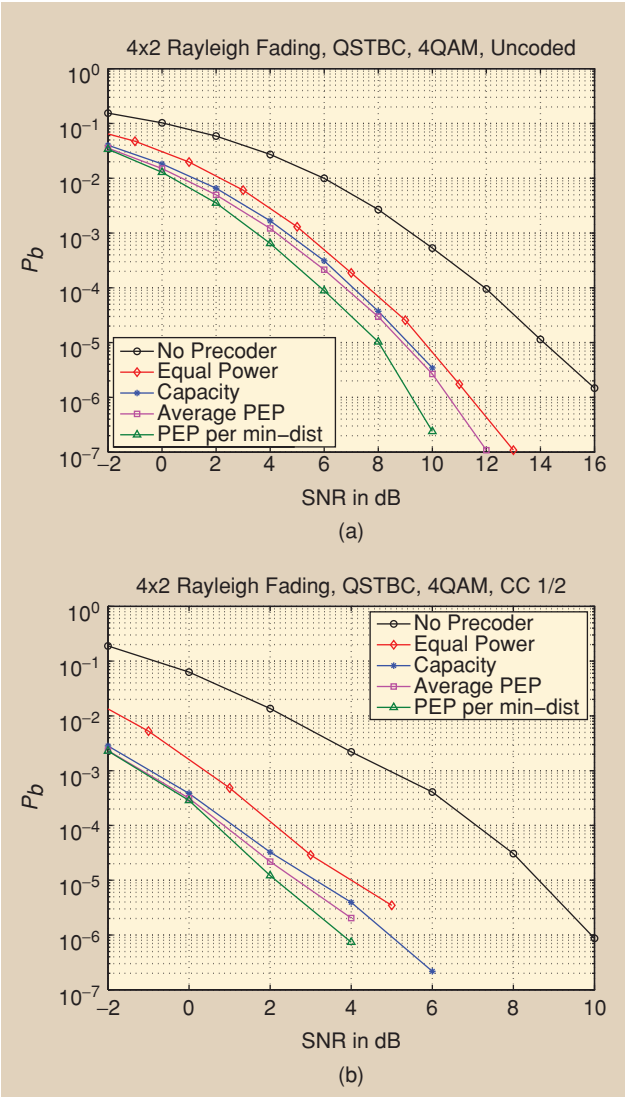
Precoding with perfect CSIT can be viewed as the ideal case for reference. Although precoding with perfect CSIT has been studied under different criteria, comparative performance of the different designs using the same system setup can draw some useful observations. For these simulations, the channel is assumed to be i.i.d. Rayleigh fading ($H_m = 0$ and $R_0 = I$).

Figure 11 shows the error rate performance of the different precoders. All three precoder designs achieve substantial gains, measured in both uncoded (without the convolutional code) and coded (with the convolutional code) domains, with larger gain in the latter (up to 6 dB SNR gain at 10^{-4} coded bit-error-rate). Such a gain is consistent with the analytical capacity gain (8). Since the QSTBC provides only partial diversity, some additional diversity gain is obtained by the precoder, evident through the higher slopes of the precoded error curves in the uncoded systems. In both uncoded and coded systems, however, most of the precoding gain appears in the form of a coding gain. This coding gain is attributed to the optimal beam directions and the water-filling-type power allocation. To differentiate the gain from each effect, a two-beam precoder with the optimal directions, given by the channel right singular vectors, but equal beam-power allocation is also studied. Results show that with perfect CSIT, optimal beam directions alone achieve a significant portion of the precoding gain. A water-filling-type power allocation further improves the gain, especially at low SNRs. Thus, both the precoder beam directions and the power allocation contribute to the performance gain.

These results also reveal only minor performance differences among precoder designs according to the three criteria. The minimum-distance PEP precoder, which also maximizes the received SNR, achieves the best gain here, attributed to the small number of receive antennas. The other two precoders, based on the capacity and the average PEP criteria, perform similarly. This relative performance order is dependent on the CSIT and the system configurations, including the number of antennas, channel coding, and the STBC; it may change for a different system.

