Capacity Limits of MIMO Systems

Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal, and Sriram Vishwanath

I. INTRODUCTION

In this chapter we consider the Shannon capacity limits of single-user and multi-user MIMO systems. These fundamental limits dictate the maximum data rates that can be transmitted simultaneously to all users over the MIMO channel with asymptotically small error probability, assuming no constraints on delay or complexity of the encoder and decoder. Much of the initial excitement about MIMO systems was due to pioneering work by Winters [132], Foschini [29], and Telatar [113] predicting remarkable capacity growth for wireless systems with multiple antennas when the channel exhibits rich scattering and its variations can be accurately tracked. This initial promise of exceptional spectral efficiency almost “for free” resulted in an explosion of research and commercial activity to characterize the theoretical and practical issues associated with MIMO systems. However, these predictions are based on somewhat unrealistic assumptions about the underlying time-varying channel model and how well it can be tracked at the receiver as well as at the transmitter. More realistic assumptions can dramatically impact the potential capacity gains of MIMO techniques. This chapter provides a comprehensive summary of MIMO Shannon capacity for both single-user and multi-user systems with and without fading under different assumptions about what is known at the transmitter(s) and receiver(s).

We first provide some background on Shannon capacity and mutual information, and then apply these ideas to the single-user AWGN MIMO channel. We next consider MIMO fading channels, describe the different capacity definitions that arise when the channel is time-varying, and present the MIMO capacity under these different definitions. These results indicate that the capacity gain obtained from multiple antennas heavily depends on the available channel information at either the receiver or transmitter, the channel SNR, and the correlation between the channel gains on each antenna element. We then focus attention on the capacity region of multiuser channels: in particular the multiple access and broadcast channels. In contrast to single-user MIMO channels, capacity results for these multi-user MIMO channels can be quite difficult to obtain, especially when the channel is not known perfectly at all transmitters and receivers. We will show that with perfect channel knowledge, the capacity region of the MIMO multiple access and broadcast channels are intimately related via a duality transformation, which can greatly simplify capacity analysis. This transformation facilitates finding the transmission strategies that achieve a point on the boundary of the MIMO multiple access channel (MAC) capacity region in terms of the transmission strategies of the MIMO broadcast channel (BC) capacity region, and vice-versa. We then consider MIMO cellular systems with frequency reuse, where the base stations cooperate. With cooperation the base stations act as a spatially distributed antenna array, and transmission strategies that exploit this structure exhibit significant capacity gains. The chapter concludes with a discussion of capacity results for wireless ad hoc networks where nodes either have multiple antennas or cooperate to form multiple antenna transmitters and/or receivers. Open problems in this field abound and are discussed throughout the chapter.

A table of abbreviations used throughout the chapter is given in Table I.

CSI	Channel State Information
CDI	Channel Distribution Information
CSIT	Transmitter Channel State Information
CSIR	Receiver Channel State Information
CDIT	Transmitter Channel Distribution Information
CDIR	Receiver Channel Distribution Information
ZMSW	Zero Mean Spatially White
CMI	Channel Mean Information
CCI	Channel Covariance Information
QCI	Quantized Channel Information
DPC	Dirty Paper Coding
MAC	Multiple-Access Channel
BC	Broadcast Channel

TABLE I

TABLE OF ABBREVIATIONS.

II. Mutual Information and Shannon Capacity

Channel capacity was pioneered by Claude Shannon in the late 1940s, using a mathematical theory of communication based on the notion of mutual information between the input and output of a channel [104–106]. Shannon defined capacity as the mutual information maximized over all possible input distributions. The significance of this mathematical construct was Shannon's coding theorem and converse, which prove that a code exists that can achieve a data rate asymptotically close to capacity with negligible probability of error, and that any data rate higher than capacity cannot be achieved without an error probability bounded away from zero.

For a memoryless time-invariant single-user channel with random input X and random output Y, the channel's **mutual information** is defined as

$$I(X;Y) = \int_{S_x, S_y} \log\left(\frac{f(y|x)}{f(y)}\right) dF(x,y), \tag{1}$$

where the integral is taken over the supports S_x, S_y of the random variables X and Y, respectively, $f(y)$ denotes the probability density function (p.d.f.) of Y, and $F(x)$ and $F(x,y)$ denote the cumulative distribution functions (c.d.f.s) of X and X, Y, respectively. The log function is typically with respect to base 2, in which case the units of mutual information are bits per channel use. Assuing all the random variables posssess a p.d.f., mutual information can also be written in terms of the **differential entropy** of the channel output and conditional output as $I(X;Y) = h(Y) - h(Y|X)$, where $h(Y) = -\int_{S_y} f(y)\log f(y)dy$ and $h(Y|X) = -\int_{S_x,S_y} f(x,y)\log f(y|x)dxdy$. Shannon proved that channel capacity for most channels equals the mutual information of the channel maximized over all possible input distributions:

$$C = \max_{f(x)} I(X;Y) = \max_{f(x)} \int_{S_x,S_y} f(x,y)\log\left(\frac{f(x,y)}{f(x)f(y)}\right). \tag{2}$$

For a time-invariant AWGN channel with received SNR γ, the maximizing input distribution is Gaussian, which results in the channel capacity

$$C = B\log_2(1+\gamma). \tag{3}$$

The definition of entropy and mutual information is the same when the channel input and output are vectors instead of scalars, as in the MIMO channel. Thus, the Shannon capacity of the MIMO AWGN channel is based on its maximum mutual information, as described in the next section.

When the channel is time-varying channel capacity has multiple definitions, depending on what is known about the channel state or its distribution at the transmitter and/or receiver. These definitions have different operational meanings. Specifically, when the instantaneous channel gains, also called the *channel state information* (CSI), are known perfectly at both transmitter and receiver, the transmitter can adapt its transmission strategy (rate and/or power) relative to the instantaneous channel state. In this case the Shannon (ergodic) capacity is the maximum mutual information averaged over all channel states. Ergodic capacity is an appropriate capacity metric for channels that vary quickly, where the channel is ergodic over the duration of one codeword. In this case rates approaching ergodic capacity can be achieved with each codeword transmission. Ergodic capacity is typically achieved using an adaptive transmission policy where the power and data rate vary relative to the channel state variations [34]. Alternatively, this ergodic capacity can be achieved when only transmit power is varied [10].

An alternate capacity definition for time-varying channels with perfect transmitter and receiver CSI is outage capacity. Outage capacity requires a fixed data rate in all non-outage channel states, which is needed for applications with delay-constrained data where the data rate cannot depend on channel variations (except in outage states, where no data is transmitted). The average rate associated with outage capacity is typically smaller than ergodic capacity due to the additional constraint associated with this definition. Outage capacity is the appropriate capacity metric in slowly-varying channels, where the channel coherence time exceeds the duration of a codeword. In this case each codeword experiences only one channel state: if the channel state is not good enough to support the desired rate then an outage is declared. Note that ergodic capacity is not a relevant metric for slowly varying channels since each codeword is affected by a single channel realization. Similarly, outage capacity is not an appropriate capacity metric for channels that vary quickly: since the channel experiences all possible channel states over the duration of a codeword, there is no notion of poor states where an outage must be declared. Outage capacity under perfect CSIT and CSIR has been studied for single antenna channels [12] [39] [77] but this work has yet to be extended to MIMO channels.

When only the channel distribution is known at the transmitter (receiver) the transmission (reception) strategy is based on the channel distribution instead of the instantaneous channel state. The channel coefficients are typically assumed to be jointly Gaussian, so the channel distribution is specified by the channel mean and covariance matrices. We will refer to knowledge of the channel distribution as *channel distribution information* (CDI): CDI at the transmitter is abbreviated as CDIT, and CDI at the receiver is denoted by CDIR. We assume throughout the chapter that CDI is always perfect, so there is no mismatch between the CDI at the transmitter or receiver and the true channel distribution. When only the receiver has perfect CSI the transmitter must maintain a fixed-rate transmission strategy optimized with respect to its CDI. In this case ergodic capacity defines the rate that can be achieved based on averaging over all channel states [113], and this metric is relevant for channels that vary quickly so that codeword transmissions are affected by all possible channel states. Alternatively, the transmitter can send at a rate that cannot be supported by all channel states: in these poor channel states the receiver declares an outage and the transmitted data is lost. In this scenario each transmission rate has an outage probability associated with it and capacity is measured relative to outage probability[1] (capacity CDF) [29]. An excellent tutorial on fading channel capacity for single antenna channels can be found in [4]. This chapter extends these results to MIMO systems.

[1]Note that an outage under perfect CSI at the receiver only is different than an outage when both transmitter and receiver have perfect CSI. Under receiver CSI, an outage only occurs when the transmitted data cannot be reliably decoded at the receiver, so that data is lost. When both the transmitter and receiver have perfect CSI the channel is not used during outage (no service), so no data is lost.

In multi-user channels capacity becomes a K-dimensional region defining the set of all rate vectors $(R_1, \ldots, R_K)$ simultaneously achievable by all K users. The multiple capacity definitions for time-varying channels under different transmitter and receiver CSI and CDI assumptions extend to the capacity region of the MAC and BC in the obvious way, as we will describe in Section IV. However, these MIMO multi-user capacity regions, even for time-invariant channels, are difficult to find. Few capacity results exist for time-varying multi-user MIMO channels under the realistic assumption that the transmitter(s) and/or receiver(s) have CDI only. The results become even more sparse for more complex systems such as cellular and ad-hoc wireless networks, as will be seen in Sections V and VI. Thus, there are many open problems in these areas, as will be highlighted in the related sections.

III. Single User MIMO

In this section we focus on the capacity of single-user MIMO channels. While most wireless systems today support multiple users, single-user results are still of much interest for the insight they provide and their application to channelized systems where users are allocated orthogonal resources (time, frequency bands, etc.). MIMO channel capacity is also much easier to derive for single users than for multiple users. Indeed, single-user MIMO capacity results are known for many cases where the corresponding multi-user problems remain unsolved. In particular, very little is known about multi-user capacity without the assumption of perfect channel state information at the transmitter (CSIT) and at the receiver (CSIR). While there remain many open problems in obtaining the single-user capacity under general assumptions of CSI and CDI, for several interesting cases the solution is known. This section will discuss fundamental capacity limits for single-user MIMO channels with particular focus on special cases of CDI at the transmitter as well as the receiver. We begin with a description of the channel model and the different CSI and CDI models we consider, along with their motivation.

A. Channel and Side Information Model

Consider a transmitter with M transmit antennas, and a receiver with N receive antennas. The channel can be represented by the $N \times M$ matrix $\mathbf{H}$ of channel gains h_{ij} representing the gain from transmit antenna j to receive antenna i. The $N \times 1$ received signal $\mathbf{y}$ is equal to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{4}$$

where $\mathbf{x}$ is the $M \times 1$ transmitted vector and $\mathbf{n}$ is the $N \times 1$ additive white circularly symmetric complex Gaussian noise vector[2], normalized so that its covariance matrix is the identity matrix. The normalization of any non-singular noise covariance matrix $\mathbf{K_w}$ to fit the above model is as straightforward as multiplying the received vector $\mathbf{y}$ with $\mathbf{K_w}^{-1/2}$ to yield the effective channel $\mathbf{K_w}^{-1/2}\mathbf{H}$ and a white noise vector. The channel state information is the channel matrix $\mathbf{H}$ and/or its distribution.

The transmitter is assumed to be subject to an average power constraint of P across all transmit antennas, i.e. $E[\mathbf{x^H x}] \leq \mathbf{P}$. Since the noise power is normalized to unity, we commonly refer to the power constraint P as the SNR.

A.1 Perfect CSIR and CSIT

With perfect CSIT or CSIR, the channel matrix $\mathbf{H}$ is assumed to be known perfectly and instantaneously at the transmitter or receiver, respectively. When the transmitter or receiver knows the channel state perfectly, we also assume that it knows the distribution of this state perfectly, since the distribution can be obtained from the state observations.

A.2 Perfect CSIR and CDIT

The perfect CSIR and CDIT model is motivated by the scenario where the channel state can be accurately tracked at the receiver and the statistical channel model at the transmitter is based on channel distribution information fed back from the receiver. This distribution model is typically based on receiver estimates of the channel state and the uncertainty and delay associated with these estimates. Figure 1 illustrates the underlying communication model in this scenario, where $\tilde{\mathcal{N}}$ denotes the complex Gaussian distribution.

The salient features of the model are as follows. The channel distribution is defined by the parameter θ and, conditioned on this parameter, the channel realizations $\mathbf{H}$ at different time instants are independently and identically distributed (i.i.d.). Since the channel statistics will change over time due to mobility of the transmitter, receiver, and the scattering environment, we assume that θ is time-varying. Note that the statistical model depends on the time-scale of interest. For example, in the short term the channel coefficients may have a non-zero mean and one set of correlations reflecting the geometry of the particular propagation environment. However, over a long term the channel coefficients may be described as zero-mean and uncorrelated due to the averaging over several propagation environments. For this reason,

[2]A complex Gaussian vector $\mathbf{x}$ is circularly symmetric if for any $\theta \in [0, 2\pi]$, the distribution of $\mathbf{x}$ is the same as the distribution of $e^{j\theta}\mathbf{x}$

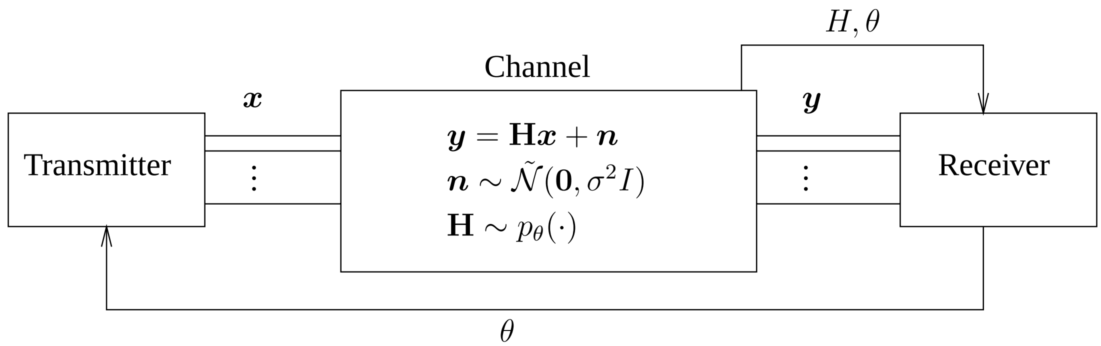

Fig. 1. MIMO channel with perfect CSIR and distribution feedback.

uncorrelated, zero-mean channel coefficients are commonly assumed for the channel distribution in the absence of distribution feedback or when it is not possible to adapt to the short-term channel statistics. However, if the transmitter receives frequent updates of θ and it can adapt to these time-varying short-term channel statistics then capacity is increased relative to the transmission strategy associated with just the long-term channel statistics. In other words, adapting the transmission strategy to the short-term channel statistics increases capacity. In the literature adaptation to the short-term channel statistics (the feedback model of Figure 1) is referred to by many names including mean and covariance feedback, quantized feedback, imperfect feedback, and partial CSI [124] [56] [54] [50] [67] [65] [86] [109]. The feedback channel is assumed to be free from noise. This makes the CDIT a deterministic function of the CDIR and allows optimal codes to be constructed directly over the input alphabet [10]. We assume a power constraint such that for each realization of θ, the conditional average transmit power is constrained as $\mathbb{E}\left[||\boldsymbol{u}||^2|\Theta = \theta\right] \leq P$.

The ergodic capacity C of the system in Figure 1 is the capacity $C(\theta)$ averaged over the different θ realizations:

$$C = \mathbb{E}_\theta[C(\theta)],$$

where $C(\theta)$ is the ergodic capacity of the channel shown in Figure 2. This figure represents a MIMO channel with perfect CSI at the receiver and only CDI about the *constant* distribution θ at the transmitter. Channel capacity calculations generally implicitly assume CDI at both the transmitter and receiver except for special channel classes, such as the compound channel or arbitrarily varying channel [5] [24] [25]. This implicit knowledge of θ is justified by the fact that the channel coefficients are typically modeled based on their long-term average distribution. Alternatively, θ can be obtained by the feedback model of Figure 1. The distribution feedback model of Figure 1 along with the system model of Figure 2 lead

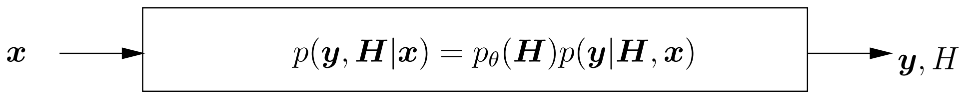

Fig. 2. MIMO channel with perfect CSIR and CDIT (θ fixed).

to various capacity results under different distribution (θ) models. The availability of CDI at either the transmitter or receiver is explicitly indicated to contrast with the case where CSI is also available.

Computation of $C(\theta)$ for general $p_\theta(\cdot)$ is a hard problem. With the exception of the quantized channel information model, almost all research in this area has focused on three special cases for this distribution: zero-mean spatially white channels, spatially white channels with nonzero mean, and zero-mean channels with non-white channel covariance. In all three cases the channel coefficients are modeled as complex jointly Gaussian random variables. Under the zero-mean spatially white (ZMSW) model, the channel mean is zero and the channel covariance is modeled as white, i.e. the channel elements are assumed to be i.i.d. random variables. This model typically captures the long-term average distribution of the channel coefficients averaged over multiple propagation environments. Under the channel mean information (CMI) model, the mean of the channel distribution is nonzero while the covariance is modeled as white with a constant scale factor. This model is motivated by a system where the delay in the feedback leads to an imperfect estimate at the transmitter, so the CMI reflects the outdated channel measurement and the constant factor reflects the estimation error. Under the channel covariance information (CCI) model, the channel is assumed to be varying too rapidly to track its mean, so the mean is set to zero and the information regarding the relative geometry of the propagation paths is captured by a non-white covariance matrix. Based on the underlying system model shown in Figure 1, in the literature the CMI model is also called mean feedback and the CCI model is also called covariance feedback. Mathematically, the three distribution models for $\mathbf{H}$ can be described as follows:

ZMSW : $\mathbb{E}[\mathbf{H}] = \mathbf{0}, \quad \mathbf{H} = \mathbf{H^w}$

CMI : $\mathbb{E}[\mathbf{H}] = \overline{\mathbf{H}}, \quad \mathbf{H} = \overline{\mathbf{H}} + \sqrt{\alpha}\mathbf{H^w}$

CCI : $\mathbb{E}[\mathbf{H}] = \mathbf{0}, \quad \mathbf{H} = (\mathbf{R^r})^{1/2}\mathbf{H^w}(\mathbf{R^t})^{1/2}.$

Here $\mathbf{H^w}$ is an $N \times M$ matrix of i.i.d. zero mean, unit variance complex circularly symmetric Gaussian random variables. In the CMI model, the channel mean $\overline{\mathbf{H}}$ and α are constants that may be interpreted as the channel estimate based on the feedback, and the variance of the estimation error, respectively. In the CCI model, $\mathbf{R^r}$ and $\mathbf{R^t}$ are referred to as the receive and transmit antenna correlation matrices, respectively. Although not completely general, this simple correlation model has been validated through field measurements as a sufficiently accurate representation of the fade correlations seen in actual cellular systems [17]. Under CMI the channel mean $\overline{\mathbf{H}}$ and the variance of the estimation error α are assumed known when there is CDI, and under CCI the transmit and receive covariance matrices $\mathbf{R}^{\mathbf{r}}$ and $\mathbf{R}^{\mathbf{t}}$ are assumed known when there is CDI.

In addition to the CMI, CCI, and ZMSW models of CDIT, recent research has explored the effects of quantized channel state information (QCI) at the transmitter based on a finite bit rate feedback channel. In this model, the receiver is assumed to have perfect CSI and feeds back a B-bit quantization of the channel instantiation to the transmitter. This model is most applicable to relatively slow fading scenarios, where the receiver feeds back quantized CSI at the beginning of each block, thereby allowing the transmitter to adapt. Note that this is a very practical model, as many wireless systems have a low-rate feedback link from the receiver to the transmitter. With B bits of QCI, a predetermined set of $N = 2^B$ quantization vectors is used to represent the channel at the transmitter. Finding the best quantization vectors is equivalent to the Grassmannian packing of subspaces within a vector space to maximize the minimum distance between them [87] [79]. Conditioned on the QCI, the channel distribution assumed at the transmitter is constrained within the Voronoi region of the quantization vector used to represent the channel.

A.3 CDIT and CDIR

In highly mobile channels the assumption of perfect CSI at the receiver can be unrealistic. This motivates system models where both transmitter and receiver only have information about the channel distribution. Even for a rapidly fluctuating channel where reliable channel estimation is not possible, it might be possible for the receiver to track the short-term distribution of the channel fades, as the channel distribution changes much more slowly than the channel itself. The estimated distribution can be made available to the transmitter through a feedback channel. Figure 3 illustrates the underlying communication model.

Note that the estimation of the channel statistics at the receiver is captured in the model as a genie that provides the receiver with the correct channel distribution. The feedback channel represents the same information being made available to the transmitter simultaneously. This model is slightly optimistic because in practice the receiver estimates θ only from the received signal $\boldsymbol{v}$ and therefore will not have a perfect estimate.