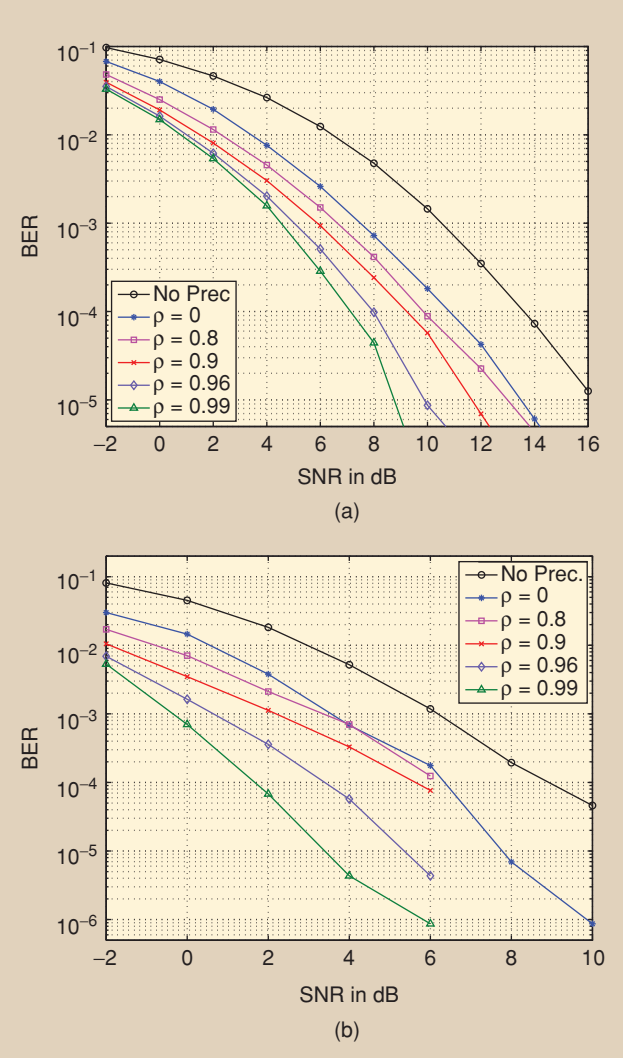
PRECODING WITH DYNAMIC CSIT

This section examines precoding performance with dynamic CSIT. For the system capacity and average-PEP criteria, unfortunately, no analytical solutions exist for the optimal precoders. The optimal precoder based on the minimum-distance PEP (30) is used. The transmit correlation matrix has the eigenvalues [2.717, 0.997, 0.237, 0.049], representing a relatively strong correlation with the condition number of 55.5, and the channel mean has $K = 0.1$. System performance is obtained for different values of the estimate quality ρ between 0 and 1. The error probabilities are averaged over multiple initial channel measurements H_0, randomly and independently drawn from the simulated channel distribution, and multiple channel estimates given each initial measurement.



[FIG11] Precoding performance with perfect CSIT. (a) Uncoded and (b) coded.

Performance results for different values of ρ are given in Figure 12. The precoding gain increases with a better CSIT quality. Depending on ρ, the gain ranges between statistical CSIT and perfect CSIT gains. When $\rho = 0$, the precoder achieves a performance gain based on statistical CSIT alone (channel mean and covariance information); as ρ approaches 1, the precoding gain increases to that with perfect CSIT. It is also noted through simulations that in dynamic CSIT, the initial channel measurement H_0 helps increase the precoding gain over the statistical CSIT gain only when its correlation with the current channel is sufficiently strong: $\rho \geq 0.6$; otherwise, precoding on the channel statistics alone can extract most of the gain. Furthermore, when the CSIT is imperfect ($\rho < 1$), the precoder does not provide diversity gain, in contrast to the perfect CSIT case. This observation is confirmed by analyses showing that the high-SNR asymptotic BER slope is independent of ρ for $\rho \neq 1$ [55]. Thus with partial CSIT ($\rho < 1$), the precoder only achieves an SNR gain, and the system transmit diversity is determined by the ST code (and the channel code if that exists).