As in the previous subsection the ergodic capacity turns out to be the expected value (expectation over θ) of the ergodic capacity $C(\theta)$, where $C(\theta)$ is the ergodic capacity of the channel in Figure 4. In this figure θ is constant and known at both the transmitter and receiver (CDIT and CDIR). As in the previous

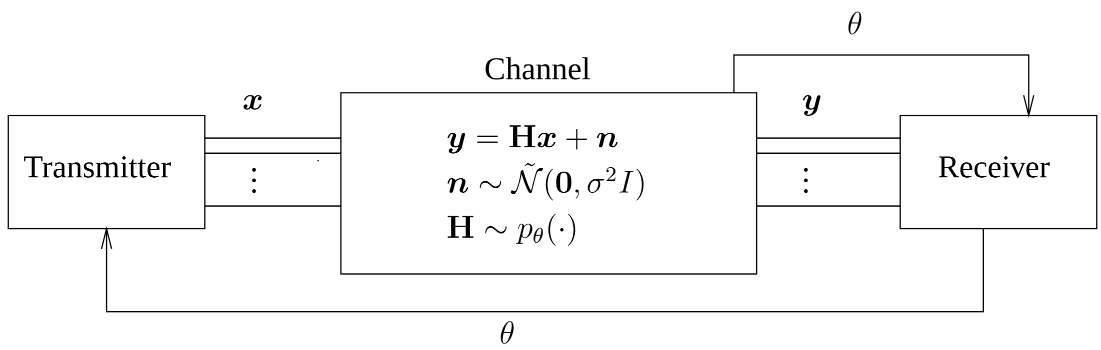

Fig. 3. MIMO channel with CDIR and distribution feedback.

section, the computation of $C(\theta)$ is difficult for general θ and the capacity investigations are limited mainly to the same channel distribution models described in the previous subsection: the ZMSW, CMI, CCI and QCI models.

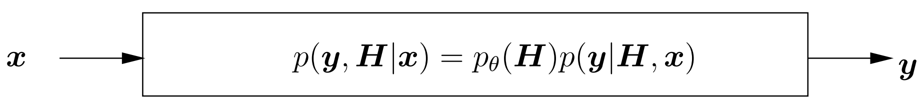

Fig. 4. MIMO channel with CDIT and CDIR (θ fixed).

Next, we summarize single user MIMO capacity results under various assumptions on CSI and CDI.

B. Constant MIMO Channel Capacity

When the channel is constant and known perfectly at the transmitter and the receiver, the capacity (maximum mutual information) is

$$C = \max_{\mathbf{Q}\,:\,\text{Tr}(\mathbf{Q})=P} \log \left|\mathbf{I}_N + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathbf{H}}\right| \tag{5}$$

where the optimization is over the input covariance matrix $\mathbf{Q}$, which is $M \times M$ and must be positive semi-definite by definition. Using the singular value decomposition (SVD) of the $M \times N$ matrix $\mathbf{H}$, this channel can be converted into $\min(M, N)$ parallel, non-interfering single-input/single-output channels [113] [35]. The SVD allows us to rewrite $\mathbf{H}$ as $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, where $\mathbf{U}$ is $M \times M$ and unitary, $\mathbf{V}$ is $N \times N$ and unitary, and $\mathbf{\Sigma}$ is $M \times N$ and diagonal with non-negative entries. The diagonal elements of the matrix $\mathbf{\Sigma}$, denoted by σ_i, are the singular values of $\mathbf{H}$ and are assumed to be in descending order (i.e. $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_{\min(M,N)}$). The matrix $\mathbf{H}$ has exactly R_H positive singular values, where R_H is the rank of $\mathbf{H}$, which by basic principles satisfies $R_H \leq \min(M, N)$.

The MIMO channel is converted into parallel, non-interfering channels by pre-multiplying the input by

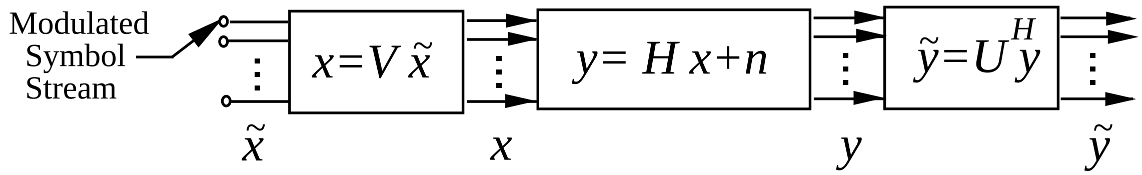

Fig. 5. MIMO Channel Decomposition

the matrix $\mathbf{V}$ (i.e. transmit precoding) and post-multiplying the output by the matrix $\mathbf{U}^H$. This conversion is illustrated in Fig. 5. The input $\mathbf{x}$ to the channel is generated by multiplying the data stream $\tilde{\mathbf{x}}$ by the matrix $\mathbf{V}$, i.e. $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$. Since $\mathbf{V}$ is a unitary matrix, this is a power preserving linear transformation, i.e. $E[||\mathbf{x}||^2] = E[||\tilde{\mathbf{x}}||^2]$. The vector $\mathbf{x}$ is fed into the channel and the output $\mathbf{y}$ is multiplied by the matrix $\mathbf{U}^H$, resulting in $\tilde{\mathbf{y}} = \mathbf{U}^H\mathbf{y}$, which can be expanded as:

$$\begin{aligned}
\tilde{\mathbf{y}} &= \mathbf{U}^H(\mathbf{H}\mathbf{x} + \mathbf{n}) \\
&= \mathbf{U}^H(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H(\mathbf{V}\tilde{\mathbf{x}}) + \mathbf{n}) \\
&= \boldsymbol{\Sigma}\tilde{\mathbf{x}} + \mathbf{U}^H\mathbf{n} \\
&= \boldsymbol{\Sigma}\tilde{\mathbf{x}} + \tilde{\mathbf{n}},
\end{aligned}$$

where $\tilde{\mathbf{n}} = \mathbf{U}^H\mathbf{n}$. Since $\mathbf{U}$ is unitary and $\mathbf{n}$ is a spatially white complex Gaussian, $\tilde{\mathbf{n}}$ and $\mathbf{n}$ have the same distribution. Using the fact that $\boldsymbol{\Sigma}$ is diagonal, we have $\tilde{\mathbf{y}}_i = \sigma_i\tilde{\mathbf{x}}_i + \tilde{\mathbf{n}}_i$ for $i = 1, \ldots, \min(M, N)$. Since R_H of the singular values σ_i are strictly positive, the result is R_H parallel, non-interfering channels. The parallel channels are commonly referred to as the *eigenmodes* of the channel, because the singular values of $\mathbf{H}$ are equal to the square root of the eigenvalues of the matrix $\mathbf{H}\mathbf{H}^H$.

Because the parallel channels are of different quality, the water-filling algorithm can be used to optimally allocate power over the parallel channels, leading to the following allocation:

$$P_i = \left(\mu - \frac{1}{\sigma_i^2}\right)^+, \quad 1 \leq i \leq R_H \tag{6}$$

where P_i is the power of $\tilde{\mathbf{x}}_i$, x^+ is defined as $\max(x, 0)$, and the waterfill level μ is chosen such that $\sum_{i=1}^{R_H} P_i = P$. Capacity is therefore achieved by choosing each component $\tilde{\mathbf{x}}_i$ according to an independent Gaussian distribution with power P_i. The covariance which achieves the maximum in (5) (i.e. the covariance of the capacity-achieving input) is $\mathbf{Q} = \mathbf{V}\mathbf{P}\mathbf{V}^H$, where the $M \times M$ matrix $\mathbf{P}$ is defined as $\mathbf{P} = \text{diag}(P_1, \ldots, P_{R_H}, 0, \ldots, 0)$. The resulting capacity is given by

$$C = \sum_{i}^{R_H} (\log(\mu\sigma_i^2))^+. \tag{7}$$

At low SNR's, the waterfilling algorithm allocates all power to the strongest of the R_H parallel channels (i.e. $P_1 = P$ and $P_i = 0$ for $i \neq 1$). At high SNR, the waterfilling algorithm allocates approximately equal power to each of the R_H, and a first order approximation of the capacity at high SNR is $C \approx R_H \log_2(P) + O(1)$, where the constant term depends on the singular values of $\mathbf{H}$. From this approximation, one can see that every increase of 3 dB in transmit power leads to an increase of approximately R_H bps/Hz in spectral efficiency; this contrasts with single antenna systems, where every 3 dB of power only leads to an additional bit of spectral efficiency. For this reason the pre-log term of R_H is referred to as the *multiplexing gain* in the literature, as it determines the multiplicative capacity increase of MIMO relative to SISO systems.

There are a number of intuitive observations that can made regarding the capacity achieving transmission scheme. First, note that transmit precoding by the matrix $\mathbf{V}$ "aligns" the inputs with the eigenmodes of the channel. By aligning the inputs with the eigenmodes, simple post-multiplication of the output signal results in independent (noisy) observations of each of the inputs, i.e. a complete decoupling of the different data streams. If the transmitter did not perform this alignment process, e.g. transmitted independent inputs on each of the M transmit antennas, then the receiver would not be able to completely decouple the multiple data streams, which leads to an effective loss in SNR of each of the streams and thus is not capacity achieving. In addition, the transmitter performs waterfilling across the different eigenmodes of the channel in order to take advantage of channels of different quality. As expected, the channels with the highest SNR are loaded with the most power and rate.

A key advantage of the decomposition of the MIMO channel into parallel non-interfering channels is that the decoding complexity is only linear in R_H, the rank of the channel. When it is not possible to perform such a decomposition (e.g. when the transmitter does not have perfect knowledge of the matrix $\mathbf{H}$), decoding complexity is typically exponential in R_H.

Although the constant channel model is relatively easy to analyze, wireless channels in practice are not fixed or constant. Instead, due to the changing propagation environment wireless channels vary over time, assuming values over a continuum. The capacity of fading channels is investigated next.

C. Fading MIMO Channel Capacity

With slow fading, the channel may remain approximately constant long enough to allow reliable estimation of the channel state at the receiver (perfect CSIR) and timely feedback of this state information to the transmitter (perfect CSIT). However, in systems with moderate to high user mobility, the system designer is inevitably faced with channels that change rapidly. Fading models where only the channel distribution is available to the receiver (CDIR) and/or transmitter (CDIT) are more applicable to such systems. Capacity results under various assumptions regarding CSI and CDI are summarized in this section.

C.1 Capacity with Perfect CSIT and Perfect CSIR

Perfect CSIT and perfect CSIR model a fading channel that changes slowly enough to be reliably measured by the receiver and fed back to the transmitter without significant delay. The ergodic capacity of a flat-fading channel with perfect CSIT and CSIR is simply the average of the capacities achieved with each channel realization. The capacity for each channel realization is given by the constant channel capacity expression in the previous section. Thus the fading MIMO channel capacity assuming perfect channel knowledge at both transmitter and receiver is,

$$C = \mathbb{E}_{\mathbf{H}} \left[\max_{\mathbf{Q}(\mathbf{H}) \,:\, \mathrm{Tr}(\mathbf{Q}(\mathbf{H}))=P} \log \left| \mathbf{I}_N + \mathbf{H}\mathbf{Q}(\mathbf{H})\mathbf{H}^{\mathbf{H}} \right| \right]. \tag{8}$$

The covariance of the input is written as $\mathbf{Q}(\mathbf{H})$ to emphasize the fact that the covariance can be changed as a function of the channel realization. In fact, the covariance for each channel realization is chosen using the waterfilling procedure described in Chapter III-B. Thus, each MIMO channel realization is decomposed into parallel channels, and waterfilling is performed over both space and time, i.e. over the $\min(M, N)$ eigenmodes in each state and across the fading distribution. Some results on the computation of the waterfilling level and the corresponding capacity are given in [57]. Note that the capacity expression in (8) is valid for *any* fading distribution[3].

Obtaining CSIT can be rather difficult in time-varying channels, as it generally requires either high rate feedback from the receiver, or time-division duplex (TDD) operation on a sufficiently fast scale. However, there are are both capacity and implementation benefits relative to having only CDIT. In the next section we study the capacity with CSIR and CDIT, and briefly compare this scenario to CSIT/CSIR for the ZMSW model.

C.2 Capacity with Perfect CSIR and CDIT: ZMSW Model

For the case of perfect CSIR and a ZMSW channel distribution at the transmitter, the channel matrix $\mathbf{H}$ is assumed to have i.i.d. complex Gaussian entries (i.e. $\mathbf{H} \sim \mathbf{H}^{\mathbf{w}}$). As described in Section II, the two

[3]Additionally, this capacity only depends on the stationary distribution of the fading process and is thus independent of any memory in the fading process. We investigate only memoryless fading processes in this chapter, but this assumption does not affect the capacity if perfect CSI is available at both transmitter and receiver.

relevant capacity definitions in this case are capacity versus outage (capacity CDF) and ergodic capacity. For any given input covariance matrix the input distribution that achieves the ergodic capacity is shown in [30] and [113] to be complex vector Gaussian, mainly because the vector Gaussian distribution maximizes the entropy for any given covariance matrix. This leads to the transmitter optimization problem, i.e., finding the optimum input covariance matrix to maximize ergodic capacity subject to a transmit power (trace of the input covariance matrix) constraint. Mathematically, the problem is to characterize the optimum $\mathbf{Q}$ to maximize

$$C = \max_{\mathbf{Q}\,:\,\mathrm{Tr}(\mathbf{Q})=P} C(\mathbf{Q}), \tag{9}$$

where

$$C(\mathbf{Q}) \triangleq \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_N + \mathbf{H}\mathbf{Q}\mathbf{H}^{\mathbf{H}}\right|\right] \tag{10}$$

is the maximum achievable rate when using input covariance matrix $\mathbb{E}[\mathbf{x}\mathbf{x}^{\mathbf{H}}] = \mathbf{Q}$ and the expectation is with respect to the channel matrix $\mathbf{H}$. The rate $C(\mathbf{Q})$ is achieved by transmitting independent complex circular Gaussian symbols along the eigenvectors of $\mathbf{Q}$. The powers allocated to each eigenvector are given by the eigenvalues of $\mathbf{Q}$.

When there is CSIR and CSIT, as discussed in the previous section, the transmitter can use its instantaneous knowledge of $\mathbf{H}$ to align its transmission (i.e. the input covariance matrix) with the eigenmodes of the channel $\mathbf{H}$. When the transmitter does not know the instantaneous channel realization and only has knowledge of the fading distribution, it is not possible to align the input (whose covariance is fixed for all time) with every possible realization of the channel. In addition to not being able to identify the eigenmodes of the channel, the transmitter is also not able to identify the directions in which the channel is stronger in the sense of delivering more power. Therefore, one might intuitively expect the optimal strategy to involve transmitting power in all spatial directions, without any form of power control. In fact, it is shown in [113] and [30] that the optimum transmit strategy is to indeed transmit power in all spatial directions with equal power allocated to each direction. More specifically, the optimum input covariance matrix that maximizes ergodic capacity is the scaled identity matrix, i.e. $\mathbf{Q} = \frac{P}{M}\mathbf{I}_M$, and thus the transmit power is divided equally among all the transmit antennas. Thus the ergodic capacity is given by:

$$C = \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_N + \frac{P}{M}\mathbf{H}\mathbf{H}^{\mathbf{H}}\right|\right]. \tag{11}$$

An integral form of this expectation involving Laguerre polynomials is derived in [113].

Results on the asymptotic behavior of ergodic capacity as either the SNR or the number of antennas are taken to infinity are useful for gaining intuition. Though these results are asymptotic, they illustrate many

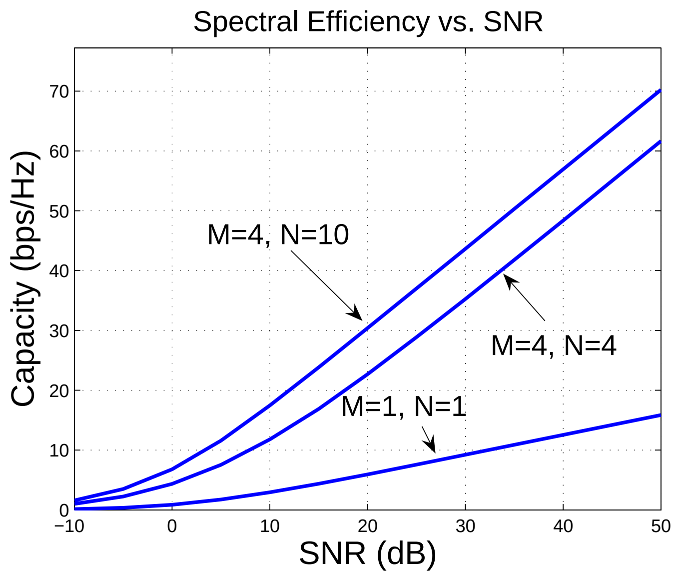

Fig. 6. Ergodic Capacity with CSIR/CDIT vs. SNR

features that are applicable for moderate SNR values and relatively small antenna arrays. If M and N are fixed and the SNR (P) is taken to infinity, capacity grows approximately as $C \approx \min(M, N) \log_2 P + O(1)$. Thus the ergodic capacity has a multiplexing gain of $\min(M, N)$, i.e. each 3 dB of SNR leads to an increase of $\min(M, N)$ bps/Hz in spectral efficiency[4]. Plots of ergodic capacity for 1×1, 4×4 and 4×10 systems are shown in Figure 6. Notice that the linear growth of capacity with respect to SNR begins around 10 dB, which is quite reasonable. The 1×1 system has a slope of 1 bit/3 dB, whereas the 4×4 and 4×10 systems both have a slope of 4 bits/3 dB. Notice also that the 4×4 and 4×10 systems have the same slope but different constant terms, i.e. the 4×10 system has a power gain relative to the 4×4 system. Expressions for the constant term (for the ZMSW model as well as the CCI and CMI model) can be found in the literature on the high SNR behavior of MIMO channels, e.g. in [82] [115].

It is also possible to study the asymptotic behavior as the number of transmit and/or receive antennas is taken to infinity while the transmit power is kept fixed. If the number of transmit antennas (M) is taken to infinity while keeping N fixed, the capacity is bounded in M and converges to $N \log(1 + P)$ [113]. This is due to the fact that a fixed amount of transmit power is equally divided between more and more antennas. If the number of receive antennas N is taken to infinity while keeping M fixed, the capacity does indeed go to infinity approximately as $\log(N)$. The key difference is that adding receive antennas increases the amount of received power, while adding transmit antennas does not do so because power

[4]This is related to the fact that the rank of $\mathbf{H}$ is $\min(M, N)$ with probability one for the ZMSW model.

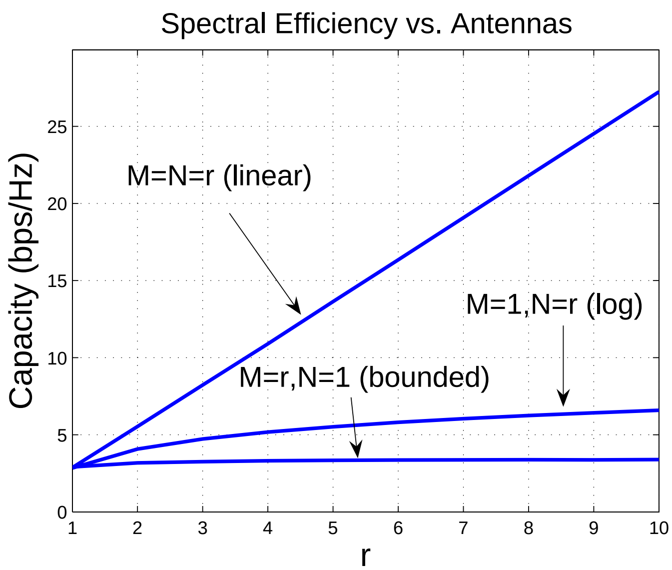

Fig. 7. Ergodic Capacity with CSIR/CDIT vs. Antennas

is split between all transmit antennas. If M and N are simultaneously taken to infinity, capacity is seen to grow *linearly* with $\min(M, N)$, i.e. $C \approx \min(M, N) \cdot c$, where c is a constant depending on the ratio of M and N and on the SNR. Expressions for the growth rate constant can be found in [113] [43]. In summary, increasing the number of receive antennas leads to logarithmic growth in capacity, while simultaneously increasing the number of receive and transmit antennas leads to linear growth. Increasing the number of transmit antennas, on the other hand, provides only a bounded increase in capacity. These points are illustrated in Fig. 7, where capacity versus the number of antennas is plotted. In the linear curve, the number of transmit and receive antennas are set equal to the x-axis parameter r. In the second curve, only the number of receive antennas is increased (i.e. $N = r$) while M is kept at one, leading to logarithmic growth. In the final curve, only the number of transmit antennas is increased (i.e. $M = r$) while N is kept at one, leading to only bounded growth. One crucial point to take away from this section is that the capacity of an $r \times r$ MIMO system is approximately r times the capacity of a SISO system at *any SNR*. This approximation has a maximum error of around 10%, and thus is extremely accurate.