[FIG12] Performance with dynamic CSIT for a precoded 4×1 system using OSTBC. (a) Uncoded and (b) coded.

With perfect CSIT ($\rho = 1$), the precoder also delivers the maximum transmit diversity gain of order N.

For comparison, we also study a single-beam scheme that relies only on the initial channel measurement, shown in Figure 13. This scheme coincides with the optimal PEP precoder for perfect CSIT ($\rho = 0.99$ in the simulation). For other ρ values, however, the scheme performs poorly. It looses all transmit diversity regardless of the STBC and performs worse than no precoding at high SNRs. The optimal precoder exploiting dynamic CSIT, on the other hand, provides gain at all SNRs for all ρ. This result demonstrates the robustness of the dynamic CSIT model.

Figure 14 shows the regions of different number of active (non-zero power) precoding beams, as a function of the channel K factor and the SNR. A higher K factor leads to fewer beams, whereas a higher SNR leads to more beams. The thresholds in K factor for different beam regions increase with the SNR; at very low K factors, however, the regions appear to depend little on K but only on the SNR. Other design criteria may lead to different precoding beam regions.

These numerical results illustrate significant precoding gains. The gain depends on the CSIT, the number of antennas, the system configuration (encoder and receiver), and the SNR. It usually increases with better CSIT, quantified by the estimate quality ρ in the dynamic CSIT model, and with more antennas. The gain, however, is less dependent on the design criteria: similar BER performance among precoders based on different criteria has been observed numerically [56]. While with perfect CSIT, the precoders achieve the maximum transmit diversity, the main benefit of precoding in all CSIT scenarios comes from the SNR gain (also called the coding gain). Two factors contribute to the precoding gain: the optimal beam directions and the water-filling type power allocation. Both of these result in an SNR advantage.

PRECODING APPLICATIONS IN EMERGING WIRELESS STANDARDS

Precoding has been successfully integrated into the IEEE 802.16e standard for broadband mobile wireless metropolitan networks (WiMax). Both open- and close-loop techniques are included. In the open-loop technique, a subset of users are scheduled to transmit a sounding signal. The base station then estimates the channels for these users and determines the CSIT for precoding use after transmit-receive RF calibration. In the closed-loop technique, the precoder uses either an initial channel measurement or the channel statistics. The users measure the channel using the forward-link preambles or pilots, then feed back the best codeword, usually a unitary-fit, representing this channel measurement from a codebook, along with a time-to-live parameter. The precoder uses the unitary-fit until the time-to-live expires; thereafter, it relies on the channel statistics information, which is updated at a much slower rate and is always valid.

MIMO is expected to enter the IEEE 802.11n standard for wireless local area networks (WLAN), with support for both ST

coding and spatial multiplexing. The current precoding proposals use an open-loop method, based on the reciprocity principle implying that the best beam on reception must be the best beam for transmission. The base uses preformed beams for receiving and transmitting and records the beam(s) with the best signal strength on reception from each user, then uses the same beam(s) during the next transmission to that user.

The 3GPP standard uses a closed-loop beamforming technique, based on the quantized feedback of the channel phase and amplitude. Precoding is under discussion in High-Speed Downlink Packet Access (HSDPA) for mobile communication. Channel-sounding appears to be the preferred technique for obtaining CSIT.

CONCLUSION

MIMO LINEAR PRECODING

This article has provided an overview of linear precoding techniques for exploiting CSIT in single-user, frequency-flat MIMO wireless systems. It discusses principles and methods for acquiring the CSIT, including open- and closed-loop techniques, and related issues such as sources of error, system overhead, and complexity. A dynamic CSIT model is formulated as an estimate of the channel at the transmit time with the associated error covariance. Dynamic CSIT can be obtained using a potentially outdated channel measurement, the first- and second-order channel spatial statistics, and the channel temporal correlation, thereby taking into account the channel time-variation. This CSIT model delivers robust precoding gain for different CSIT qualities.

Information theoretic foundation establishes the optimality of a linear precoder in exploiting dynamic CSIT. A linear precoder functions as the combination of an input shaper and a multimode beamformer that contains orthogonal beam-directions, each with a defined beam-power. We discuss linear precoder solutions under different design criteria for several CSIT scenarios: perfect CSIT, mean CSIT, covariance CSIT, and statistical CSIT, as parts of the dynamic CSIT model. Simulation examples, using a spatial rate one QSTBC transmission, demonstrate that precoding can improve error performance significantly. For higher spatial rate transmissions (such as spatial multiplexing), although not discussed in this article, precoding can also improve the capacity and error performance at all SNRs for systems with more transmit than receive antennas, and at low SNRs for others.

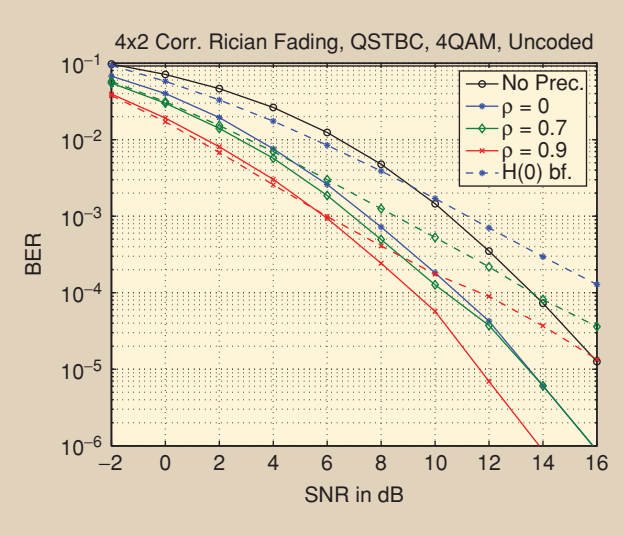
The essential value of precoding in exploiting CSIT is to add an SNR gain. This gain is achieved by the optimal eigenbeam patterns and the spatial power allocation across these beams. The optimal beam patterns (directions) can contribute to a significant part of the gain, but the power allocation becomes increasingly important as the SNR decreases. Both features help increase the transmission rate (the system capacity) and reduce the error probability. If the CSIT is perfect, precoding can also deliver a diversity gain; in addition, it helps reduce receiver complexity for higher spatial-rates by allowing parallel channel transmissions.

RELATED RESULTS

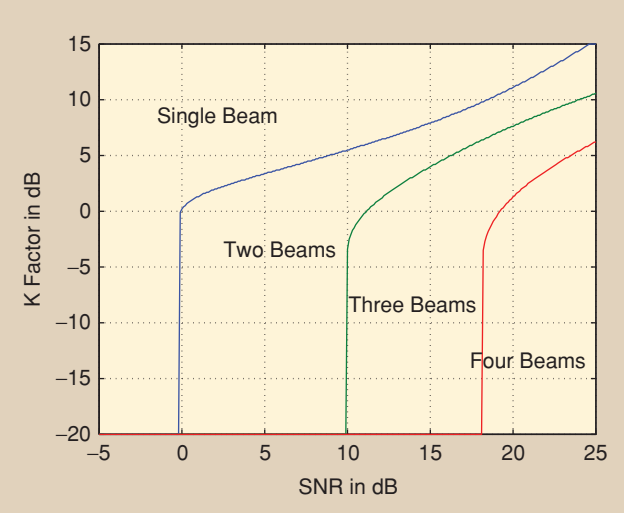
Looking beyond the scope of this article, precoding theory has been developed for other types of CSIT, limited-feedback scenarios, frequency-selective channels, and multiuser communications. We now briefly mention a few key ideas and selected references for interested readers.