In general, vector codebooks are needed to achieve capacity on a channel with multiple inputs. The decoding complexity for vector codebooks can increase exponentially with the number of inputs. Therefore capacity achieving schemes that use scalar codes are of much practical interest. For the two transmit one receive antenna ZMSW case, the Alamouti coding scheme is an extremely simple method to achieve capacity (** reference to Calderbank chapter**). Such a scheme, however, cannot be generalized to an arbitrary number of transmit and receive antennas. BLAST (Bell Labs Layered Space Time) is a well known layered architecture which can be used to achieve (or come close to) capacity for an arbitrary number of transmit and receive antennas, and practical implementations have even been shown to provide enormous capacity gains over single antenna systems. For example, at 1% outage, 12 dB SNR, and with 12 antennas, the spectral efficiency is shown to be 32 b/s/Hz as opposed to the spectral efficiencies of around 1 b/s/Hz achieved in present day single antenna systems. While initial results on BLAST assumed uncorrelated and frequency flat fading, practical channels exhibit both correlated fading as well as frequency selectivity. The need to estimate the capacity gains of BLAST for practical systems in the presence of channel fade correlations and frequency selective fading sparked off measurement campaigns reported in [85] [33]. The measured capacities are found to be about 30% smaller than would be anticipated from an idealized model. However, the capacity gains over single antenna systems are still overwhelming. Different low-complexity receiver structures are analyzed in detail in Chapter *** Biglieri & Taricco ***.

In the previous section an expression for the capacity with perfect CSIR and CSIT is given. Of course, capacity with CSIR/CSIT must be larger than with CSIR/CDIT for the ZMSW model. Note that the multiplexing gain is $\min(M, N)$ with either CSIT or CDIT; thus, transmitter CSI can only provide a power or rate gain relative to CDIT. In general CSIT provides the most benefit relative to CDIT at low SNR's (for any number of antennas), and at all SNR's when the number of transmit antennas is strictly larger than the number of receive antennas. This comparison is discussed in more detail in Section 1.2.1 (*** Paulraj-Vu chapter ****).

It is conjectured in [113] that the optimal input covariance matrix that maximizes capacity versus outage is a diagonal matrix with the power equally distributed among a *subset* of the transmit antennas. The principal observation is that as the capacity CDF becomes steeper, capacity versus outage increases for low outage probabilities and decreases for high outage probabilities. This is reflected in the fact that the higher the outage probability, the smaller the number of transmit antennas that should be used. As the transmit power is shared equally between more antennas the expectation of C increases (so the ergodic capacity increases) but the tails of its distribution decay faster. While this improves capacity versus outage for low outage probabilities, the capacity versus outage for high outages is decreased. Usually we are interested in low outage probabilities[1] and therefore the usual intuition for outage capacity is that it

[1] The capacity for high outage probabilities become relevant for schemes that transmit only to the best user. For such schemes, it is shown in [8] that increasing the number of transmit antennas reduces the average sum capacity.

increases as the diversity order of the channel increases, i.e. as the capacity CDF becomes steeper.

C.3 Capacity with Perfect CSIR and CDIT: CMI, CCI, and QCI Models

For MIMO channels the capacity improvement resulting from some knowledge of the short-term channel statistics at the transmitter have been shown to be substantial, igniting much interest in the capacity of MIMO channels with perfect CSIR and CDIT under general distribution models. In this section we focus on the cases of CMI, CCI and QCI channel distributions, corresponding to distribution feedback of the channel mean, covariance matrix or quantized information of the instantaneous channel state, respectively. Key results on the capacity of such channels can be found in [50, 54, 56, 65, 67, 79, 86, 87, 89, 90, 109, 112, 116, 124].

Mathematically the problem is defined by (9) and (10), with the distribution on $\mathbf{H}$ determined by the CMI, CCI, or QCI. The optimum input covariance matrix in general can be a full rank matrix which implies either vector coding across the antenna array or transmission of several scalar codes in parallel with successive interference cancellation at the receiver. Limiting the rank of the input covariance matrix to unity, called *beamforming*, essentially leads to a scalar coded system which has a significantly lower complexity for typical array sizes.

The complexity versus capacity tradeoff is an interesting aspect of capacity results under CDIT. The ability to use scalar codes to achieve capacity under CDIT for different channel distribution models, also called optimality of beamforming, captures this tradeoff and has been the topic of much research in itself. Note that vector coding refers to fully unconstrained signaling schemes for the memoryless MIMO Gaussian channel. Every symbol period, a channel use corresponds to the transmission of a vector symbol comprised of the inputs to each transmit antenna. Ideally, while decoding vector codewords the receiver needs to take into account the dependencies in both space and time dimensions and therefore the complexity of vector decoding grows exponentially in the number of transmit antennas. A lower complexity implementation of the vector coding strategy is also possible in the form of several scalar codewords being transmitted in parallel. It is shown in [50] that without loss of capacity, any input covariance matrix, regardless of its rank, can be treated as several scalar codewords encoded independently at the transmitter and decoded successively at the receiver by subtracting out the contribution from previously decoded codewords at each stage. However, well known problems associated with successive decoding and interference subtraction, e.g. error propagation, render this approach unsuitable for use in practical systems. It is in this context that the question of optimality of beamforming becomes important. Beamforming transforms the MIMO channel into a single input single output (SISO) channel. Thus, well established scalar codec technology can be used to approach capacity and since there is only one beam, interference cancellation is not needed. In the summary given below we include the results on both the transmitter optimization problem as well as the optimality of beamforming. We first discuss multiple input single output channels, followed by multiple input/multiple output channels. Notice that if there is perfect CSIR, a single input multiple output channel can be converted into a SISO channel by use of maximal-ratio combining at the receiver, and therefore we need not consider such channels.

Multiple Input Single Output Channels

We first consider systems that use a single receive antenna and multiple transmit antennas. The channel matrix is rank one. With perfect CSIT and CSIR, for every channel matrix realization it is possible to identify the only non-zero eigenmode of the channel accurately and beamform along that mode. On the other hand, with perfect CSIR and CDIT under the ZMSW model, the optimal input covariance matrix is a multiple of the identity matrix. Thus, the inability of the transmitter to identify the non-zero channel eigenmode forces a strategy where the power is equally distributed in all directions.

For a system using a single receive antenna and multiple transmit antennas, the transmitter optimization problem under CSIR and CDIT is solved for the distribution models of CMI and CCI. For the CMI model ($\mathbf{H} \sim \tilde{\mathcal{N}}(\overline{\mathbf{H}}, \alpha\mathbf{I})$) the principal eigenvector of the optimal input covariance matrix $\mathbf{Q^o}$ is along the channel mean vector and the eigenvalues corresponding to the remaining eigenvectors are equal [124]. When beamforming is optimal, all power is allocated to the principal eigenvector. For the CCI model ($\mathbf{H} \sim \tilde{\mathcal{N}}(\mathbf{0}, \mathbf{R^t})$) the eigenvectors of the optimal input covariance matrix $\mathbf{Q^o}$ are along the eigenvectors of the transmit fade covariance matrix and the eigenvalues are in the same order as the corresponding eigenvalues of the transmit fade covariance matrix [124]. A general condition that is both necessary and sufficient for optimality of beamforming can be obtained for both the CMI and CCI models by simply taking the derivative of the capacity expression [54].

The optimality conditions are plotted in Figure 8. For the CCI model the optimality of beamforming depends on the two largest eigenvalues λ_1, λ_2 of the transmit fade covariance matrix and the transmit power P. Beamforming is found to be optimal when the two largest eigenvalues of the transmit covariance matrix are sufficiently disparate or the transmit power P is sufficiently low. Since beamforming corresponds to using the principal eigenmode alone, this is reminiscent of waterpouring solutions where only the deepest level gets all the water when it is sufficiently deeper than the next deepest level and when the quantity of water is small enough. For the CMI model the optimality of beamforming is found to depend on transmit power P and the *quality of feedback* associated with the mean information, which is

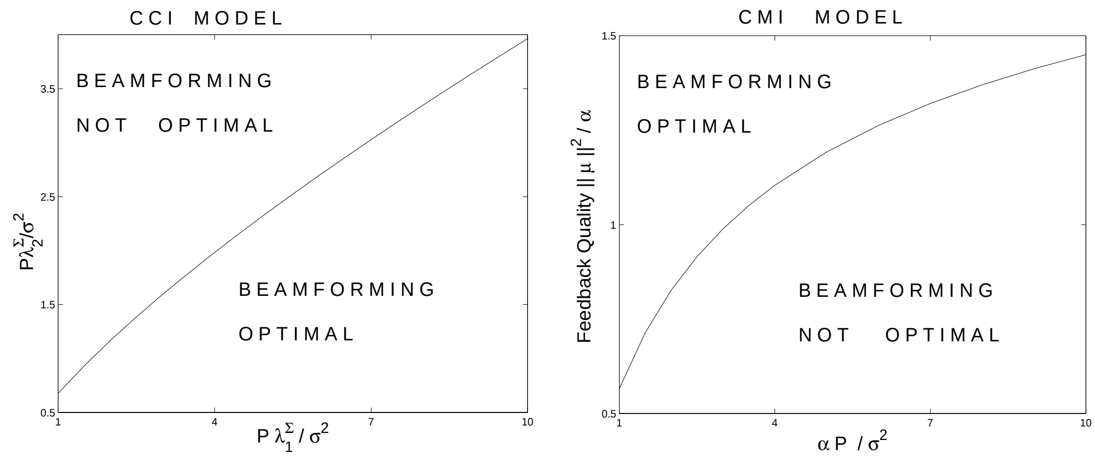

Fig. 8. Conditions for optimality of beamforming.

defined mathematically as the ratio $\frac{||\mu||^2}{\alpha}$. Here, $\mu = ||\overline{\mathbf{H}}||$ is the norm of the channel mean vector. As the transmit power P is decreased or the quality of feedback improves beamforming becomes optimal. As mentioned earlier, for perfect CSIT (uncertainty $\alpha \to 0$ so quality of feedback $\to \infty$) the optimal input strategy is beamforming, while in the absence of mean feedback (quality of feedback $\to 0$ so the CMI model becomes the ZMSW model), the optimal input covariance has full rank, i.e. beamforming is necessarily sub-optimal.

Most research on quantized channel state information has assumed a beamforming transmit strategy where the transmitter forms a beam along the quantized channel vector. Beamforming is known to achieve ergodic capacity when the number of transmit antennas is equal to the number of quantization vectors. This is also known as the antenna selection scenario. In general, for symmetric quantization regions, e.g. if the quantization region consists of all channel vectors that make a maximum angle of $\theta_{\max}$ or less with the quantization vector then for $\theta_{\max} \leq 45^0$, beamforming is optimal not only for ergodic capacity but for outage capacity as well, regardless of the number of quantization vectors or the number of transmit antennas. The outage capacity with quantized beamforming approaches the perfect CSIT case as $(t-1)2^{-B/(t-1)}$, where B is the number of feedback bits and t is the number of transmit antennas. In general, the optimality of beamforming for both ergodic capacity as well as outage capacity depends on the number of quantization vectors as well as the symmetry of the Voronoi regions [112].

Next we summarize the analogous capacity results for MIMO channels.

Multiple Input Multiple Output Channels

With multiple transmit and receive antennas, capacity with CSIR and CDIT under the CCI model with spatially white fading at the receiver ($\mathbf{R^r} = \mathbf{I}$) shows similar behavior as the single receive antenna case. The capacity achieving input covariance matrix has the same eigenvectors as the transmit fade covariance matrix and the eigenvalues are in the same order as the corresponding eigenvalues of the transmit fade covariance matrix [50] [67]. A direct differentiation of the capacity expression yields a necessary and sufficient condition for optimality of beamforming in this case as well. While the receive fade correlation matrix does not affect the eigenvectors of the optimal input covariance matrix, it does affect the eigenvalues as well as the corresponding capacity. The general condition for optimality of beamforming depends upon the two largest eigenvalues of the transmit covariance matrix and all the eigenvalues of the receive covariance matrix.

Transmitter optimization under the CMI model with multiple transmit and receive antennas is also similar to the single receive antenna case. The eigenvectors of the capacity achieving input covariance matrix coincide with the eigenvectors of $\overline{\mathbf{H}}^H\overline{\mathbf{H}}$, where $\overline{\mathbf{H}}$ is the mean value of the random channel matrix [50] [119] [45]. The capacity for the CMI model has been shown to be monotonic in the singular values of $\overline{\mathbf{H}}$ [45].

These results summarize our discussion of channel capacity with CDIT and perfect CSIR under different channel distribution models. From these results we notice that the benefits of adapting to distribution information regarding CMI or CCI fed back from the receiver to the transmitter are two-fold. Not only does the capacity increase with more information about the channel distribution, but this feedback also allows the transmitter to identify the stronger channel modes and achieve this higher capacity with simple scalar codewords.

We conclude this subsection with a discussion on how capacity grows as the number of antennas increases. With perfect CSIR and CDIT under the ZMSW channel distribution, it was shown in [30, 113] that the channel capacity grows linearly with $\min(M, N)$. This linear increase occurs whether the transmitter knows the channel perfectly (perfect CSIT) or only knows its distribution (CDIT). The proportionality constant of this linear increase, called the rate of growth, has also been characterized in [18,42,110,113]. With perfect CSIR and CSIT, the rate of growth of capacity with $\min(M, N)$ is reduced by channel fading correlations at high SNR but is increased at low SNR. The mutual information under CSIR increases linearly with min(M, N) even when a spatially-white transmission strategy is used on a correlated fading channel, although the slope is reduced relative to the uncorrelated fading channel. As we will see in the next section, the assumption of perfect CSIR is crucial for the linear growth behavior of capacity with the number of antennas. Interestingly, it has been shown in [16] that the effect of correlation is not significant when the maximum correlation between pairs of antenna elements is less than 0.5.

In the next section we explore the capacity when only CDI is available at the transmitter and the receiver.

C.4 Capacity with CDIT and CDIR: ZMSW Model

We saw in the last section that with perfect CSIR, channel capacity grows linearly with the minimum of the number of transmit and receive antennas. However, reliable channel estimation may not be possible for a mobile receiver that experiences rapid fluctuations of the channel coefficients. Since user mobility is the principal driving force for wireless communication systems, the capacity behavior with CDIT and CDIR under the ZMSW distribution model (i.e. $\mathbf{H}$ is distributed as $\mathbf{H}^{\mathbf{w}}$ with no knowledge of $\mathbf{H}$ at either the receiver or transmitter) is of particular interest. In this section we summarize some MIMO capacity results in this area.

With CDIR and CDIT and the ZMSW model, in a block fading scenario the channel matrix components are modeled as i.i.d. complex Gaussian random variables that remain constant for a coherence interval of T symbol periods, after which they change to another independent realization. In the block fading model, the effective input to the channel is the input over the duration of the length T block. Capacity is achieved when the $T \times M$ transmitted signal matrix is equal to the product of two statistically independent matrices: a $T \times T$ isotropically distributed unitary matrix multiplied with a certain $T \times M$ random matrix that is diagonal, real, and nonnegative [83]. This result enables the computation of capacity for many interesting cases. Not unexpectedly, for a fixed number of antennas, as the length of the coherence interval T increases, the capacity approaches the capacity that would be obtained if the receiver knew the propagation coefficients. However, there is a surprising result obtained for this channel model: in contrast to the linear growth of capacity with min(M, N) under the perfect CSIR assumption, in the absence of CSIT and CSIR, capacity does not increase at all as the number of transmit antennas is increased beyond the length of the coherence interval T. At high SNRs capacity is achieved using no more than $M^\star =$ min$(M, N, \lfloor T/2 \rfloor)$ transmit antennas [145]. In particular, having more transmit antennas than receive antennas does not provide any capacity increase at high SNR. For each 3-db SNR increase, the capacity gain is $M^\star \left(1 - \frac{M^\star}{T}\right)$.

Crucial to these results is the assumption of a block fading model, i.e. the channel fade coefficients are assumed to be constant for a block of T symbol durations. A more realistic model is that of continuous fading within a block, where it is assumed that within each independent T-symbol block, the fading coefficients have an arbitrary time correlation. If the correlation vanishes beyond some lag τ, called the *correlation time* of the fading, then it is shown in [84] that increasing the number of transmit antennas beyond min(τ, T) antennas does not increase capacity. It is important to note that the assumption of independent fading between blocks has a significant impact on capacity behavior. By contrast, it is shown in [74] that without independent fading between blocks, the capacity at high SNR for the CDIT/CDIR model with the ZMSW distribution grows only double logarithmically in SNR. This result is shown to hold under very general conditions, even allowing for memory and partial receiver side information.

C.5 Capacity with CDIT and CDIR: CCI Model

The results described in the previous section assume a somewhat pessimistic model for the channel distribution. That is because most channels, when averaged over a relatively small area, have either a nonzero mean or a nonwhite covariance. Thus, if these distribution parameters can be tracked, the channel distribution corresponds to either the CMI or CCI model.

With CDIT and CDIR under the CCI distribution model, the channel matrix components are modeled as spatially correlated complex Gaussian random variables that remain constant for a coherence interval of T symbol periods, after which they change to another independent realization based on the spatial correlation model. The channel correlations are assumed to be known at the transmitter and receiver. As in the case of spatially white fading (ZMSW model), with the CCI model the capacity is achieved when the $T \times M$ transmitted signal matrix is equal to the product of a $T \times T$ isotropically distributed unitary matrix, a statistically independent $T \times M$ random matrix that is diagonal, real, and nonnegative, and the matrix of the eigenvectors of the transmit fade covariance matrix $\mathbf{R^t}$ [52]. The channel capacity is independent of the smallest $(M - T)^+$ eigenvalues of the transmit fade covariance matrix as well as the eigenvectors of the transmit and receive fade covariance matrices $\mathbf{R^t}$ and $\mathbf{R^r}$. Also, in contrast to the results for the spatially white fading model, where adding more transmit antennas beyond the coherence interval length ($M > T$) does not increase capacity, additional transmit antennas always increase capacity as long as their channel fading coefficients are spatially correlated. Thus, in contrast to the results in favor of independent fades with perfect CSIR, these results indicate that with CCI at the transmitter and the receiver, transmit fade correlations can be beneficial, making the case for minimizing the spacing between transmit antennas when dealing with highly mobile, fast fading channels that cannot be accurately measured. Mathematically, for fast fading channels ($T = 1$), capacity is a Schur-concave function of the vector of eigenvalues of the transmit fade correlation matrix. The maximum possible capacity gain due to transmitter fade correlations is shown to be $10 \log_{10} M$ db in terms of power.

C.6 Capacity with Correlated Fading

The impact of channel correlations on the capacity of a MIMO channel is of interest because the channels encountered in practice invariably exhibit non-zero correlations in time and space. Temporal correlations are those that exist between the channel matrix realizations at different time instants. Spatial correlations are those that exist between the elements of the channel matrix for each realization.

We first consider the impact of temporal correlations. In general, if perfect CSIR is not assumed, a single letter capacity characterization for a temporally correlated channel is hard to obtain because the correlations introduce memory in the channel. Typically if channel memory is limited to a block of length τ then we need to consider the τ symbol extension of the channel. Because the τ symbol extension of the channel has an input alphabet size that is exponentially larger (exponential in the memory τ), the complexity of the input optimization problem for channels with memory can increase exponentially with τ. While a complete characterization of the impact of temporal correlations is not known, it is easy to see that temporal correlations will increase the capacity when no CSIR is available. This is because the channel correlations allow some amount of channel estimation that is not possible in a memoryless channel. It makes sense intuitively that channel capacity assuming temporal correlation cannot be smaller than that of a temporally uncorrelated channel since, by interleaving codewords, it is often possible to transform the correlated channel into an uncorrelated one.

Capacity characterization with temporal correlations is significantly simpler when the channel realizations are assumed known perfectly to the receiver (perfect CSIR). In this case, temporal channel correlations do not affect ergodic capacity. This is because, conditioned on the channel knowledge available at the receiver, the channel randomness is only due to the additive noise, which is memoryless from one symbol to next. Therefore, with perfect CSIR, a single letter characterization of the capacity is possible and the ergodic capacity depends only on the marginal distribution of a single realization of the channel matrix. Note that while capacity with perfect CSIR is independent of the temporal correlations, the performance of practical coding schemes is affected by the temporal correlations. On the one hand, strong temporal correlations signify a slowly varying channel that would require longer codewords to realize the ergodic capacity. On the other hand, many low complexity coding schemes such as orthogonal space-time codes rely on the channel remaining constant over several symbols (e.g. the Alamouti scheme) and therefore may perform better for slowly varying (high temporal correlation) channels.

Next we discuss the impact of spatial correlations. Spatial correlations are a function of the scattering environment and the antenna spacing. Roughly speaking, correlation between fades experienced by different antennas decreases as the density of scatterers in the vicinity increases or as the spacing between the antennas increases. For example, elevated base stations located in relatively unobstructed surroundings have a larger decorrelating distance between antennas than an indoor mobile device surrounded by scatterers. Since a mobile unit is more size-constrained than a base station, the two factors offset each other.

Completely general models for spatial correlations are often analytically intractable and remain an active area of research. The most commonly used model is the Kronecker product form

$$\mathbf{H} = \mathbf{R}_r^{1/2}\mathbf{H}_w\mathbf{R}_t^{1/2}, \tag{12}$$

where $\mathbf{H}_w$ is an i.i.d. Rayleigh fading channel, and $\mathbf{R}_t$, $\mathbf{R}_r$ are the transmit and receive correlation matrices, respectively. This model is attractive for its analytical tractability and also shown to be reasonably accurate through field measurements. The impact of transmit matrix correlation has been explored in some detail while the impact of receiver matrix correlation is not as easily characterized. Here, using the Kronecker product model as our reference, we discuss the intuition behind the impact of transmitter matrix correlation on the capacity, and also provide references in the literature that analytically validate this intuition. First let us consider the case of perfect CSIR. In particular, to develop some intuition, let us consider the cases of a perfectly correlated channel (rank$(\mathbf{R}_t) = 1$) and a perfectly uncorrelated channel $(\mathbf{R}_t = \mathbf{I}_M)$, as well as the two extremes of high and low SNR. Notice that the perfectly correlated channel has unit rank. Therefore, correlation can make the channel matrix rank-deficient. This is an important limitation at high SNR where the rank of the channel matrix determines the capacity (multiplexing gain). On the other hand, consider a MIMO system at low SNR. At low SNR, if the transmitter can identify the strongest eigenmode of the channel, the optimal transmit strategy is to beamform along the strongest channel mode. Thus, at low SNR the channel capacity is limited by the strength of the strongest channel mode and the ability of the transmitter to beamform in that direction; the rank of the channel matrix does not play an important role in capacity at low SNR. Moreover, correlation enhances the principal eigenmode of the channel at the cost of the remaining eigenmodes. In other words, correlation reduces the multiplexing gain but enhances beamforming along the principle eigenmode. In light of these observations, one would expect that if both the transmitter and the receiver know the channel state (perfect CSIT, perfect CSIR) then transmit matrix correlation will increase capacity at low SNR where beamforming is the optimal policy, and will reduce capacity at high SNR where the the multiplexing gain is the limiting factor. Now suppose the transmitter does not have any CSIT and is forced to split its power uniformly among all transmit antennas. In this case, correlation does not help at low SNR because the transmitter is not able to identify the strongest channel mode along which to direct its transmit power. At high SNR, however, uniform power allocation is nearly optimal (for exceptions to this rule see [81]) and the impact of correlation would be the same as with perfect CSIT. Thus, with uniform power allocation, transmit matrix correlation would reduce capacity both at high SNR as well as at low SNR. Finally, consider the case where the transmitter possesses no CSIT but knows the correlation structure through CDIT. This allows the transmitter to identify the strong channel modes and beamform to them at low SNR. Thus, correlation would be helpful at low SNR. The high SNR case is more complicated. On the one hand, knowledge of the channel correlation structure allows the transmitter to identify the strong channel modes. In fact for a perfectly correlated channel, knowledge of the channel correlations is as good as perfect CSIT, since there is only one non-zero eigenmode (the principal eigenvector of $\mathbf{R}_t$). However, this is offset by the loss in the multiplexing gain of the channel. Because these two factors work in opposition, the impact of transmit matrix correlation is not as easily characterized for a MIMO channel. However, suppose there is only one receive antenna. In that case the spatial multiplexing gain is unity whether the channel is correlated or uncorrelated. Indeed, in this case correlation increases capacity because the transmitter is able to identify the strong channel eigenmodes. Analytical results verifying these intuitive explanations are provided for the MISO channel in [66] and for the MIMO channel in [18].

C.7 Frequency Selective Fading Channels

While flat fading is a realistic assumption for narrowband systems where the signal bandwidth is smaller than the channel coherence bandwidth, broadband communications involve channels that experience frequency selective fading. Research on the capacity of MIMO systems with frequency selective fading typically takes the approach of dividing the channel bandwidth into parallel flat fading channels, and constructing an overall block diagonal channel matrix with the diagonal blocks given by the channel matrices corresponding to each of these subchannels. Under perfect CSIR and CSIT, the total power constraint then leads to the usual closed-form waterfilling solution. Note that the waterfill is done simultaneously over both space and frequency. Even SISO frequency selective fading channels can be represented by the MIMO system model (4) in this manner [95]. For MIMO systems, the matrix channel model is derived in [7] based on an analysis of the capacity behavior of OFDM-based MIMO channels in broadband fading environments. Under the assumption of perfect CSIR and CDIT for the ZMSW model, it is shown that in the MIMO case, unlike the SISO case, frequency selective fading channels may provide advantages over flat fading channels not only in terms of ergodic capacity but also in terms of capacity versus outage. In other words, MIMO frequency selective fading channels are shown to provide both higher diversity gain and higher multiplexing gain than MIMO flat-fading channels. The measurements in [85] show that frequency selectivity makes the CDF of the capacity steeper and, thus, increases the capacity for a given outage as compared with the flat-frequency case, but the influence on the ergodic capacity is small.

C.8 Training for Multiple Antenna Systems

The results summarized in the previous sections indicate that CSI plays a crucial role in the capacity of MIMO systems. In particular, the capacity results in the absence of CSIR are strikingly different and often quite pessimistic compared to those that assume perfect CSIR. To recapitulate, with perfect CSIR and CDIT MIMO channel capacity is known to increase linearly with min(M, N) when the CDIT assumes the ZMSW or CCI distribution models. However in fast fading, when the channel changes so rapidly that it cannot be estimated reliably at the receiver (CDIR only), the capacity does not increase with the number of transmit antennas at all for $M > T$, where T is the channel decorrelation time. Also at high SNR under the ZMSW distribution model, capacity with perfect CSIR and CDIT increases logarithmically with SNR, while the capacity with CDIR and CDIT increases only double logarithmically with SNR. Thus, CSIR is critical for obtaining the high capacity benefits of multiple antenna wireless links. CSIR is often obtained by sending known training symbols to the receiver. However, with too little training the channel estimates are poor, whereas with too much training there is no time for data transmission before the channel changes. So the key question to ask is how much training is needed in multiple antenna wireless links [41]. It turns out that when the training and data powers are allowed to vary, the optimal number of training symbols is equal to the number of transmit antennas, which is also the smallest training interval length that guarantees meaningful estimates of the channel matrix. When the training and data powers are instead required to be equal, the optimal training duration may be longer than the number of antennas. Interestingly, while training-based schemes can be optimal at high SNR, they are suboptimal at low SNR.

C.9 Application to Matrix Channels

Note that the MIMO capacity results described in the prior sections are applicable to any channel described by a matrix. Matrix channels describe not only multiantenna systems but also channels with crosstalk [140] and wideband channels [120]. While the focus of this chapter is on flat and frequency-selective fading channels, the same capacity analysis can be applied to obtain the capacity of these matrix channels as well.

D. Open Problems in Single-User MIMO

The results summarized in this section form the basis of our understanding of channel capacity under different CSI and CDI assumptions. These results serve as useful indicators for the benefits of incorporating training and feedback schemes in a MIMO wireless link to obtain CSIR/CDIT and CSIT/CDIT, respectively. However, our knowledge of MIMO capacity with CDI only is still far from complete, even for single user systems. We conclude this section by pointing out some of the many open problems. (1) Combined CCI and CMI: Capacity under CDIT and perfect CSIR is unsolved under a combined CCI and CMI distribution model, even with a single receive antenna. (2) CCI: With perfect CSIR and CDIT capacity is not known under the CCI model for completely general (i.e. non-separable) spatial correlations. (3) CDIR: Almost all cases with only CDIR are open problems. (4) Outage capacity: Most results for CDI only at either the transmitter or receiver are for ergodic capacity. Capacity versus outage has proven to be less analytically tractable than ergodic capacity and contains an abundance of open problems.

IV. MULTI-USER MIMO

In this section we give capacity results for the two basic multi-user MIMO channel models: the MIMO multiple-access channel (MAC) and the MIMO broadcast channel (BC). The MIMO MAC consists of many multiple-antenna transmitters sending to a single multiple-antenna receiver, and the MIMO BC consists of one multiple-antenna transmitter sending to many multiple-antenna receivers. In cellular-type architectures (e.g. cellular networks or wireless local-area networks), the MAC, or uplink, models the channel from mobile devices to the base station, and the BC, or downlink, models the channel from base station to mobile devices. The uplink and downlink channels are illustrated in Fig. 9. Multiple antennas are becoming increasingly common in such systems (consider IEEE 802.11n or IEEE 802.16), and thus it is important to understand the fundamental limits of such channels.

The channel capacity of a point-to-point MIMO channel is a real number which is the fundamental limit on reliable communication: any rate strictly smaller than capacity is achievable, i.e. there exist codes at such a rate that can achieve any desired probability of error, and any rate strictly larger than capacity is not achievable in the sense that the probability of error is bounded away from zero. For multi-user channels, channel capacity has a similar definition, but capacity is a region of rates (i.e. a set of rate vectors in K-dimensional space) instead of a single number. In the MAC, each transmitter is assumed to have an independent message for the base station, and thus a different rate is associated with each transmitter. In the BC, the transmitter is assumed to have a different (and independent) message for

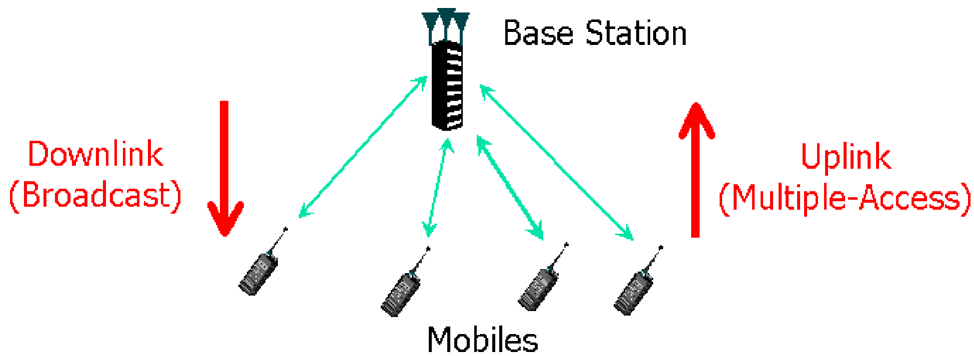

Fig. 9. Uplink/Downlink Channels in Cellular Systems

each of the receivers[5], and similarly a different rate is associated with each transmission. The capacity region is therefore defined as the set of user rates that can *simultaneously* be achieved with arbitrarily small probability of error. It is important to note that points on the boundary of the capacity region are typically obtained by sending messages to all users simultaneously; user rates achieved with schemes such as TDMA, where at a given time only a single mobile communicates with the base station or access point on either the uplink or downlink, are contained in the capacity region, but are generally strictly sub-optimal. In fact, contrary to the method in which most systems are currently designed, tremendous capacity benefits can be achieved by simultaneously transmitting to multiple users (on the downlink) or having multiple users simultaneously transmit (on the uplink) without any separation in the time, frequency, or space domain.

The capacity benefits of multi-user MIMO can be even greater than in the single-user setting discussed in Section III. In single-user systems, multiple antennas are required at both transmitter and receiver in order to realize linear capacity gains (i.e. having capacity scale as $\min(M, N)$, where M and N are the number of transmit and receive antennas, respectively). In the uplink and downlink channels, however, it is sufficient to deploy multiple antennas at only the access point in order to achieve a similar linear increase in capacity. In this scenario, if we let M denote the number of access point antennas, N the number of antennas at each mobile, and K the number of mobiles, then sum rate capacity (the maximum throughput, or the point in the capacity region that maximizes the sum of all rates) increases linearly with $\min(M, NK)$ as the number of antennas and users is increased. Thus, having a large number of mobiles can make up for deploying a small number of antennas at each mobile. This is a key point for

[5]In networking terms, this is referred to as unicast. In the multicast scenario, which is not discussed here, there is a single common message which all receivers wish to receive [90] [60] [108] [70] [68].

space-limited mobile devices.

Similar to Section III, we give results on the capacity of the MIMO MAC and MIMO BC for different assumptions about the amount of channel state information available at the transmitters and receivers. Multi-user detection or successive interference cancellation can be used to achieve the capacity of the MIMO MAC. This channel is well understood from an information theoretic point of view, and thus many results are available for the MIMO MAC. The broadcast channel, on the other hand, remains one of the key open problems in information theory. However, there has been a great deal of recent progress on the MIMO BC, and the capacity region is known in some scenarios. A clever pre-processing technique called *dirty paper coding* achieves the capacity of the MIMO BC when the channel is fixed. Interestingly, the MIMO MAC and MIMO BC have been shown to be duals, as we will discuss in Section IV-C.3.

A. System Model

To describe the MAC and BC models, we consider a cellular-type system in which the base station has M antennas, and each of the K mobiles has N antennas. The downlink (or forward channel) of this system is a MIMO BC, and the uplink (or reverse channel) is a MIMO MAC. We will use $\mathbf{H}_i$ to denote the *downlink* channel matrix from the base station to User i. Assuming that the same channel is used on the uplink and downlink (i.e. a TDD system), the *uplink* matrix of User i is $\mathbf{H}_i^H$. A picture of the system model is shown in Fig. 10 [6].

In the MAC, let $\mathbf{x_k} \in \mathbb{C}^{N\times 1}$ be the transmitted signal of User (i.e. mobile) k. Let $\mathbf{y_{MAC}} \in \mathbb{C}^{M\times 1}$ denote the received signal and $\mathbf{n} \in \mathbb{C}^{M\times 1}$ the noise vector where $\mathbf{n} \sim \tilde{\mathcal{N}}(0, \mathbf{I})$ is circularly symmetric complex Gaussian with identity covariance matrix. The received signal at the base station is then equal to:

$$\begin{aligned}\mathbf{y_{MAC}} &= \mathbf{H_1}^H\mathbf{x}_1 + \ldots + \mathbf{H}_K^H\mathbf{x}_K + \mathbf{n}\\ &= \mathbf{H}^H\begin{bmatrix}\mathbf{x}_1\\ \vdots\\ \mathbf{x}_K\end{bmatrix} + \mathbf{n} \quad \text{where} \quad \mathbf{H}^H = \begin{bmatrix}\mathbf{H_1}^H & \ldots & \mathbf{H}_K^H\end{bmatrix}.\end{aligned}$$

In the MAC, each user (i.e. mobile) is subject to an individual power constraint of P_k. The transmit covariance matrix of User k is defined to be $\mathbf{Q}_k \triangleq \mathbb{E}[\mathbf{x}_k\mathbf{x}_k^H]$. The power constraint implies $\mathrm{Tr}(\mathbf{Q}_k) \leq P_k$

[6]The system is assumed to be TDD for mathematical simplicity, and in order to introduce the concept of MAC-BC duality. However, this assumption need not be true in order for the results of this section to hold true. In addition, we consider the dual uplink channel to be the conjugate transpose of the downlink channel. The true channels should only be related through the transpose operation, but we add the conjugate (which does not affect capacity) for mathematical simplicity.

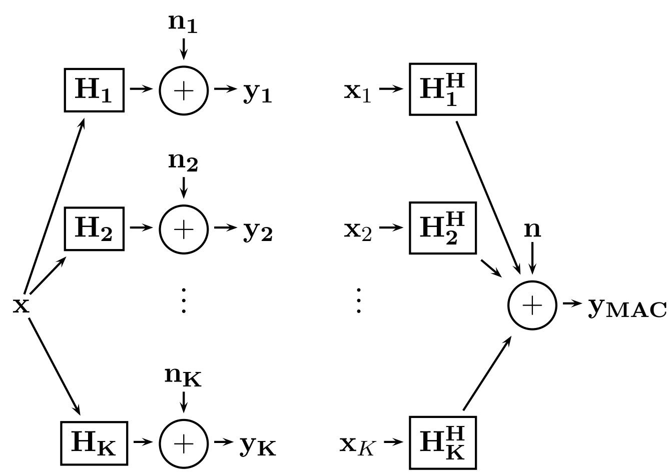

Fig. 10. System models of the MIMO BC (left) and the MIMO MAC (right) channels

for $k = 1, \ldots, K$.

In the BC, let $\mathbf{x} \in \mathbb{C}^{M \times 1}$ denote the transmitted vector signal (from the base station) and let $\mathbf{y_k} \in \mathbb{C}^{N \times 1}$ be the received signal at receiver (i.e. mobile) k. The noise at receiver k is represented by $\mathbf{n_k} \in \mathbb{C}^{N \times 1}$ and is assumed to be circularly symmetric complex Gaussian noise ($\mathbf{n_k} \sim \tilde{N}(0, \mathbf{I})$). The received signal of User k is equal to:

$$\mathbf{y_k} = \mathbf{H_k x} + \mathbf{n_k}. \tag{13}$$

The transmit covariance matrix of the input signal is $\mathbf{\Sigma}_x \triangleq \mathbb{E}[\mathbf{x}\mathbf{x}^H]$. The base station is subject to an average power constraint P, which implies Tr$(\mathbf{\Sigma}_x) \leq P$.

In the MAC, each transmitter is assumed to have an independent data stream (i.e. a sequence of messages) for the receiver. Thus, a different data rate is associated with each transmitter, and the capacity region is a K-dimensional region. Similarly, in the BC the transmitter has an independent message for each of the receivers, and the capacity region is therefore also a K-dimensional region.

Though we consider only multiple antenna channels here, it is important to note that the multiple-antenna MAC and BC with $M > 1$ and $N = 1$ can be used to model single-antenna CDMA systems. The channel vectors (on either the downlink or uplink) represent the spreading codes of the users, with M being the length of each code. A number of works for the uplink channel further study this connection (c.f. [127]).

B. MIMO Multiple-Access Channel

In this section we summarize capacity results on the multiple-antenna MAC. We first provide general results on the capacity region of multiple-access channels to provide some background. We then analyze the constant channel MIMO MAC, followed by the fading channel. Since the capacity region of a general MAC is known, the expressions for the capacity of a constant MAC are quite straightforward. For the fading case, one must consider different assumptions about the CSI and CDI available at the transmitter and receiver.

B.1 Multiple Access Channel Capacity

The capacity region of the general multiple-access channel was first derived in the 1970s [2] [78]. The capacity region of the two transmitter single antenna AWGN MAC (i.e. $M = N = 1$) is given by [23]:

$$\begin{aligned} R_1 &\leq \log(1 + |h_1|^2 P_1) \\ R_2 &\leq \log(1 + |h_2|^2 P_2) \\ R_1 + R_2 &\leq \log(1 + |h_1|^2 P_1 + |h_2|^2 P_2). \end{aligned}$$

We provide the two user capacity region for simplicity; this formula can easily be extended to a MAC with an arbitrary number of transmitters.

In order to achieve the capacity region, each transmitter uses a Gaussian codebook (as in a point-to-point AWGN channel) and the receiver employs either multi-user detection or successive interference cancellation to decode the codes of every transmitter. The capacity region clearly corresponds to a pentagonal region. Consider the corner point $(R_1 = \log(1+|h_1|^2 P_1),\ R_2 = \log(1+\frac{|h_2|^2 P_2}{|h_1|^2 P_1 + 1}))$. In order to achieve this point, the receiver first decodes the message from transmitter 2 while treating the codeword from transmitter 1 (which is normally distributed with power $|h_1|^2 P_1$) as an extra source of additive noise. The receiver subtracts the decoded message from the received signal, and then decodes the message from transmitter 1. During this second decoding operation, note that there is no interference, i.e. transmitter 1 communicates over a "clean" channel to the receiver. The other corner point of the capacity region $((R_1 = \log(1 + \frac{|h_1|^2 P_1}{|h_2|^2 P_2 + 1}),\ R_2 = \log(1 + |h_2|^2 P_2))$ can similarly be achieved by first decoding transmitter 1's signal, followed by the codeword of transmitter 2. The same strategy is also optimal for the MIMO channel, which we discuss next.

B.2 Constant Channel

For any set of power constraints $\mathbf{P} = (P_1, \ldots, P_K)$, the capacity of the MIMO MAC (denoted $\mathcal{C}_{\rm MAC}(\mathbf{P}; \mathbf{H^H})$) is given by:

$$\mathcal{C}_{\rm MAC}(\mathbf{P}; \mathbf{H^H}) \triangleq \bigcup_{\{\mathbf{Q}_i \geq 0, \ \mathrm{Tr}(\mathbf{Q}_i) \leq P_i \ \ \forall i\}} \left\{ \begin{array}{l} (R_1, \ldots, R_K): \\ \sum_{i \in S} R_i \leq \log \left| \mathbf{I}_M + \sum_{i \in S} \mathbf{H}_i^H \mathbf{Q}_i \mathbf{H}_i \right| \ \ \forall S \subseteq \{1, \ldots, K\} \end{array} \right\} \quad (14)$$

The variable S refers to a subset of $\{1, \ldots, K\}$. The i-th user transmits a zero-mean Gaussian vector with spatial covariance matrix $\mathbf{Q}_i$. Each set of covariance matrices $(\mathbf{Q}_1, \ldots, \mathbf{Q}_K)$ corresponds to a K-dimensional polyhedron of achievable rates (i.e. $\{(R_1, \ldots, R_K):$ $\sum_{i \in S} R_i \leq \log \left| \mathbf{I}_M + \sum_{i \in S} \mathbf{H}_i^H \mathbf{Q}_i \mathbf{H}_i \right| \ \ \forall S \subseteq \{1, \ldots, K\}\}$), and the capacity region is equal to the union over all covariance matrices satisfying the trace constraints of all such polyhedrons. The corner points of each polyhedron can be achieved by *successive decoding*, in which users' signals are successively decoded and subtracted out of the received signal. For the two-user case, each set of covariance matrices corresponds to a pentagon, similar in form to the capacity region of the scalar Gaussian MAC. The corner point where $R_1 = \log \left| \mathbf{I}_M + \mathbf{H}_1^H \mathbf{Q}_1 \mathbf{H}_1 \right|$ and $R_2 = \log \left| \mathbf{I}_M + \mathbf{H}_1^H \mathbf{Q}_1 \mathbf{H}_1 + \mathbf{H}_2^H \mathbf{Q}_2 \mathbf{H}_2 \right| - R_1 = \log \left| \mathbf{I}_M + (\mathbf{I}_M + \mathbf{H}_1^H \mathbf{Q}_1 \mathbf{H}_1)^{-1} \mathbf{H}_2^H \mathbf{Q}_2 \mathbf{H}_2 \right|$ corresponds to decoding $\mathbf{x}_2$ first while treating $\mathbf{x}_1$ as noise, then subtracting $\mathbf{x}_2$ from $\mathbf{y}_{MAC}$, and then decoding User 1. Successive decoding can reduce a complex multi-user detection problem into a series of single-user detection steps [37]. Note that the expression for this corner point is simply the matrix version of the expression for the single antenna MAC given in the previous section.