While this article models CSIT as an estimate of the entire channel and its error covariance, there are also other types of less complete CSIT. For example, in a high-K channel, the K factor and the antenna-to-antenna phase statistics offer near-optimal precoding performance using beamforming and antenna power allocation [59]. A known channel condition number suggests adapting the transmission spatial rate [60]. These types of parametric CSIT help reduce the amount of overhead incurred in CSIT acquisition.

For feedback techniques in slow time-varying channels, the problem of overhead also motivates data compression techniques to minimize feedback. The compressed feedback



[FIG13] Performance comparison between optimal PEP precoding and H_0 beamforming.



[FIG14] Regions of different number of active precoding beams.

information can be, for example, selected and important channel information for precoding [33], the index of the precoder from a codebook [61], [62], or quantized channel information [63], [64] (and references therein). Often in such cases, the feedback information is tied to the specific precoding technique. The feedback overhead has motivated the area of finite-rate or limited feedback precoding (see [65] for an overview).

When the channel is frequency-selective, the precoder can also exploit this selectivity and become frequency-dependent. For single-carrier systems, nonlinear precoding techniques using spatial extensions of the Tomlinson-Harashima precoder can be employed [66], [67]. For multicarrier systems such as OFDM, frequency-flat precoding techniques discussed in this article can be applied on a per tone (subcarrier) basis. To reduce feedback overhead in OFDM, the CSIT feedback is sampled and interpolated in the frequency domain [68]. Exploiting the OFDM structure and tone correlation results in precoders with frequency-dependent eigen-beam directions and frequency-beam dependent power allocation [69], [70].

In wireless multiuser communications, partial CSIT is also highly relevant, since the channel time-variation makes it impractical to have perfect CSIT at all users. Initial research has shown that the loss of degrees of freedom due to no CSIT reduces the capacity region of an isotropic vector broadcast channel to that of a scalar one [71]. Imperfect CSIT also severely reduces the growth of the sum-rate broadcast capacity at high SNRs [72]. Schemes such as opportunistic scheduling, which requires only an SNR feedback, can achieve an optimal throughput growth-rate in broadcast channels with a large number of users [73]. With finite-rate feedback, however, the feedback rate needs to be increased with the SNR to achieve the full multiplexing gain [74].

With these results, precoding techniques that exploit partial CSIT continue to be an important research area with direct practical applications.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for closely reading the manuscript and providing detailed comments, which helped improve the quality of this paper. This work was supported in part by the Rambus Corporation Stanford Graduate Fellowship and the Intel Foundation Ph.D. Fellowship. It was also supported in part by NSF Contract DMS-0354674-001 and ONR Contract N00014-02-0088.

AUTHORS

Mai Vu (maivu@seas.harvard.edu) received a B.Eng. degree in computer systems engineering from the Royal Melbourne Institute of Technology (RMIT), Australia in 1997, an M.S.E. degree in electrical engineering from the University of Melbourne, Australia in 1999, and M.S. and Ph.D. degrees, both in electrical engineering from Stanford University, CA in 2006. She is currently a Lecturer in engineering sciences at the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA. Her research interests are signal processing for communications, wireless networks, information theory, and convex optimization. Her Ph.D. focus has been on precoding techniques for exploiting partial channel knowledge at the transmitter in MIMO wireless systems. She has been a recipient of several awards including the Australian Institute of Engineers Award at RMIT, the Rambus Corporation Stanford Graduate Fellowship and the Intel Foundation Ph.D. Fellowship at Stanford University.

Arogyaswami Paulraj (apaulraj@stanford.edu) received the Ph.D. degree from the Indian Institute of Technology, New Delhi in 1973. He is currently a professor with the Department of Electrical Engineering, Stanford University, where he supervises the Smart Antennas Research Group, working on applications of ST techniques for wireless communications. His nonacademic positions have included Head, Sonar Division, Naval Oceanographic Laboratory, Cochin, India; Director, Center for Artificial Intelligence and Robotics, India; Director, Center for Development of Advanced Computing, India; Chief Scientist, Bharat Electronics,India; CTO and Founder, Iospan Wireless Inc., Cofounder and CTO of Beceem Communications Inc. His research has emphasized estimation theory, sensor signal processing, parallel computer architectures/algorithms, and ST wireless communications. His engineering experience has included development of sonar systems, massively parallel computers, and broadband wireless systems. He has won several awards for his research and engineering contributions, including the IEEE Signal Processing Society's Technical Achievement Award. He is the author of more than 300 research papers and holds twenty patents. He is a Fellow of the IEEE and a member of both the National Academy of Engineering (NAE) and the Indian National Academy of Engineering.