The capacity region of a MIMO MAC for the single transmit antenna case ($N = 1$) is shown in Fig. 11. When $N = 1$, the covariance matrix of each transmitter is a scalar equal to the transmitted power. Clearly, each user should transmit at full power. Thus, the capacity region for a K-user MAC for $N = 1$ is the set of all rate vectors $(R_1, \ldots, R_K)$ satisfying:

$$\sum_{i \in S} R_i \leq \log \left| \mathbf{I}_M + \sum_{i \in S} \mathbf{H}_i^H P_i \mathbf{H}_i \right| \quad \forall S \subseteq \{1, \ldots, K\}. \quad (15)$$

For the two-user case, this reduces to the simple pentagon seen in Fig. 11.

When $N > 1$, a union must be taken over all covariance matrices. Intuitively, the set of covariance matrices that maximize R_1 are different from the set of covariance matrices that maximize the sum rate. In Figure 12, a MAC capacity region for $N > 1$ is shown. Notice that the region is equal to the union of pentagons (each pentagon corresponding to a different set of transmit covariance matrices), a few of which are shown with dashed lines in the figure. The boundary of the capacity region is in general curved, except

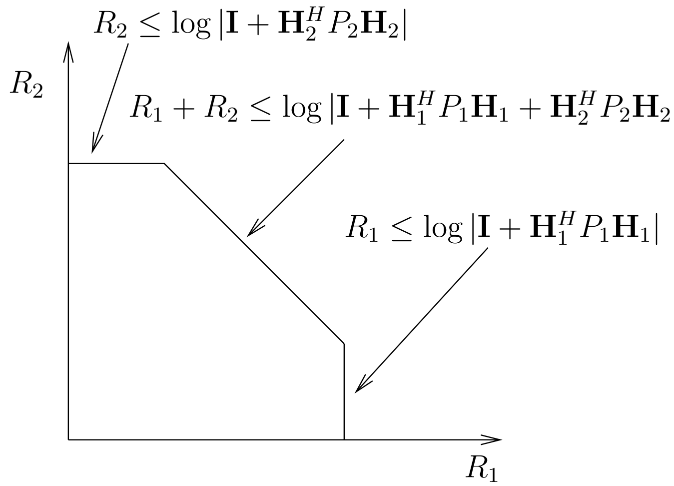

Fig. 11. Capacity region of MIMO MAC for $N = 1$.

at the sum rate point, where the boundary is a straight line, and along the planes where the single-user capacities are achieved [141]. Each point on the curved portion of the boundary is achieved by a *different* set of covariance matrices. At point A, User 1 is decoded last and achieves its single-user capacity by choosing $\mathbf{Q}_1$ as a water-fill of the channel $\mathbf{H}_1$ (i.e. the input that achieves the capacity of the constant MIMO channel from X_1 to Y, as described in Chapter III-B). User 2 is decoded first, in the presence of interference from User 1, so $\mathbf{Q}_2$ is chosen as a waterfill of the channel $\mathbf{H}_2$ and the interference from User 1. The sum-rate corner points B and C are the two corner points of the pentagon corresponding to the sum-rate optimal covariance matrices $\mathbf{Q}_1^{sum}$ and $\mathbf{Q}_2^{sum}$. Note that these covariances are not individually optimal for either user, but instead are optimal in terms of the total amount of rate that the receiver can decode. At point B User 1 is decoded last whereas at point C User 2 is decoded last. Successive decoding can be used to achieve the corner points of the sum capacity plane, and time division between different decoding orders or multi-user detection can be used to achieve interior points.

Next, we focus on characterizing the optimal covariance matrices $(\mathbf{Q}_1, \ldots, \mathbf{Q}_K)$ that achieve different points on the boundary of the MIMO MAC capacity region. Since the MAC capacity region is convex, it is well known from convex theory that the boundary of the capacity region can be fully characterized by maximizing the function $\mu_1 R_1 + \cdots + \mu_K R_K$ over all rate vectors in the capacity region and for all non-negative priorities $(\mu_1, \ldots, \mu_K)$ such that $\sum_{i=1}^K \mu_i = 1$. For a fixed set of priorities $(\mu_1, \ldots, \mu_K)$, this is equivalent to finding the point on the capacity region boundary that is tangent to a line whose slope is defined by the priorities. An example of this tangent line is shown in Fig. 12. The structure of the MAC capacity region implies that all boundary points of the capacity region (except for the plane

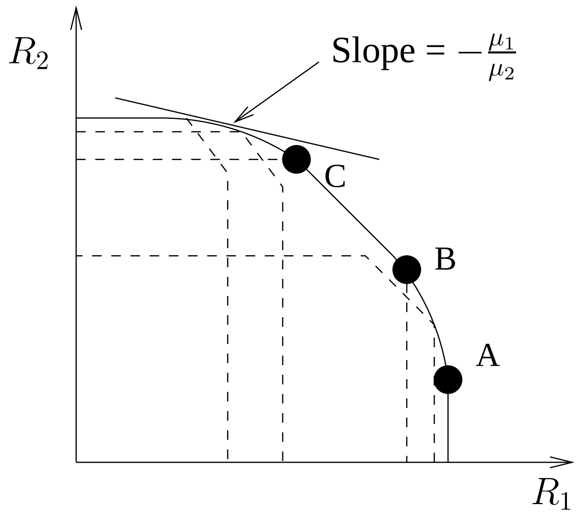

Fig. 12. Capacity region of MIMO MAC for $N > 1$.

defining the sum capacity) are corner points of polyhedrons corresponding to different sets of covariance matrices. Furthermore, the corner point should correspond to successive decoding in order of *increasing* priority, i.e. the user with the highest priority should be decoded last and therefore sees no interference [114] [122]. Thus, the problem of finding the boundary point on the capacity region associated with priorities $\mu_1, \ldots, \mu_K$ assumed to be in descending order (users can be arbitrarily re-numbered to satisfy this condition) can be written as

$$\max_{\mathbf{Q}_1,\ldots,\mathbf{Q}_K} \mu_K \log \left| \mathbf{I}_M + \sum_{l=1}^{K} \mathbf{H}_l^H \mathbf{Q}_l \mathbf{H}_l \right| + \sum_{i=1}^{K-1} (\mu_i - \mu_{i+1}) \log \left| \mathbf{I}_M + \sum_{l=1}^{i} \mathbf{H}_l^H \mathbf{Q}_l \mathbf{H}_l \right|, \tag{16}$$

subject to power constraints on the trace of each of the covariance matrices. Note that the covariances that maximize the function above are the *optimal* covariances. The optimization problem is convex, and thus efficient numerical tools can be employed to solve it [9]. Iterative waterfilling is an extremely efficient algorithm that finds the sum-rate maximizing (i.e. $\mu_1 = \cdots = \mu_K$) covariance matrices [141]. This algorithm is a coordinate descent method, where each user greedily optimizes his own transmit covariance while treating interference from all the other users as additional noise.

A key point regarding the sum rate capacity of the MIMO MAC is its first order growth term. It is in fact easy to see that a MIMO MAC with K transmitters with N antennas each, and a single M antenna receiver, is closely related to an NK transmit antenna, M receive antenna point-to-point MIMO channel. Furthermore, these two channels have the same multiplexing gain, i.e. the sum capacity of the MIMO MAC can be approximated as $\min(M, NK)\log(SNR)$ [61]. Thus, for systems with large number of users, capacity can be increased almost linearly by increasing the number of base station antennas. This is a key benefit of MIMO in multi-user systems.

B.3 Fading Channels

As in the single-user case, the capacity of the fading MIMO MAC depends on the definition of capacity and the availability of CSI and CDI at the transmitters and the receiver. The capacity with perfect CSIT and CSIR is very well studied, as is the capacity with perfect CSIR and CDIT under the ZMSW model. However, little is known about the capacity of the MIMO MAC with CDIT under the CMI or CCI distribution models. Some results on the optimum distribution for the single antenna case with CDIT and CDIR under the ZMSW distribution can be found in [101].

With perfect CSIR and CSIT the system can be viewed as a set of *parallel* non interfering MIMO MACs (one for each fading state) sharing a common power constraint. Thus, the ergodic capacity region can be obtained as an average of these parallel MIMO MAC capacity regions [142], where the averaging is done with respect to the channel statistics. The iterative waterfilling algorithm of [141] extends to this case, with joint space and time waterfilling.

The capacity region of a single-antenna MAC with perfect CSIR and CDIT was found in [31, 102]. These results can easily be extended to MIMO channels. In this scenario, Gaussian inputs are optimal, and the ergodic capacity region is equal to the time average of the capacity obtained at each fading instant with a constant transmit policy (i.e. the input covariance matrix for each transmitter is fixed for all time). Thus, the ergodic capacity region is given by

$$\bigcup_{\{\mathbf{Q}_i \geq 0,\ \mathrm{Tr}(\mathbf{Q}_i) \leq P_i \ \ \forall i\}} \left\{ (R_1, \ldots, R_K) : \sum_{i \in S} R_i \leq \mathbb{E}_{\mathbf{H}} \left[\log \left| \mathbf{I}_M + \sum_{i \in S} \mathbf{H}_i^H \mathbf{Q}_i \mathbf{H}_i \right| \right] \ \ \forall S \subseteq \{1, \ldots, K\} \right\}.$$

Note that this is identical to the expression for the capacity of the constant MIMO MAC with the addition of the expected value over the distribution of the channels. The boundary of the capacity region can be characterized by the maximization in (16), but taking the expectation over the fading distribution can make this problem computationally difficult.

If the channel matrices $\mathbf{H}_i$ are ZMSW and each user has the same power constraint (and there is perfect CSIR and CDIT), then the optimal covariances are scaled versions of the identity matrix, i.e. $\mathbf{Q}_i = \frac{P_i}{N}\mathbf{I}$ [113]. In this scenario a single set of covariance matrices achieves the entire capacity region. This is not in general true for other fading distributions, i.e. different covariances may be required to achieve different points on the capacity region boundary, as is the case for the MIMO MAC with no fading. The sum rate capacity of the MAC (achieved using scaled identity covariance matrices) is equal to:

$$\begin{aligned}
\mathcal{C}_{\mathrm{MAC}}^{sum}(\mathbf{P};\mathbf{H^H}) &= \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_M + \sum_{i=1}^{K}\mathbf{H}_i^H\left(\frac{P_i}{N}\mathbf{I}_N\right)\mathbf{H}_i\right|\right] & (17)\\
&= \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_M + \frac{P_i}{N}\mathbf{H}^H\mathbf{H}\right|\right], & (18)
\end{aligned}$$

where P_i is the i-th transmitter's power constraint and we have assumed $P_i = P$ for all i. Note that this expression is exactly the ergodic capacity of the single-user MIMO channel with NK transmit antennas, M receive antennas, the ZMSW distribution model, and perfect CSIR and CDIT, as given in (10). Therefore, the lack of cooperation between the K transmitters does not reduce capacity under this fading model, and the MAC achieves the same capacity as the fully cooperative model. This implies that the MAC sum capacity scales as $\min(M, NK)$ with perfect CSIR and CDIT. The MIMO MAC with perfect CSIR and CDIT under the CCI and CMI models has also been investigated, but only limited results are known [55] [44].

For MIMO multiple-access channels, it is generally sufficient to have perfect CSI at the receiver and CSIT is not crucial to obtain capacity benefits from MIMO. This is because for fading multiple-access channels, independent Gaussian codewords can be transmitted from each antenna (for each user). The subsequent sum rate is given by the right hand side of (18). This technique is optimal when all users have the same power constraint and each channel is ZMSW, but is generally sub-optimal in other scenarios. However, this simple transmission scheme clearly achieves a sum capacity similar to the capacity of the point-to-point MIMO channel from the NK aggregate transmit antennas to an M antenna receiver, i.e. sum capacity grows as $\min(M, NK)\log(SNR)$. Thus for systems with large numbers of users, increasing the number of receive antennas at the base station (M) while keeping the number of mobile antennas (N) constant can lead to linear growth.

Though CSIR and CDIT are sufficient to achieve the linear capacity growth described above, the availability of CSIT increases capacity further, though not in the first order growth of min(M, NK). As seen for point-to-point MIMO channels, CSIT gives the largest performance increases at low SNR's, and the value of CSIT disappears at high SNR. Furthermore, if the number of base station antennas (M) and the number of users (K) are taken to infinity at a fixed ratio, the gain due to CSIT relative to CDIT vanishes [127].

C. *MIMO Broadcast Channel*

In this section we summarize capacity results on the multiple-antenna BC. We first discuss the general broadcast channel and the single antenna broadcast channel, followed by an explanation of dirty paper coding for the MIMO BC. We then establish a duality relationship between the MAC and the BC, which leads into sections on the capacity of the MIMO BC.

C.1 Broadcast Channel Capacity

Unlike the multiple-access channel, a general expression for the capacity region of the broadcast channel is unknown. In fact, this is one of the most fundamental unanswered questions in multi-user information theory. However, the capacity regions for certain classes of broadcast channels are known. Amongst those is the class of degraded broadcast channels, which are channels where receivers can be absolutely ranked in terms of their channel quality [23]. The single antenna AWGN BC falls into this class (users are ranked in terms of the absolute value of their channel strength), and the capacity region of a two user channel is given by all rate pairs satisfying:

$$\begin{aligned} R_1 &\leq \log(1+\alpha|h_1|^2 P) \\ R_2 &\leq \log\left(1+\frac{|h_2|^2(1-\alpha)P}{\alpha|h_2|^2P+1}\right) \end{aligned}$$

for some $\alpha \in [0,1]$. Without loss of generality we have assumed $||h_1|^2| > ||h_2|^2|$. In order to achieve the capacity region, the codewords for the different receivers are superimposed on one another, and successive interference cancellation is used at the receivers. For the AWGN channel, the transmitter generates independent Gaussian codewords for each receiver (with power αP for receiver 1, and power $(1-\alpha)P$ for receiver 2) and transmits the sum of these codewords. Receiver 1, which has the larger channel gain, first decodes the codeword intended for receiver 2, subtracts this from the received signal, and then decodes its intended codewords. Receiver 2 is not able to first decode the codeword for receiver 1 because $||h_1|^2| > ||h_2|^2|$ by assumption. Thus, receiver 2 treats the codeword for receiver 1 as additional noise (with power αP) while decoding its intended codeword.

However, when the transmitter has more than one antenna, the Gaussian broadcast channel is generally non-degraded. This is because matrix channels can only be partially ordered, i.e. there is no absolute ordering of channels. As a result, the successive decoding technique which is capacity-achieving for the single antenna broadcast channel is not effective in the MIMO setting, and an alternative technique must be used, as we describe next.

C.2 Dirty Paper Coding Achievable Rate Region

As noted earlier, successive cancellation cannot effectively be performed by receivers in a MIMO BC to reduce multi-user interference. However, *dirty paper coding* (DPC) can be used at the transmitter to essentially pre-subtract multi-user interference [19] [11] [138]. The resulting rates mimic what one would expect from successive cancellation. The basic premise of dirty paper coding can be illustrated by considering a point-to-point AWGN channel with additive interference: $y = x + s + n$, where x is the transmitted signal (subject to power constraint P), y is the received signal, n is the Gaussian noise, and s is additive interference. If the transmitter but not the receiver has perfect, non-causal knowledge of the interference s, then the capacity of the channel is the same as if there was no additive interference (i.e. $\log(1 + \frac{P}{N})$). Dirty paper coding is a technique that allows non-causally known interference to be "pre-subtracted" at the transmitter, but in such a way that the transmit power is not increased. A more practical and general technique to perform interference pre-subtraction based on nested lattice codes is provided in [27].

In the MIMO BC, dirty paper coding can be applied at the transmitter when choosing codewords for different receivers. The transmitter first picks a codeword (i.e. $\mathbf{x_1}$) for receiver 1. The transmitter then chooses a codeword for receiver 2 (i.e. $\mathbf{x_2}$) with full (non-causal) knowledge of the codeword intended for receiver 1. Therefore the codeword of User 1 can be pre-subtracted such that receiver 2 does not see the codeword intended for receiver 1 as interference. Similarly, the codeword for receiver 3 is chosen such that receiver 3 does not see the signals intended for receivers 1 and 2 (i.e $\mathbf{x_1} + \mathbf{x_2}$) as interference. This process continues for all K receivers. If $\mathbf{x}_i$ is chosen according to $N(0, \mathbf{\Sigma}_i)$ and User $\pi(1)$ is encoded first, followed by User $\pi(2)$, etc., the following is an achievable rate vector:

$$R_{\pi(i)} = \log \frac{\left|\mathbf{I}_N + \mathbf{H}_{\pi(i)}(\sum_{j\geq i} \mathbf{\Sigma}_{\pi(j)})\mathbf{H}_{\pi(i)}^H\right|}{\left|\mathbf{I}_N + \mathbf{H}_{\pi(i)}(\sum_{j>i} \mathbf{\Sigma}_{\pi(j)})\mathbf{H}_{\pi(i)}^H\right|} \quad i = 1, \dots, K. \tag{19}$$

Different rates can be achieved by varying the input covariance matrices and the encoding order. Notice that the same rates would be achieved if receiver $\pi(i)$ knew the transmitted signals $\mathbf{x}_{\pi(\mathbf{1})}, \dots, \mathbf{x}_{\pi(\mathbf{i-1})}$. The dirty paper region $\mathcal{C}_{\mathrm{DPC}}(P, \mathbf{H})$ is defined as the convex hull of the union of all such rate vectors over all positive semi-definite covariance matrices $\mathbf{\Sigma}_1, \dots, \mathbf{\Sigma}_K$ such that Tr$(\mathbf{\Sigma}_1 + \dots \mathbf{\Sigma}_K) =$ Tr$(\mathbf{\Sigma}_x) \leq P$

and over all permutations $(\pi(1), \ldots, \pi(K))$:

$$\mathcal{C}_{\text{DPC}}(P, \mathbf{H}) \triangleq Co\left(\bigcup_{\pi, \mathbf{\Sigma}_i} \mathbf{R}(\pi, \mathbf{\Sigma}_i)\right) \tag{20}$$

where $\mathbf{R}(\pi, \mathbf{\Sigma}_i)$ are the set of rates given by (19). The transmitted signal is $\mathbf{x} = \mathbf{x_1} + \ldots + \mathbf{x_K}$ and the input covariance matrices are of the form $\mathbf{\Sigma}_i = \mathbb{E}[\mathbf{x_i}\mathbf{x_i}^H]$. From the dirty paper result we find that $\mathbf{x}_1, \ldots, \mathbf{x}_K$ are uncorrelated, which implies $\mathbf{\Sigma}_x = \mathbf{\Sigma}_1 + \ldots + \mathbf{\Sigma}_K$.

One important feature to notice about the dirty paper rate equations in (19) is that the rate equations are neither a concave nor convex function of the covariance matrices, which makes finding the dirty paper region a difficult numerical problem.

C.3 MAC-BC Duality

A key component in establishing capacity results for the MIMO BC is the duality relationship between the MIMO BC and the MIMO MAC. The duality relationship refers to the fact that the dirty paper rate region of the multi-antenna BC with power constraint P is equal to the union of capacity regions of the dual MAC, where the union is taken over all individual power constraints that sum to P [123]:

$$\mathcal{C}_{\text{DPC}}(P, \mathbf{H}) = \bigcup_{\mathbf{P}: \sum_{i=1}^{K} P_i = P} \mathcal{C}_{MAC}(P_1, \ldots, P_K, \mathbf{H}^H). \tag{21}$$

This is the multiple-antenna extension of the previously established duality between the scalar Gaussian broadcast and multiple-access channels [64]. In addition to the relationship between the two rate regions, for any set of covariance matrices in the MAC/BC, [123] provides an explicit set of transformations to find covariance matrices in the BC/MAC that achieve the same rates. The union of MAC capacity regions in (21) is easily seen to be the same expression as in (14) but with the constraint $\sum_{i=1}^{K} \text{Tr}(\mathbf{Q}_i) \leq P$ instead of $\text{Tr}(\mathbf{Q}_i) \leq P_i \;\; \forall i$ (i.e. a sum constraint instead of individual constraints). In establishing this duality, it is shown that every rate vector in the dual MAC capacity region can be achieved in the MIMO BC. Each set of MAC covariance matrices corresponds to the polyhedron of rates described in (15). Each corner point of the polyhedron is achievable in the MAC using successive decoding with a specific order, and is also achievable in the MIMO BC using DPC with the *opposite* encoding order.