REFERENCES

[1] E. Telatar, "Capacity of multi-antenna Gaussian channels," Bell Lab. Tech. Memo., Oct. 1995. [Online]. Available: http://mars.bell-labs.com/papers/proof/ (*Eur. Trans. Telecommun. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.)

[2] G.J. Foschini and M.J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[3] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, UK: Cambridge Univ. Press, 2003.

[4] C. Shannon, "Channels with side information at the transmitter," *IBM J. Res. Develop.*, vol. 2, no. 4, pp. 289–293, Oct. 1958.

[5] G. Caire and S.S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007–2019, Sept. 1999.

[6] M. Skoglund and G. Jöngren, "On the capacity of a multiple-antenna communication link with channel side information," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 395–405, Apr. 2003.

[7] A. Narula, M. Lopez, M. Trott, and G. Wornell, "Efficient use of side information in multiple antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1423–1436, Oct. 1998.

[8] M. Vu and A. Paulraj, "On the Capacity of MIMO wireless channels with dynamic CSIT," *IEEE J. Select. Areas. Commun.*, (Special issue on optimization of MIMO transceivers for realistic communication networks), to appear Sept. 2007.

[9] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, pp. 1986–1992, Nov. 1997.

[10] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect feedback," *IEEE Trans. Inform. Theory*, vol. 47, no. 6, pp. 2632–2639, Sept. 2001.

[11] S. Venkatesan, S. Simon, and R. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," *Proc. IEEE Veh. Tech. Conf.*, vol. 3, pp. 1767–1771, Oct. 2003.

[12] D. Hösli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," in *Proc. 5th Int. ITG Conf. Source and Channel Coding*, Jan. 2004.

[13] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, June 2003.

[14] S. Jafar and A. Goldsmith, "Transmitter optimization and optimality of beamforming for multiple antenna systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1165–1175, July 2004.

[15] E. Jorswieck and H. Boche, "Channel capacity and capacity-range of beamforming in MIMO wireless systems under correlated fading with covariance feedback," *IEEE Trans. Wireless. Commun.*, vol. 3, no. 5, pp 1543–1553, Sept. 2004.

[16] E. Jorswieck and H. Boche, "Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 12, pp. 3440–3453, Dec. 2004.

[17] A. Tulino, A. Lozano, and S. Verdu, "Capacity-achieving input covariance for single user multi-antenna channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, pp. 662–671, Mar. 2006.

[18] G. Jöngren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 611–627, Mar. 2002.

[19] S. Zhou and G. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2599–2613, Oct. 2002.

[20] E. Jorswieck, A. Sezgin, H. Boche, and E. Costa, "Optimal transmit strategies in MIMO Ricean channels with MMSE receiver," in *Proc. Veh. Tech. Conf.*, Sept. 2004, vol. 5, pp. 3787–3791.

[21] L. Liu and H. Jafarkhani, "Application of quasi-orthogonal space-time block codes in beamforming," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 54–63, Jan. 2005.

[22] H. Sampath and A. Paulraj, "Linear precoding for space-time coded systems with known fading," *IEEE Commun. Lett.*, vol. 6, no. 6, pp. 239–241, June 2002.

[23] S. Zhou and G. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel correlations," *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1673–1690, July. 2003.

[24] T. Haustein and H. Boche, "Optimal power allocation for MSE and bit-loading in MIMO systems and the impact of correlation," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Apr. 2003, vol. 4, pp. 405–408.

[25] M. Vu and A. Paulraj, "Optimal linear precoders for MIMO wireless correlated channels with non-zero mean in space-time coded systems," *IEEE Trans. Signal Processing*, vol. 54, no. 6, pp. 2318–2332, June 2006.

[26] 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, no. 12, pp. 2198–2206, Dec. 2001.

[27] A. Scaglione, P. Stoica, S. Barbarossa, G. Giannakis, and H. Sampath, "Optimal designs for space-time linear precoders and decoders," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1051–1064, May 2002.

[28] D. Palomar, J. Cioffi, and M. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.

[29] Y.W. Ding, T.N. Davidson, Z.Q. Luo, and K.M. Wong, "Minimum BER block precoders for zero-forcing equalization," *IEEE Trans. Signal Processing*, vol. 51, no. 9, pp. 2410–2423, Sept. 2003.

[30] S. Zhou and G.B. Giannakis, "MIMO communications with partial channel state information," in *Space-Time Processing for MIMO Communications*, A.B. Gershman and N. Sidiropoulos, Eds. New York: Wiley, 2005.

[31] T. Kim, G. Jöngren, and M. Skoglund, "Weighted space-time bit-interleaved coded modulation," in *Proc. IEEE Inform. Theory Workshop*, Oct. 2004, pp. 375–380.

[32] A. Bourdoux, B. Come, and N. Khaled, "Non-reciprocal transceivers in OFDM/SDMA systems: Impact and mitigation," in *Proc. Radio and Wireless Conf.*, Aug. 2003, pp. 183–186.

[33] A. Hottinen, O. Tirkkonen, and R. Wichman, *Multi-Antenna Transceiver Techniques for 3G and Beyond*. New York: Wiley, 2003.

[34] T. Rappaport, *Wireless Communications: Principles and Practice*. Englewood Cliffs, NJ: Prentice Hall PTR, 1996.

[35] W. Jakes, *Microwave Mobile Communications*. Piscataway, NJ: IEEE Press, 1994.

[36] T. Kailath, A. Sayed, and H. Hassibi, *Linear Estimation*. Englewood Cliffs, NJ: Prentice Hall, 2000.

[37] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[38] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[39] V. Tarokh, N. Seshadri, and R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[40] H. Jafarkhani, "A quasi-orthogonal space time block code," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 1–4, Jan. 2001.

[41] O. Tirkkonen, A. Boariu, and A. Hottinen, "Minimal non-orthogonality rate 1 space-time block code for 3+ tx antennas," in *Proc. IEEE ISSSTA2000*, Sept. 2000, vol.2, pp. 429–432.

[42] G. Turin, "On optimal diversity reception, II," *IRE Trans. Commun. Syst.*, vol. 10, no. 1, pp. 22–31, Mar. 1962.

[43] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[44] H. Yao and G. Wornell, "Structured space-time block codes with optimal diversity-multiplexing tradeoff and minimum delay," in *Proc. IEEE Global Telecom. Conf.*, Dec. 2003, vol. 4, pp. 1941–1945.

[45] H. Gamal, G. Caire, and M. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 6, pp. 968–985, June 2004.

[46] R. Narasimhan, "Finite-SNR diversity-multiplexing tradeoff for correlated Rayleigh and Rician MIMO channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3965–3979, Sept. 2006.

[47] D. Shiu, G. Foschini, M. Gans, and J. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[48] K. Yu, M. Bengtsson, B. Ottersten, D. McNamara, P. Karlsson, and M. Beach, "Second order statistics of NLOS indoor MIMO channels based on 5.2 GHz measurements," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2001, vol. 1, pp. 25–29.

[49] J. Kermoal, L. Schumacher, K. Pedersen, P. Mogensen, and F. Frederiksen, "A stochastic MIMO radio channel model with experimental validation," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1211–1226, Aug. 2002.

[50] D. Bliss, A. Chan, and N. Chang, "MIMO wireless communication channel phenomenology," *IEEE Trans. Antennas Propag.*, vol. 52, no. 8, pp. 2073–2082, Aug. 2004.