The MAC-BC duality is very useful from a numerical standpoint because the dirty paper region cannot be characterized in terms of a convex optimization problem, whereas the boundary of the dual MAC capacity region can be cast as a convex optimization problem. As result, the optimal MAC covariances can be found using standard convex optimization techniques and then transformed to the corresponding

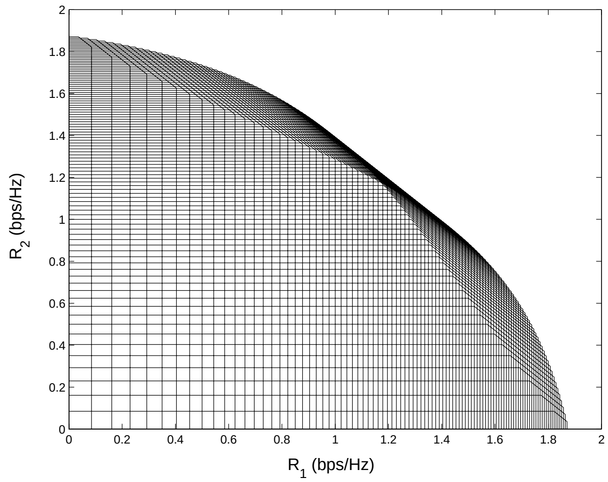

Fig. 13. Dirty paper rate region, $\mathbf{H}_1 = [1 \;\; 0.5]$, $\mathbf{H}_2 = [0.5 \;\; 1]$, $P = 10$.

optimal BC covariances using the MAC-BC transformations given in [123].

The dirty paper rate region is shown in Fig. 13 for a channel with 2 users, $M = 2$, and $N = 1$. Notice that the dirty paper rate region shown in Fig. 13 is actually a union of MAC regions, where each MAC region corresponds to a different set of individual power constraints. Since $N = 1$, each of the MAC regions is a pentagon, as discussed in Section IV-B.2. Similar to the MAC capacity region, the boundary of the dirty paper coding region is curved, except at the sum-rate maximizing portion of the boundary.

Duality also allows the MIMO MAC capacity region to be expressed as an intersection of the dual dirty paper BC rate regions [123, Corollary 1]:

$$\mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K, \mathbf{H}^H) = \bigcap_{\boldsymbol{\alpha} > 0} \mathcal{C}_{\mathrm{DPC}} \left(\sum_{i=1}^{K} \frac{P_i}{\alpha_i}; [\sqrt{\alpha_1} \mathbf{H}_1^T \cdots \sqrt{\alpha_K} \mathbf{H}_K^T]^T \right). \tag{22}$$

C.4 Constant Channel Capacity

In the previous section we described how the technique of dirty paper coding can be applied to the MIMO BC to pre-subtract interference at the transmitter. This strategy has in fact been shown to achieve the capacity region of the MIMO BC [131]. The optimality of dirty paper coding was first shown for sum rate capacity in [11] [123] [139] [126], and later extended to the full capacity region [131]. Since DPC achieves the capacity region (denoted as $\mathcal{C}_{BC}(P, \mathbf{H})$), the capacity region can be given in terms of (20), i.e. $\mathcal{C}_{BC}(P, \mathbf{H}) = \mathcal{C}_{DPC}(P, \mathbf{H})$. However, this form of the capacity region is difficult to find numerically because the rate equations are not concave functions of the input covariance matrices. By the MAC-BC duality explained in Section IV-C.3, a simpler expression of the capacity region of the MIMO BC can given in terms of the sum power MIMO MAC:

$$
\begin{aligned}
\mathcal{C}_{\mathrm{BC}}(P;\mathbf{H}) &= \mathcal{C}_{\mathrm{MAC}}(P;\mathbf{H}^{\mathbf{H}}) \\
&= \bigcup_{\{\mathbf{Q}_i \geq 0,\ \sum_{i=1}^{K} \mathrm{Tr}(\mathbf{Q}_i) \leq P\}} \left\{ \begin{array}{l} (R_1,\ldots,R_K): \\ \sum_{i\in S} R_i \leq \log\left|\mathbf{I}_M + \sum_{i\in S}\mathbf{H}_i^H \mathbf{Q}_i \mathbf{H}_i\right| \quad \forall S\subseteq\{1,\ldots,K\} \end{array} \right\}, \quad (23)
\end{aligned}
$$

where the $\mathbf{Q}_i$'s are the dual MAC covariance matrices. As stated earlier, a key property of this characterization is that the rate equations are concave functions of the input covariance matrices. Thus, power convex optimization algorithms can be used to numerically compute the MIMO BC capacity region.

The capacity region of the MIMO BC is a convex region, and thus its boundary can be found by maximizing the function $\mu_1 R_1 + \cdots + \mu_K R_K$ over all rate vectors in the capacity region and for all non-negative priorities $(\mu_1,\ldots,\mu_K)$ such that $\sum_{i=1}^{K}\mu_i = 1$. Since the MIMO BC capacity region is best described in terms of the MIMO MAC (with sum power constraint) capacity region, the same intuition holds regarding finding the boundary of the capacity region. In fact, the boundary of the MIMO BC capacity region can be found by solving the maximization in (16) subject to a sum power constraint on the covariance matrices. Standard convex optimization techniques can be used to solve this optimization, and a specific instance of such an algorithm is provided in [129]. As mentioned earlier, the encoding order with DPC is the opposite of the decoding order in the MAC. Thus, on the boundary of the capacity region, encoding should be done in order of *decreasing* priority, i.e. the user with the smallest priority should be encoded last and benefit from the most interference cancellation. For sum rate capacity (i.e. $\mu_1 = \cdots = \mu_K$), more efficient algorithms based on the KKT conditions exist [63] [136] [71]. Of particular interest is the sum power iterative waterfilling algorithm in [63] which is related to the iterative waterfilling technique for the MAC, with the difference that all the users simultaneously waterfill with respect to one power constraint.

Similar to the MIMO MAC, as SNR goes to infinity the sum rate capacity of the MIMO BC grows approximately as $\min(M,NK)\log(SNR)$. In other words, a MIMO BC with perfect CSIR and CSIT has a multiplexing gain of $\min(M,NK)$: the same as a point-to-point $M \times NK$ MIMO system. Additionally, if SNR is fixed but the number of transmit antennas and the number of receivers (with $N = 1$) are taken to infinity with $M \geq K$, then the sum capacity under the ZMSW model has the same growth constant as a $K \times M$ point-to-point MIMO channel (notice that there are K transmit antennas in the equivalent point-to-point channel, whereas K represents the receivers in the MIMO BC) with perfect CSIR and CDIT [43].

C.5 Fading Channels

For fading MIMO BC's, the capacity region depends crucially on the CSI available at the transmitter and receivers. As we discuss below, the degree of channel knowledge at the transmitter is particularly important.

With perfect CSIR and CSIT, the MIMO BC can be split into parallel channels with an overall power constraint [137] (see [76] for a treatment of the single-antenna channel). Clearly, the full multiplexing gain of $\min(M, NK)$ is achievable in this scenario.

The scenario where there is perfect CSIR but only CDIT is perhaps the most practical as well as interesting situation. However, the capacity and even the multiplexing gain in this scenario is in general unknown. Furthermore, the technique of dirty paper coding requires perfect CSIT in order for multi-user interference to be cancelled perfectly. Thus, it is still not clear what transmission strategy should be used for such fading channels.

One special case of CSIR and CDIT for which capacity is known is when all receivers have the same channel distribution (e.g. all ZMSW) and the same number of antennas. In this situation the K channels are *statistically* identical, i.e. $p(Y_1 = y|x) = \cdots = p(Y_K = y|x)$ for all x, y. [7] This implies that if any of the K receivers can decode a codeword, then receiver 1 can also decode the same codeword. This, in turn, implies that receiver 1 can decode every message sent by the transmitter. Therefore, the sum rate capacity of the MIMO BC is bounded by the capacity of the channel from the transmitter to any single receiver. The capacity region is therefore given by $C_{BC}(P) = \{(R_1, \ldots, R_K) : \sum_{i=1}^{K} R_i \leq C_1(P)\}$, where $C_1(P)$ is the point-to-point capacity from the transmitter to any receiver. If each of the channels follows the ZMSW model, the capacity is given by (11): $C_1(P) = \mathbb{E}_{\mathbf{H}}\left[\log\left|\mathbf{I}_N + \frac{P}{M}\mathbf{H_1}\mathbf{H_1^H}\right|\right]$. The rates achievable in the MIMO BC are therefore limited by the point-to-point capacity from transmitter to any of the receivers. In other words, there is no multi-user MIMO benefit in this scenario. Since $C_1(P)$ has a multiplexing gain of only $\min(M, N)$, the sum rate capacity also has a multiplexing gain of only $\min(M, N)$, as opposed to the $\min(M, NK)$ possible with perfect CSIT. Note that this is in contrast to the MIMO MAC, where CSIR and CDIT are sufficient in order to achieve the full multiplexing gain of $\min(M, NK)$. This idea is further investigated in [51], where it is shown that the multiplexing gain in the MIMO BC is lost whenever the channel fading distributions are spatially isotropic, i.e. the transmitter

[7] Notice that the statistical equivalence of the K receivers depends crucially on the CDIT, and the lack of CSIT. If there is CSIT, one must consider $p(y_i|x, H)$ instead of $p(y_i|x)$, and the statistical equivalence is lost, even if each of the channels follow the same model. Also note that with CDIT the statistical equivalence does not depend on the ZMSW model; the distributions of $\mathbf{H}_1, \ldots, \mathbf{H}_K$ need only be identical.

has no information regarding the spatial *direction* of each of the K channels.

If there is perfect CSIR and CDIT and the users do not have statistically identical channels, then very little is known about the capacity region. For example, consider a system under the CMI model with $K = 2$, $M = 2$, and $N = 1$. If $E[\mathbf{H}_1] \neq E[\mathbf{H_2}]$ (i.e. non-identical means), then the channels of the two users are not statistically identical. Therefore, the argument given above does not apply and it is not clear what the capacity region or sum rate capacity of this channel is. However, recent work has shown that the multiplexing gain of this channel is strictly smaller than that with perfect CSIR and CSIT [75].

The addition of limited feedback from receivers to transmitters can significantly increase the MIMO BC capacity region. In this setting, perfect CSIR and partial CSIT (achieved via a feedback channel) are considered. In [107], M random and orthogonal beamforming vectors are transmitted. The receivers measure the SINR on each of the beams and feed back the measurement values to the transmitter. The transmitter then transmits to the users with the highest SINR values. It is shown that this scheme achieves the full multiplexing gain when there are a large number of receivers experiencing fading according to the ZMSW model. Thus, this limited feedback is sufficient to overcome the barriers of the CSIR/CDIT model. The MIMO BC has also been considered in the finite rate feedback model, whereby each receiver feeds back a quantized version of its channel instantiation to the transmitter (see QCI model in Section III-A for a discussion of this model in the point-to-point setting). In this setting, if the number of feedback bits per user is increased in proportion to the system SNR, then the full multiplexing gain is achievable for any number of receivers [59].

C.6 Sub-Optimal Methods

Though DPC is capacity-achieving for the constant MIMO BC, its complexity is rather high because implementing DPC requires the use of near-optimal vector quantizers at both the transmitter and receivers [28]. Therefore, it is also of great interest to study the behavior of less complex, perhaps sub-optimal transmission schemes. In this section we examine the performance of two classes of schemes: orthogonal transmission (i.e. TDMA), and multi-user beamforming without interference cancellation.

TDMA (time-division multiple-access) has the inherent advantage of low complexity, because the transmitter transmits to only a single user at a time and the MIMO BC is essentially reduced to a point-to-point MIMO channel. The TDMA rate region is given by:

$$\mathcal{R}_{TDMA}(\mathbf{H}, P) \triangleq \left\{ (R_1, \ldots, R_K) : \sum_{i=1}^{K} \frac{R_i}{C(\mathbf{H}_i, P)} \leq 1 \right\}, \tag{24}$$

where $C(\mathbf{H}_i, P)$ denotes the single-user capacity of the i-th user subject to power constraint P.

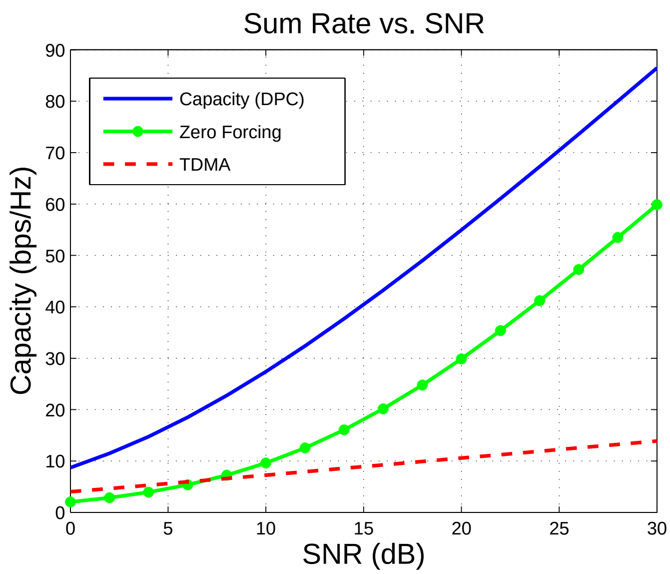

Fig. 14. Sum rate versus SNR for $M = 10$, $N = 1$, and $K = 10$ system

Note that other orthogonal allocations of resources such as FDMA are equivalent to TDMA from a capacity standpoint. The main drawback of TDMA is that the multiplexing gain is only $\min(M, N)$, as opposed to the $\min(M, NK)$ possible using DPC. As a result of this, TDMA can yield rates up to a factor of $\min(M, K)$ times smaller than DPC [61]. In fact, this gap is particularly pronounced in large systems with a small number of receive antennas operating at high SNR's.

One advantage of TDMA is that it can be easily used in fading channels with CSIR and CDIT. Point-to-point MIMO capacity is well known in this scenario and, as stated earlier, TDMA is in fact optimal in this scenario if the fading distributions of each of the users are identical. In general, TDMA is expected to perform well (i.e. close to capacity) even when channels are not statistically identical but when the CDIT is not very informative (e.g. CMI with very weak line-of-sight components).

Multi-user (or linear) beamforming is another sub-optimal transmission technique for the MIMO BC. With this scheme, a different beamforming direction(s) is chosen for each receiver and the sum of independent codewords is transmitted using the different beamforming directions. Since no interference cancellation is performed, multi-user interference is treated as noise. Note that this scheme differs from capacity-achieving DPC only in that no interference pre-subtraction is performed at the transmitter; the beamforming structure with DPC is in fact optimal when $N = 1$. Thus, the only difference from capacity-achieving DPC is that no multi-user interference is cancelled. Since beamforming is typically implemented through linear transmit and receive filters, beamforming is also referred to as linear processing.

For the single receive antenna channel ($N = 1$), a duality between the MAC and BC also exists for beamforming [6] [126], and the rates achievable using beamforming are most easily expressed in terms of the dual MAC as:

$$C_{BF}(\mathbf{H}_1, \ldots, \mathbf{H}_K, P) = \max_{\{P_i:\ \sum_{i=1}^K P_i \leq P\}} \sum_{j=1}^{K} \log \frac{\left|\mathbf{I}_M + \sum_{i=1}^K \mathbf{H}_i^H P_i \mathbf{H}_i\right|}{\left|\mathbf{I}_M + \sum_{i \neq j} \mathbf{H}_i^H P_i \mathbf{H}_i\right|}. \tag{25}$$

This maximization, however, cannot be rephrased as a convex optimization problem, and as a result is extremely difficult to compute numerically. Though it is difficult to find the optimal beamforming strategy, it is possible to compute sub-optimal strategies, and it is easy to see that beamforming does provide the full multiplexing gain of $\min(M, NK)$. Zero-forcing beamforming, in which the transmitter multiplies the vector of data symbols by the inverse of the channel to eliminate all multi-user interference, is one example of sub-optimal multi-user beamforming. Some asymptotic results regarding beamforming are given in [61, 128, 135]. A number of works have also studied the selection of beamforming vectors when receivers have multiple antennas (c.f. [111]). A novel technique combining beamforming and turbo coding has also been proposed [92] [93].

Beamforming can be used in fading channels with CSIR/CDIT, but it is difficult to choose good beamforming directions for each user unless the transmitter has very good estimates of the channels. Thus, beamforming is not expected to perform well unless the CDIT contains a high degree of information regarding the current channel (e.g. CMI with strong line-of-sight components).

In Fig. 14 sum rate is plotted versus SNR for a 10 transmit antenna, 10 receiver (each with a single antenna) system with perfect CSIR and CSIT. The sum rate capacity (achieved using DPC), zero-forcing beamforming, and TDMA rates are plotted. Notice that both the sum rate capacity and zero forcing curves achieve the full multiplexing gain, i.e. have a slope of $\min(M, K) = 10$ bits / 3 dB. Though zero-forcing achieves the correct slope, there is a substantial power penalty (approximately 8.3 dB) to using this sub-optimal method [58]. TDMA, on the other hand, only achieves a multiplexing gain of unity, corresponding to a slope of only 1 bit / 3 dB. Both DPC and zero-forcing provide a substantial increase over TDMA, even at moderate SNR values.

D. Open Problems in Multi-User MIMO

Multi-User MIMO has been a primary focus of research in recent years, mainly due to the large number of open problems in this area. Some of these are:

1. MIMO BC with perfect CSIR and CDIT: Capacity is only known when the channels of all users have

the same distribution. When this condition is not met, however, little is known regarding the capacity.

2. CDIT and CDIR: Since perfect CSI is rarely possible, a study of capacity with CDI at both the transmitter(s) and receiver(s) for both MAC and BC is of great practical relevance.

3. Non-DPC techniques for BC: Dirty paper coding is a very powerful capacity-achieving scheme, but it appears quite difficult to implement in practice. Thus, non-DPC multi-user transmissions schemes for the downlink (such as downlink beamforming [96]) are also of practical relevance. In addition, performing DPC (or some variant) with imperfect CSIT or CDIT is still challenging.

V. MULTICELL MIMO

The MAC and the BC are information theoretic abstractions of the uplink and the downlink of a single cell in a cellular system. However a cellular system consists of many cells with channels (timeslots, bandwidth, or codes) reused at spatially separated locations. Due to the fundamental nature of wireless propagation, transmissions in a cell are not limited to within that cell, and thus there is intercell interference between users and base stations that use the same channels. The majority of current systems are interference limited rather than noise limited. As a result, it is not sufficient to exclusively study single-cell models; multi-cell environments must be explicitly considered in order to accurately assess the benefits of MIMO technology.

In this chapter we provide an overview of information theoretic results for multi-cell environments. Analysis of the capacity of the cellular network explicitly taking into account the presence of multiple cells, multiple users, and multiple antennas along with the possibility of cooperation between base stations is inevitably a hard problem. Moreover, this setting can coincide with several long-standing unsolved problems in network information theory. Although difficult, analysis of multicell MIMO capacity is also of the utmost importance, since it defines a benchmark that can be used to gauge the efficiency of practical schemes. Work on capacity for multi-cell environments can be grouped in two broad categories, with one group assuming that base stations cannot cooperate (as is done in current systems), and with more recent work considering multi-cellular environments where base stations are allowed some level of cooperation (i.e. cooperative transmission and/or reception). We will consider capacity results for both cases, and also discuss some system level issues associated with multicell systems with multiple antennas, which give rise to additional open problems.

A. Multicell MIMO without Base Station Cooperation

The traditional analysis of capacity of cellular systems has assumed that neither base stations nor users can cooperate. Note that current cellular systems do cooperate in some sense (e.g. handoff), but mostly in terms of networking rather than in transmission or reception of data. In this chapter we use "cooperation" to imply cooperative transmission and reception, i.e. cooperation at the physical layer. If no cooperation is allowed, the channel becomes an interference channel, in which there are multiple transmitters and multiple receivers communicating over a common medium, but each transmitter only wishes to communicate with a single receiver. Unfortunately, Shannon capacity of channels with interference is a long-standing open problem in information theory [23, 118]; in fact, even the capacity region of a two transmitter, two receiver interference channel with no fading and single antenna elements is not fully known [20]. Some results for the multiple antenna interference channel are given in Section VI on MIMO ad-hoc networks.

A more promising and well studied approach is to treat all out-of-cell interference as an additional source of Gaussian noise. The Gaussian assumption can be viewed as a worst-case assumption about the interference, since exploiting known structure of the interference can help in decoding the desired signals. By treating the interference as Gaussian noise, the capacity of both the uplink and downlink can be determined using the single-cell analysis of Section IV. The capacity of a single antenna cellular system uplink with fading that treats interference as Gaussian noise was obtained in [102] for both one and two dimensional cellular grids. These capacity results show that with or without fading, when intercell interference is nonnegligible, an orthogonal multiple access method (e.g. TDMA) within a cell is optimal. This is also the case when channel-inversion power control is used within a cell. Moreover, in some cases partial or full orthogonalization of channels assigned to different cells can increase capacity. There has also been a body of work that has studied multiple transmit and receive antenna arrays in cellular systems [133] [14] [15] [80]. In this work, the spatial structure of out-of-cell interference is considered. Note that AWGN is normally considered to be spatially white. However, out-of-cell interference generally has some statistical structure, i.e. a spatial covariance, which significantly affects capacity results. One interesting conclusion that can be drawn from this work is that it is sometimes beneficial to limit the number of transmit antennas, because the receiver must simultaneously use its antennas to cancel out-of-cell interference and decode the desired signal. When performing these operations, it is beneficial to have the number of receive antennas be as large as the aggregate number of transmit antennas, i.e. including the transmit antennas of strong interfering users.