[51] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[52] W. Weichselberger, M. Herdin, H. Özcelik, and E. Bonek, "A stochastic MIMO channel model with joint correlation of both link ends," *IEEE Trans. Wireless Commun.*, January 2006, vol. 5, no. 1, pp. 90 –100.

[53] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2003. [Online]. Available http://www.stanford.edu/_boyd/cvxbook.html

[54] A. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[55] M. Vu, Exploiting Transmit Channel Side Information in MIMO Wireless Systems, Ph.D. dissertation, Stanford Univ., Palo Alto, CA, July 2006.

[56] M. Vu and A. Paulraj, "Precoding design," in *MIMO Wireless Communications*, E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. J. Paulraj, and H.V. Poor, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2007.

[57] J.R. Birge and F. Louveaux, *Introduction to Stochastic Programming*. New York: Springer Verlag, 1997.

[58] E. Jorswieck and H. Boche, "Performance analysis of MIMO systems in spatially correlated fading using matrix-monotone functions," *IEICE Trans. Fundamentals*, May 2006, vol. E89-A, no. 5, pp. 1454–1472.

[59] M. Vu and A. Paulraj, "Optimum space-time transmission for a high K factor wireless channel with partial channel knowledge," *Wiley J. Wireless Commun. Mobile Comput.*, vol. 4, pp. 807–816, Nov. 2004.

[60] R.W. Heath, and A. Paulraj, "Switching between diversity and multiplexing in MIMO systems," *IEEE Trans. Commun.*, vol. 53, pp. 962–968, June 2005.

[61] D. Love and R. Heath, Jr., "Limited feedback unitary precoding for orthogonal space-time block codes," *IEEE Trans. Signal Processing*, vol. 53, no. 1, pp. 64–73, Jan. 2005.

[62] D. Love and R. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing," *IEEE Trans. Inform. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.

[63] K.K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.

[64] J.C. Roh and B.D. Rao, "Design and analysis of MIMO spatial multiplexing systems with quantized feedback," *IEEE Trans. Signal Processing*, vol. 54, no. 8, pp. 2874–2886, Aug. 2006.

[65] D. Love, R. Heath, W. Santipach, and M. Honig, "What is the value of limited feedback for MIMO channels?," *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, Oct. 2004.

[66] R. Fischer, C. Stierstorfer, and J. Huber, "Precoding for point-to-multipoint transmission over MIMO ISI channels," in *Proc. Int. Zurich Sem. Commun.*, Feb. 2004, pp. 208–211.

[67] R.F. Fischer, *Precoding and Signal Shaping for Digital Transmission*. New York: Wiley, 2002.

[68] J. Choi, and R.W. Heath Jr., "Interpolation based unitary precoding for spatial multiplexing MIMO-OFDM with limited feedback," in *Proc. IEEE GLOBECOM '04*, Dec. 2004, vol. 1, pp. 214–218.

[69] P. Xia., S. Zhou, and G.B. Giannakis, "Adaptive MIMO-OFDM based on partial channel state information," *IEEE Trans. Signal Processing*, vol. 52, no. 1, pp. 202–213, Jan. 2004.

[70] E. Yoon, D. Tujkovic, A. Paulraj, "Subcarrier and power allocation for an OFDMA uplink based on tap correlation information," in *Proc. IEEE Int. Conf. Commun.*, May 2005, vol. 4, pp. 2744–2748.

[71] S. Jafar and A. Goldsmith, "Isotropic fading vector broadcast channels: The scalar upperbound and loss in degrees of freedom," *IEEE Trans. Inform. Theory*, vol. 51, no. 3, pp. 848–857, Mar. 2005.

[72] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information," in *Proc. 43rd Ann. Allerton Conf. Commun., Control, and Computing*, Sept. 2005.

[73] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[74] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inform. Theory*, Nov. 2006, vol. 52, no. 11, pp. 5045–5060. SP