B. Multicell MIMO with Base Station Cooperation

A relatively new area of research has studied cellular systems where cooperation between base stations is allowed. Since base stations are wired and fixed in location, it may be practical to allow base stations to cooperatively transmit and receive. If perfect cooperation is assumed, this allows the entire network to be viewed as a single cell with a distributed antenna array at the base station. As a result, capacity results about the single cell uplink and downlink channel, discussed in Section IV, can be applied. Note, however, that the fact that the base stations and terminals are geographically separated directly impacts the channel gains in the composite network.

The uplink capacity of cellular systems (where mobiles and base stations are assumed to each have single antennas) under the assumption of full base station cooperation, where signals received by all base stations are jointly decoded, was first investigated in [40] followed by a more comprehensive treatment in [134]. In both cases propagation between the mobiles and the base stations is characterized using an AWGN channel model (i.e. no fading is considered) with a channel gain of unity within a cell, and a gain of α, $0 \leq \alpha \leq 1$, between cells. The Wyner model of [134] considers both one and two dimensional arrays of cells, and derives the per-user capacity in both cases. It is also shown in both [134] and [40] that uplink capacity is achieved by using orthogonal multiple access techniques (e.g. TDMA) within each cell, and reusing these orthogonal channels in other cells, although this is not necessarily uniquely optimal.

The downlink channel can be modeled as a MIMO broadcast channel if perfect base station cooperation is assumed. On the downlink, since the base stations can cooperate perfectly, dirty paper coding can be used over the entire transmitted signal (i.e. across base stations) in a straightforward manner. The application of dirty paper coding to a multiple cell environment with cooperation between base stations was pioneered in recent work by Shamai and Zaidel [103]. For one antenna at each user and each base station, they show that a relatively simple application of dirty paper coding can enhance the capacity of the cellular downlink. While capacity computations are not the focus of [103], they do show that their scheme is asymptotically optimal at high SNRs. A number of other works have also studied the capacity of multi-cell downlink channels, in both the finite and asymptotic regimes [53] [1] [48]. The duality of the MAC and BC (discussed in Section IV-C.3) can also be applied to the multi-cell composite channels to relate uplink and downlink results.

One weakness of treating the multi-cell downlink channel as a standard MIMO broadcast channel is that the average transmit power across all antennas (and thus across all base stations) is bounded. In practice, the power from each base station may in fact be bounded. If these stricter constraints are enforced, it has been shown that dirty paper coding still achieves the sum capacity [71] [139]. A modified duality between the MAC and BC has also been established when such power constraints are imposed [71]. For large networks, the effect of imposing per-base power constraints is still unclear [53].

In general, results show that base station cooperation can yield significant capacity increases relative to systems without cooperation. From a research perspective, the size of large networks makes generating numerical results difficult, and as a result asymptotic results must be relied upon [1] [48]. From a practical perspective, actually achieving perfect cooperation between physically remote base stations is still very challenging. In addition, it seems that very large-scale cooperation would require tremendous complexity. As a result of these issues, it is not clear what methods of base station cooperative are practically feasible.

C. System Level Issues

The capacity results described in this section address just a few out of many interesting questions in the design of a cellular system with multiple antennas. Multiple antennas can be used not only to enhance the capacity of the system but also to drive down the probability of error through diversity combining. Recent work by Zheng and Tse [144] unravels a fundamental diversity versus multiplexing tradeoff in MIMO systems. Also, instead of using isotropic transmit antennas on the downlink and transmitting to many users, it may be simpler to use directional antennas to divide the cell into sectors and transmit to one user within each sector. The relative impact of CDIT and/or CDIR on each of these schemes is not fully understood. Although in this paper we focus on the physical layer, smart schemes to handle CDIT can also be found at higher layers. An interesting example is the idea of opportunistic beamforming [125]. In the absence of CSIT, the transmitter randomly chooses the beamforming weights. With enough users in the system, it becomes very likely that these weights will be nearly optimal for one of the users. In other words, a random beam selected by the transmitter is very likely to be pointed towards a user if there are enough users in the system. Instead of feeding back the channel coefficients to the transmitter the users simply feed back the SNRs they see with the current choice of beamforming weights. This significantly reduces the amount of feedback required. By randomly changing the weights frequently, the scheme also treats all users fairly.

VI. MIMO FOR AD-HOC NETWORKS

An ad hoc wireless network is a collection of wireless mobile nodes that self-configure to form a network without the aid of any established infrastructure, as shown in Figure 15. Without an inherent infrastructure, the mobiles handle the necessary control and networking tasks by themselves, generally through the use of distributed control algorithms. Multihop routing, whereby intermediate nodes relay packets towards their final destination, is often used since it can improve the throughput and power efficiency of the network. Note that with sufficient transmit power any node in the network can transmit a signal directly to any other node. However, such transmissions over long distances will result in a low received power, and will also cause interference to other links. Thus, links with low signal-to-interference-plus-noise (SINR) power ratios are typically not used. The SINR on different links is illustrated by the different line widths in Figure 15.

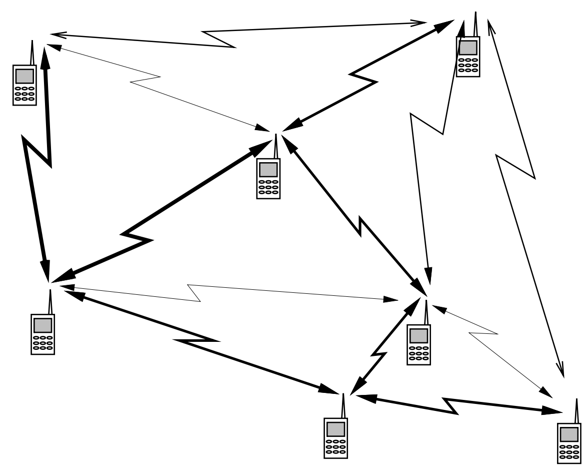

Fig. 15. Ad Hoc Network.

The fundamental capacity limits of an ad hoc wireless network - the set of maximum data rates possible between all nodes - is a highly challenging problem in information theory even when the nodes only have a single antenna. For a network of K nodes, each node can communicate with $K-1$ other nodes, so the capacity region has dimension $K(K-1)$. While rate sums across any cutset of the network are bounded by the corresponding mutual information expressions [23, Theorem 14.10.1], simplifying this formula into a tractable expression for the ad hoc network capacity region is an immensely complex problem. Given the lack of capacity results for ad hoc networks with just one antenna per node, it seems that considering multiple antennas per node will only make the problem more intractable. Moreover, as with cellular systems, multiple antennas in an ad hoc network can be used for diversity or sectorization in addition to capacity gain, and the fundamental tradeoffs between these different uses are very difficult to characterize.

One approach to get around this difficulty is to study the asymptotic behavior of the total network throughput as the number of nodes increases to infinity. Such an approach has been pioneered by Gupta and Kumar [38]. Various decentralized transmitter-receiver models have been considered in this context, including static, mobile, and fading channels with or without successive interference cancellation at the receiver. It has been found that in most static channel cases the capacity per node decays as $O(\frac{1}{\sqrt{n}})$, thus going down to zero as n grows to infinity. However, mobility and or fading is found to enhance this performance by means of exploiting opportunism, and a $O(1)$ rate per node can now be obtained [36] [26] [32]. The advantage of this analysis technique is the intuition it provides about capacity trends, with the inter-user tradeoffs being lost at this level of abstraction.

Another approach taken by researchers has been to consider networks with a few nodes that form the basic building blocks of larger ad hoc networks. The hope is that capacity analysis may be more tractable for such simple networks. The primary few-node components of ad hoc networks are the MAC, BC, relay and interference channels. The MAC and BC have already been analyzed in detail earlier in this chapter, and so this section devotes itself to the discussion of the relay and interference channels.

A. The relay channel

The relay channel, where one node aids in the communication of another node's message, is a natural and important building block of ad-hoc networks. The most elementary relay channel is modeled as a three-node system, where the transmitter communicates with the relay and the destination nodes while the relay communicates with just the destination node. The AWGN relay channel, shown in Fig. 16, is defined mathematically as

$$y = h_{sd}x + h_{rd}x_1 + n_1 \tag{26}$$

$$y_1 = h_{sr}x + n_2, \tag{27}$$

where x is the signal transmitted by the source, x_1 is the signal transmitted by the relay (which is a function of the relay's previous inputs), y is the received signal at the destination, and y_1 is the received signal at the relay. The capacity of this channel is still an open problem, with the capacity known for some special cases. One such special case is the physically degraded relay channel [22], where the destination signal is a physically degraded version of the received signal at the relay. In general, only upper and lower bounds to capacity are known. The best upper bound known is the well known cut-set upper bound, while the lower bounds are obtained using a node-cooperation strategy known as block-Markov coding [22]. In block-Markov coding, the relay uses the information received by it in the *past* to correlate its transmit signal to the transmitter's *current* signal, and the net coherent combining provides capacity enhancements over that of the point-to-point channel. This lower bound makes the assumption, however, that the relay

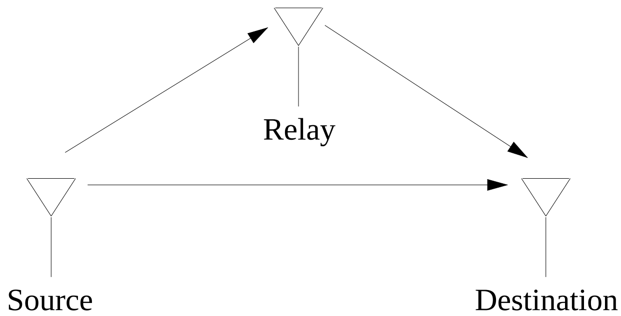

Fig. 16. The relay channel

completely decodes its received signal, and is thus called the decode-and-forward policy. There are many other techniques that the relay could use: most work to date considers relay channel transmission schemes in the following three categories: [69]:

1. *Decode-and-Forward*: In the decode and forward approach the relay decodes part or all of the codeword transmitted by the source during a block. Clearly, this transmission should be at a rate higher than the capacity of the direct link from source to destination. During the next block, the relay then transmits to the destination a re-encoded version of the message decoded during the previous block to assist the destination's decoding process. Since the source is aware of the relay's strategy and of the codeword it sent the previous block, the source and relay can *cooperate*, e.g. the source can transmit a signal identical to the relay's transmission (coherent combination). Note that if the relay and source cooperate, the source's transmission in each block consists of two parts: information regarding the previous block's message (to facilitate cooperation), and a new codeword. This class of strategies was first proposed in [22].

2. *Compress-and-Forward*: In compress and forward, the relay transmits a compressed, or quantized, version of its received signal, instead of decoding the received signal and re-encoding. This technique is related to source coding with side information.

3. *Amplify-and-Forward*: In this strategy, the relay acts as a simple linear repeater, amplifying its received signal instantaneously or during the next symbol or block. This clearly leads to some noise amplification, but even this simple strategy can yield significant performance improvements.

One of the best lower bounds known for the relay channel, given by [22, Thm. 7], uses a mix of these strategies. Unfortunately, none of the lower bounds prove to be tight even for the AWGN single-antenna relay channel, which is thus an open problem. However, the analysis techniques used for this simple relay channel have been extended to study more complex ad hoc networks using cooperative schemes, as described in Section VI-C.

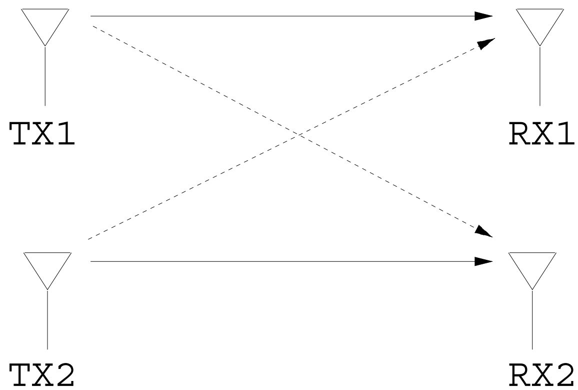

Fig. 17. The interference channel

B. The interference channel

In this section we summarize what is known about the next ad-hoc network building block - the interference channel (IFC). This is a system with two independent transmitters sending information to two receivers (Figure 17), where each transmitter has a message for a distinct receiver. The signals from the two transmitters interfere with one other and the resulting mixture is received at each receiver. For example, in the Gaussian IFC case, a weighted sum of the two transmit signals is assumed to arrive at each receiver in the presence of additive Gaussian noise. In the general discrete memoryless case, the interaction between the two signals is expressed in terms of the conditional p.d.f. $p(y_1, y_2|x_1, x_2)$. Although the capacity region of both the general case and the particular case of the Gaussian interference channel are still open problems, significant progress has been made in finding the capacity region of certain classes of these channels [21, 117].

A category of interference channels where the interference is “much stronger” than the signal, called the very-strong IFC case, was amongst the first for which capacity was determined [13]. In this case, the interference signal at each receiver is assumed to be so strong that it can be decoded and subtracted out first before decoding the intended signal. The broadest class of IFCs for which the capacity is known is the strong interference channel case [21, 97], which includes the very-strong case as a sub-class. In the strong IFC case, the interference is assumed strong enough so that it can be decoded (not necessarily first) at each receiver.

Research on the capacity limits of MIMO IFCs is still in its early stages. Interestingly, the simple strong interference result given in [21,97] for the SISO case does not generalize to the MIMO case, and a straightforward characterization of what strong interference means for MIMO IFCs is yet to be obtained.

In recent work [121] obtained strong interference conditions for SIMO interference channels, where the transmitters possess single antennas. Finding explicit strong interference conditions for the MISO and more general MIMO IFCs is still an open problem.

A special feature of MIMO IFCs is the spatial degree of freedom that is absent in the SISO case, making the channel capacity problem much more difficult and challenging. Known results from SISO IFC analysis, such as achievable regions and outer bounds, require considerable reworking before they can be replicated in the MIMO domain, and in many cases these results do not generalize to the MIMO case at all. Indeed, the capacity of MIMO IFCs is one of the areas in MIMO multiuser capacity analysis where very few results are known.

C. Cooperative Communication

In this section we summarize the work on cooperative communication, i.e. using MIMO techniques in ad-hoc networks. In addition to using such techniques when a node in the network possesses multiple antennas, clusters of nodes that are located close together can exchange information to create a *virtual* antenna array, leading to a distributed MIMO system [88]. In other words, nodes close together on the transmit side can exchange information to form a multiple-antenna transmitter, and nodes near each other on the receiver side can exchange information to form a multiple antenna receiver, as shown in Figure 18. Since each node has a different channel to each receiver, this cooperative MIMO system has performance advantages in terms of multiplexing and diversity. In addition, the multiple antennas can be used on the transmit or receive side to steer the beam in the direction of the intended receiver, thereby reducing interference and multipath.

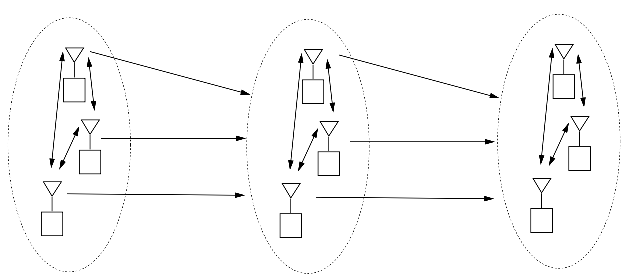

Fig. 18. Cooperative MIMO

In this section we describe extensions of the basic relay channel to the case of cooperative (i.e. MIMO) communication over ad hoc networks. We consider two different settings under the ZMSW model: perfect CSIR and CSIT, and perfect CSIR and CDIT.

C.1 Perfect CSIR and CSIT

When the channels between the sources, relays and destinations are fading, but are perfectly known, the different fading realizations can be treated as constant channels, thus reducing the problem to the analysis of non-fading relay channels. The capacity of MIMO relays with no fading is still an open problem, with early work in this area in [130]. In this paper, the authors derive upper and lower bounds on the capacity of these channels. The upper bounds are derived based on cutset bounds [23], while the lower bounds are based on path-diversity. The authors use convex optimization techniques to generate specialized algorithms to compute the lower and upper bounds.

Relaying has also been considered in a multiple-transmitter, multiple-receiver scenario (i.e. similar to an interference channel), where transmitting nodes help other nodes by sending their own messages as well as cooperating with other nodes. In addition, the receiving nodes have the ability to cooperate with each other through techniques such as amplify-and-forward. Such scenarios are beginning to be analyzed in greater detail [46] [62] [91].

C.2 Perfect CSIR and CDIT

Most work on cooperative MIMO has concentrated on the perfect CSIR and CDIT setting under the ZMSW model. Cooperation of this nature has been studied in both cellular [98] [99] [100] and ad-hoc [73] [72] settings. This is perhaps the most practical scenario for cooperation in ad-hoc or cellular networks, as obtaining full CSI at each of the transmitters would require feedback from every receiver to every transmitter.

In a cellular uplink setting, multiple transmitting nodes (i.e. mobiles) wish to communicate to a single base station. In traditional networks, mobiles transmit directly to the base station over a common channel, perhaps separated by a multi-user technique such as CDMA. Notice, however, that each of the mobiles can hear the transmission of other mobiles. Therefore, a mobile can relay the messages of other mobiles to the base station, in addition to transmitting its own message. The key advantage of doing so is the extra form of diversity obtained, referred to as *user cooperation diversity*, via the channels of other mobile devices. Even if a mobile is severely faded, there is a high probability that one of its neighboring mobiles will not be faded and thus will be able to relay the message to the base station. Therefore, a cooperation scheme such as block-Markov coding [99] increases rates relative to no cooperation. In practice, these ideas translate to combining CDMA with cooperative coding to decrease probability of error.

Under a somewhat different problem formulation based on outage capacity, [73] [72] find cooperative communication schemes for half-duplex relays, i.e., for relays that cannot transmit and receive at the same time. These cooperative schemes achieve full diversity - the diversity order is equal to the total number of transmitting nodes - but at some rate penalty. Similar to the point-to-point MIMO setting, recent work has formalized the connection between the diversity and multiplexing (i.e. rate) gain that cooperative MIMO can provide [3] [94].

Though most work on cooperative communication has concentrated on the outage formulation, there has also been some work on ergodic capacity of cooperative channels. In particular, recent work [69] has established the ergodic capacity of a class of cooperative MIMO channels. Their results prove that the decode-and-forward strategy is capacity-achieving under the ZMSW model with perfect CSIR and CDIT when the relay is sufficiently close to the source, i.e. the average SNR from the source to relay is much higher than the SNR from source to destination or relay to destination. In fact, the capacity is equal to the capacity of the channel when the source and relay nodes are assumed to fully cooperate, i.e. act as a transmit antenna array. Note that this result is surprising because the source and relay are in fact not co-located, and can only cooperate over a noisy channel.

In this situation, the lack of CSIT, and more specifically the lack of phase information, makes it optimal for the relay to decode the entire message from the source, and then re-encode it and transmit it independently of the source transmission. In other words, the source transmits Gaussian codewords at rate R, which is equal to the rate assuming full cooperation between the source and relay. The relay decodes the codeword transmitted by the source (the condition on the relay being close to the source implies that the relay-source capacity is larger enough to allow this). In the next block, the relay re-encodes the message decoded in the previous block, while the source transmits an entirely new message. The destination uses the information from the current and previous block to then decode the message sent by the source in the previous block, and so forth. This result can also be generalized to a channel where there are multiple relays.

C.3 Diversity-Multiplexing Tradeoffs

Recently, researchers have investigated the diversity and multiplexing gains in distributed MIMO (a.k.a virtual MIMO) systems [47, 49, 143]. Specifically, this work investigates whether distributed MIMO obtained by wireless relaying can mimic a multiple antenna system in terms of its diversity-multiplexing trade-off. It is easily seen by applying the min-cut max-flow bound that the spatial multiplexing gain of a distributed MIMO system is limited by the number of antennas at the initial source and the final destination, regardless of the number of intermediate relays [49, 143]. Thus, for example, if the source or the destination has only one antenna, the multiplexing gain is limited to one. This is in contrast with the fact that full diversity gain can be obtained through relays in such systems [72]. The nature of the diversity and multiplexing tradeoff for distributed MIMO systems is quite different from that in true MIMO systems [47]. It was shown in [47] that for an interference channel with a single antenna per node the multiplexing gain is limited to one even with transmitter and/or receiver cooperation. However, multiplexing gains higher than one are possible in a distributed MIMO system with multiple antennas at the nodes.

VII. SUMMARY

We have summarized recent results on the capacity of MIMO channels for both single-user and multi-user systems. The great capacity gains predicted for such systems can be realized in some cases, but realistic assumptions about channel knowledge and the underlying channel model can significantly mitigate these gains. For single user systems the capacity under perfect CSI at the transmitter and receiver is relatively straightforward and predicts that capacity grows linearly with the number of antennas. Backing off from the perfect CSI assumption makes the capacity calculation much more difficult, and the capacity gains are highly dependent on the nature of the CSI/CDI, the channel SNR, and the antenna element correlations. Specifically, assuming perfect CSIR, CSIT provides significant capacity gain at low SNRs but not much at high SNRs. The insight here is that at low SNRs it is important to put power into the appropriate eigenmodes of the system. Interestingly, with perfect CSIR and CSIT, antenna correlations are found to increase capacity at low SNR's and decrease capacity at high SNR's. Finally, under CDIT and CDIR for a zero-mean spatially white channel, at high SNRs capacity grows relative to only the double log of the SNR with the number of antennas as a constant additive term. This rather poor capacity gain would not typically justify adding more antennas. However, at lower SNRs the growth relative to the number of antennas is less pessimistic.

We also examined the capacity of MIMO broadcast and multiple access channels. The capacity region of the MIMO MAC is well-known and can be characterized as a convex optimization problem. The MAC-BC duality greatly simplifies capacity region calculations for the MIMO BC that could otherwise lead to non-convex optimization problems. These capacity and achievable regions are only known for ergodic capacity under perfect CSIT and CSIR. Relatively little is known about the MIMO MAC and BC regions under more realistic CSI assumptions. A multicell system with base station cooperation can be modeled as a MIMO BC (downlink) or MIMO MAC (uplink) where the antennas associated with each base station are pooled by the system. Exploiting this antenna structure leads to significant capacity gains over transmission strategies that send to just one user at a time, typically the user with the best channel. We also describe the rather limited results on capacity of ad-hoc networks with multiple antennas, as well as the capacity gain associated with node cooperation to form virtual MIMO channels from single-antenna nodes.

There are many open problems in this area. For single-user systems the problems are mainly associated with CDI only at either the transmitter or receiver. Most capacity regions associated with multi-user MIMO channels remain unsolved, especially ergodic capacity and capacity versus outage for the MIMO BC under perfect receiver CSI only. There are very few existing results for CDI at either the transmitter or receiver for any multi-user MIMO channel. The capacity of cellular systems with multiple antennas remains a relatively open area, in part because the single-cell problem is mostly unsolved, and in part because the Shannon capacity of a cellular system is not well-defined and depends heavily on frequency reuse assumptions and propagation models. Other fundamental tradeoffs in MIMO cellular designs, such as whether antennas should be used for sectorization, capacity gain, or diversity, are not well understood. Similar tradeoffs exist in ad hoc networks, where the capacity associated with multiple antennas is a wide open problem. In short, we have only scratched the surface in understanding the fundamental capacity limits of systems with multiple transmitter and receiver antennas as well as the implications of these limits for practical system designs. This area of research is likely to remain timely, important, and fruitful for many years to come.

REFERENCES

[1] D. Aftas, M. Bacha, J. Evans, and S. Hanly, "On the sum capacity of multiuser MIMO channels," in *Proceedings of Intern. Symposium on Inform. Theory and its Applications*, Oct. 2004.

[2] R. Ahlswede, "Multi-way communication channels," in *Proceedings 2nd Int. Symp. Information Theory*, 1973, pp. 23–52.

[3] K. Azarian, H. E. Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," submitted to IEEE Trans. Inform. Theory, 2004.

[4] E. Biglieri, J. Proakis, and S. S. (Shitz), "Fading channels: Information theoretic and communication aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[5] D. Blackwell, L. Breiman, and A. J. Thomasian, "The capacity of a class of channels," *Ann. Math. Stat.*, vol. 30, pp. 1229–1241, Dec. 1959.

[6] H. Boche and M. Schubert, "A general duality theory for uplink and downlink beamforming," in *Proceedings of IEEE Vehicular Tech. Conf.*, 2002.

[7] H. Bolcskei, D. Gesbert, and A. J. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, no. 2, pp. 225–234, Feb 2002.

[8] S. Borst and P. Whiting, “The use of diversity antennas in high-speed wireless systems: Capacity gains, fairness issues, multi-user scheduling,” *Bell Labs. Tech. Mem.*, 2001, download available at http://mars.bell-labs.com.

[9] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2003.

[10] G. Caire and 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.

[11] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna Gaussian broadcast channel,” *IEEE Trans. Inform. Theory*, vol. 49, no. 7, pp. 1691–1706, July 2003.

[12] G. Caire, G. Taricco, and E. Biglieri, “Optimum power control over fading channels,” *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1468–1489, July 1999.

[13] A. B. Carliel, “A case where interference does not reduce capacity,” *IEEE Trans. Inform. Theory*, pp. 569–570, Sept. 1975.

[14] S. Catreux, P. Driessen, and L. Greenstein, “Simulation results for an interference-limited multiple-input multiple-output cellular system,” *IEEE Commun. Letters*, vol. 4, no. 11, pp. 334–336, Nov. 2000.

[15] S. Catreux, P. Driessen, and L. Greenstein, “Attainable throughput of an interference-limited multiple-input multiple-output (MIMO) cellular system,” *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1307–1311, Aug. 2001.

[16] M. Chiani, M. Z. Win, and Z. Zanella, “On the capacity of spatially correlated MIMO Rayleigh-fading channesl,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2363–2371, Oct. 2003.

[17] D. Chizhik, J. Ling, P. Wolniansky, R. Valenzuela, N. Costa, and K. Huber, “Multiple input multiple output measurements and modeling in Manhattan,” in *Proceedings of IEEE Vehicular Tech. Conf.*, 2002.

[18] C. Chuah, D. Tse, J. Kahn, and R. Valenzuela, “Capacity scaling in MIMO wireless systems under correlated fading,” *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 637–650, March 2002.

[19] M. Costa, “Writing on dirty paper,” *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[20] M. Costa and A. El Gamal, “The capacity region of the discrete memoryless interference channel with strong interference,” *IEEE Trans. Inform. Theory*, vol. 33, pp. 710–711, Sep. 1987.

[21] M. H. M. Costa and A. A. E. Gamal, “The capacity region of the discrete memoryless interference channel with strong interference (corresp.),” *IEEE Trans. Inform. Theory*, pp. 710–711, Sept. 1987.

[22] T. Cover and A. E. Gamal, “Capacity theorems for the relay channel,” *IEEE Trans. Inform. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[24] I. Csiszar, “Arbitrarily varying channels with general alphabets and states,” *IEEE Trans. Inform. Theory*, vol. 38, no. 6, pp. 1725–1742, Nov. 1992.

[25] I. Csiszar and J. Korner, *Information theory: Coding Theorems for Discrete Memoryless Systems*. Academic Press, 1997.

[26] S. Diggavi, M. Grossglauser, and D. Tse, “Even one-dimensional mobility increases ad hoc wireless capacity,” in *Proceedings of Int. Symp. Inform. Theory*, June 2002.

[27] U. Erez, S. Shamai, and R. Zamir, “Capacity and lattice strategies for cancelling known interference,” in *Proceedings of Int. Symp. Inform. Theory and its Applications*, Nov. 2000, pp. 681–684.

[28] U. Erez and S. ten Brink, "Approaching the dirty paper limit for cancelling known interference," in *Proceedings 41st Annual Allerton Conference on Commun., Control and Computing*, 2003.

[29] G. J. Foschini, "Layered space-time architecture for wireless communication in fading environments when using multi-element antennas," *Bell Labs Techn. J.*, pp. 41–59, Autumn 1996.

[30] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun. : Kluwer Academic Press*, no. 6, pp. 311–335, 1998.

[31] R. G. Gallager, "An inequality on the capacity region of multiaccess fading channels," *Communication and Cryptography - Two Sides of one Tapestry*, boston, MA: Kluwer, 1994, pp. 129-139.

[32] A. E. Gamal, J. Mammen, B. Prabhakar, and D. Shah, "Throughput-delay trade-off in energy constrained wireless networks," in *Proceedings of Int. Symp. Inform. Theory*, July 2004.

[33] M. J. Gans, N. Amitay, Y. S. Yeh, H. Xu, T. Damen, R. A. Valenzuela, T. Sizer, R. Storz, D. Taylor, W. M. MacDonald, C. Tran, and A. Adamiecki, "Outdoor BLAST measurement system at 2.44 GHz: Calibration and initial results," *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 570–581, April 2002.

[34] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1986–1992, Nov. 1997.

[35] A. J. Goldsmith, *Wireless Communications*. Cambridge, 2005.

[36] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

[37] T. Guess and M. K. Varanasi, "Multiuser decision-feedback receivers for the general Gaussian multiple-access channel," in *Proceedings of Allerton Conf. on Commun., Control, and Comput.*, Oct 1996.

[38] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[39] S. Hanly and D. Tse, "Multiaccess fading channels–Part II:Delay-limited capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2816–2831, November 1998.

[40] S. V. Hanly and P. Whiting, "Information theory and the design of multi-receiver networks," in *IEEE 2nd Intl. Symp. Spread Spec. Tech. Apps. (ISSTA)*, Nov. 1992, pp. 103–106.

[41] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[42] B. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel-hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept. 2004.

[43] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in sum rate," in *Proceedings of 40th Annual Allerton Conference on Commun., Control and Computing*, Oct. 2002.

[44] D. Hoesli, Y.-H. Kim, and A. Lapidoth, "Monotonicity results for coherent MIMO channels," 2004, submitted to *IEEE Trans. Inform. Theory*.

[45] D. Hoesli and A. Lapidoth, "The capacity of a MIMO Ricean channel is monotonic in the singular values of the mean," in *Proceedings of the 5th International ITG Conference on source and Channel Coding (SCC), Erlangen, Nuremberg*, January 2004.

[46] A. Host-Madsen, "On the achievable rate for receiver cooperation in ad-hoc networks," in *Proceedings of IEEE Int. Symp.*

Inform. Theory, June 2004, p. 272.

[47] A. Host-Madsen and Z. Yang, “Interference and cooperation in multi-source wireless networks,” in *IEEE Communication Theory Workshop*, May 2005.

[48] H. Huang and S. Venkatesan, “Asymptotic downlink capacity of coordinated cellular networks,” in *Proceedings of Asilomar Conf. on Signals, Systems, & Comp.*, Nov. 2004, pp. 850–855.

[49] S. Jafar, “Degrees of freedom in distributed MIMO communications,” in *IEEE Communication Theory Workshop*, May 2005.

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

[51] 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, March 2005.

[52] S. Jafar and A. Goldsmith, “Multiple-antenna capacity in correlated Rayleigh fading with channel covariance information,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 990–997, May 2005.

[53] S. A. Jafar and A. Goldsmith, “Transmitter optimization for multiple antenna cellular systems,” in *Proceedings of Int. Symp. Inform. Theory*, June 2002, p. 50.

[54] S. A. Jafar and A. J. Goldsmith, “On optimality of beamforming for multiple antenna systems with imperfect feedback,” in *Proceedings of Int. Symp. Inform. Theory*, June 2001, p. 321.

[55] S. A. Jafar and A. J. Goldsmith, “Vector MAC capacity region with covariance feedback,” in *Proceedings of Int. Symp. Inform. Theory*, June 2001, p. 321.

[56] S. A. Jafar, S. Vishwanath, and A. J. Goldsmith, “Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback,” in *Proceedings of Int. Conf. Commun.*, vol. 7, 2001, pp. 2266–2270.

[57] S. Jayaweera and H. V. Poor, “Capacity of multiple-antenna systems with both receiver and transmitter channel state information,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2697–2709, Oct. 2003.

[58] N. Jindal, “High SNR analysis of MIMO broadcast channels,” in *Proceedings of Int. Symp. Inform. Theory*, Sept. 2005.

[59] N. Jindal, “MIMO broadcast channels with finite rate feedback,” in *Proceedings of IEEE Globecom*, Nov. 2005.

[60] N. Jindal and A. Goldsmith, “Optimal power allocation for parallel broadcast channels with independent and common information,” in *Proceedings of Int. Symp. Inform. Theory*, July 2004.

[61] N. Jindal and A. Goldsmith, “Dirty paper coding vs. TDMA for MIMO broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

[62] N. Jindal, U. Mitra, and A. Goldsmith, “Capacity of ad-hoc networks with node cooperation,” in *Proceedings of Int. Symp. Inform. Theory*, July 2004.

[63] N. Jindal, W. Rhee, S. Vishwanath, S. Jafar, and A. Goldsmith, “Sum power iterative water-filling for multi-antenna Gaussian broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1570–1580, April 2005.

[64] N. Jindal, S. Vishwanath, and A. Goldsmith, “On the duality of Gaussian multiple-access and broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 5, pp. 768–783, May 2004.

[65] E. Jorswieck and H. Boche, “Optimal transmission with imperfect channel state information at the transmit antenna array,” submitted to Wireless Personal Communications.

[66] E. Jorswieck and H. Boche, “Optimal power allocation and complete characterization of the impactof correlation on the

capacity of miso systems with different CSI at the transmitter," in *Proceedings of Int. Symp. Inform. Theory*, June 2003, p. 353.

[67] 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.

[68] A. Khisti, U. Erez, and G. Wornell, "A capacity theorem for co-operative multicasting in large wireless networks," in *Proceedings of Allerton Conf. on Commun., Control, and Comput.*, Oct 2004.

[69] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," submitted to IEEE Trans. Inform. Theory, 2004.

[70] P. Kumar and H. E. Gamal, "On the throughput-delay tradeoff in cellular multicast," *IEEE Trans. Inform. Theory.*, submitted.

[71] T. Lan and W. Yu, "Input optimization for multi-antenna broadcast channels and per-antenna power constraints," in *Proceedings of IEEE Globecom*, 2004.

[72] N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[73] N. Laneman and G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[74] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[75] A. Lapidoth, S. Shamai, and M. Wigger, "On the capacity of fading mimo broadcast channels with imperfect transmitter side-information," in *Proceedings of 43rd Annual Allerton Conference on Commun., Control and Computing*, Sept. 2005.

[76] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels–Part I: Ergodic capacity," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1083–1102, March 2001.

[77] L. Li, N. Jindal, and A. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1326–1347, April 2005.

[78] H. Liao, "Multiple access channels," Ph.D. dissertation, Depart. of Electrical Engineering, University of Hawaii, 1972.

[79] D. J. Love, R. W. Heath Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wi reless systems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2735–2747, Oct 2003.

[80] A. Lozano and A. Tulino, "Capacity of multiple-transmit multiple-receive antenna architectures," *IEEE Trans. Inform. Theory*, vol. 48, no. 12, pp. 3117–3128, Dec. 2002.

[81] A. Lozano, A. Tulino, and S. Verdu, "Multinantenna capacity: myths and realities," In *Space-Time Wireless Systems: From Array Processing to MIMO Communications*, Ed. H. Bolcskei, D. Gesbert, C. Papadias. and A. J. van der Veen, 2005.

[82] A. Lozano, A. Tulino, and S. Verdu, "High-SNR power offset in multi-antenna communication," in *Proceedings of IEEE Intern. Symp. on Inform. Theory*, July 2004, p. 287.

[83] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan 1999.

[84] T. Marzetta and B. Hochwald, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 46, no. 2, pp. 543–564, March 2000.

[85] A. Molisch, M. Stienbauer, M. Toeltsch, E. Bonek, and R. S. Thoma, "Capacity of MIMO systems based on measured wireless channels," *IEEE J. Select. Areas Commun.*, vol. 20, no. 3, pp. 561–569, April 2002.

[86] A. Moustakas and S. Simon, "Optimizing multiple-input single-output (MISO) communication systems with general Gaussian channels: nontrivial covariance and nonzero mean," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2770–2780, Oct. 2003.

[87] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna s ystems," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2562–2579, Oct 2003.

[88] R. U. Nabar, H. Boelsckei, and F. W. Kneubuhler, "Fading relay channels: performance limits and space-time signal design," *IEEE J. Select. Areas Commun.*, vol. 22, no. 6, pp. 1099–1109, Aug. 2004.

[89] A. Narula, , M. Trott, and G. Wornel, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Trans. Inform. Theory*, vol. 45, no. 7, Nov. 1999.

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

[91] C. Ng and A. Goldsmith, "Transmitter cooperation in ad-hoc wireless networks: does dirty-paper coding beat relaying?" in *Proceedings of IEEE Information Theory Workshop*, Oct. 2004, pp. 277–282.

[92] C. Peel, B. Hochwald, and L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[93] C. Peel, B. Hochwald, and L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, March 2005.

[94] N. Prasad and M. K. Varanasi, "Diversity and multiplexing tradeoff bounds for cooperative diversity schemes," in *Proceedings of IEEE Int. Symp. Inform. Theory*, June 2004, p. 268.

[95] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, March 1998.

[96] F. Rashid-Farrokhi, K. R. Liu, and L. Tassiulas, "Transit beamforming and power control for cellular wireless systems," *IEEE J. Select. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, Oct 1998.

[97] H. Sato, "The capacity of Gaussian interference channel under strong interference (corresp.)," *IEEE Trans. Inform. Theory*, pp. 786–788, Nov. 1981.

[98] A. Sendonaris, E. Erkip, and B. Aazhang, "Increasing uplink capacity via user cooperation diversity," in *Proceedings of Int. Symp. Inform. Theory*, Aug. 1994, p. 156.

[99] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part I: system description," *IEEE Trans. Communications*, vol. 51, no. 11, pp. 1927–1938, Nov 2003.

[100] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-Part II: implementation aspects and performance analysis," *IEEE Trans. Communications*, vol. 51, no. 11, pp. 1939–1948, Nov 2003.

[101] S. Shamai and T. L. Marzetta, "Multiuser capacity in block fading with no channel state information," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 938–942, April 2002.

[102] S. Shamai and A. D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels : Part I," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1877–1894, Nov. 1997.

[103] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proceedings of IEEE Vehicular Tech. Conf.*, May 2001-Spring, pp. 1745–1749.

[104] C. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. Journal*, pp. 379–423, 623–656,, 1948.

[105] C. Shannon, "Communications in the presence of noise," in *Proceedings of IRE*, 1949, pp. 10–21.

[106] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*. Univ. Illinois Press, 1949.

[107] 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.

[108] N. Sidiropoulos, T. Davidson, and Z. Q. Luo, "Transmit beamforming for physical layer multicasting," *IEEE Trans. Sig. Proc.*, to appear.

[109] S. Simon and A. Moustakas, "Optimizing MIMO antenna systems with channel covariance feedback," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 406–417, April 2003.

[110] P. J. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Proceedings of Int. Conf. Commun.*, April 2002, pp. 406–410.

[111] Q. Spencer, L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Sig. Proc.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[112] S. Srinivasa and S. Jafar, "Vector channel capacity with quantized feedback," in *Proceedings of IEEE International Conference on Communications (ICC)*, May 2005.

[113] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, November 1999.

[114] D. Tse and S. Hanly, "Multiaccess fading channels–Part I:Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, no. 7, pp. 2796–2815, November 1998.

[115] A. Tulino, A. Lozano, and S. Verdu, "Capacity of multi-antenna channels in the low power regime," in *Proceedings of IEEE Inform. Theory Workshop*, Oct. 2002, pp. 192–195.

[116] A. Tulino, A. Lozano, and S. Verdu, "Capacity-achieving input covariance for single-user multi-antenna channels," 2005, to appear *IEEE Trans. Wireless Commun.*

[117] E. C. van der Meulen, "A survey of multi-way channels in information theory:1961-1976," *IEEE Trans. Inform. Theory*, pp. 1–37, Jan. 1977.

[118] E. C. van der Meulen, "Some reflections on the interference channel," *Communications and Cryptography: Two sides of one tapestry, R. E. Blahut, D. J. Costello, and T. Mittelholzer, Eds. Boston, MA: Kluwer*, pp. 409–421, 1994.

[119] S. Venkatesan, S. Simon, and R. Valenzuela, "Capacity of a Gaussian MIMO channel with nonzero mean," in *Proceedings of IEEE Veh. Tech. Conf.*, Oct. 2003.

[120] S. Verdu, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319–1343, June 2002.

[121] S. Vishwanath and S. Jafar, "On the capacity of vector interference channels," in *Proc. IEEE Inform. Theory Workshop*, Oct 2004.

[122] S. Vishwanath, S. Jafar, and A. Goldsmith, "Optimum power and rate allocation strategies for multiple access fading channels," in *Proceedings of Vehicular Tech. Conf.*, May 2000, pp. 2888–2892.

[123] S. Vishwanath, N. Jindal, and A. Goldsmith, “Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[124] E. Visotsky and U. Madhow, “Space-time transmit precoding with imperfect feedback,” *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sep. 2001.

[125] P. Viswanath, D. Tse, and R. Laroia, “Opportunistic beamforming using dumb antennas,” *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1277–1294, June 2002.

[126] P. Viswanath and D. N. Tse, “Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality,” *IEEE Trans. Inform. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[127] P. Viswanath, D. N. Tse, and V. Anantharam, “Asymptotically optimal water-filling in vector multiple-access channels,” *IEEE Trans. Inform. Theory*, vol. 47, no. 1, pp. 241–267, Jan. 2001.

[128] H. Viswanathan and S. Venkatesan, “Asymptotics of sum rate for dirty paper coding and beamforming in multiple-antenna broadcast channels,” in *Proceedings of Allerton Conf. on Commun., Control, and Comput.*, Oct. 2003.

[129] H. Viswanathan, S. Venkatesan, and H. C. Huang, “Downlink capacity evaluation of cellular networks with known interference cancellation,” *IEEE J. Sel. Areas Commun.*, vol. 21, pp. 802–811, June 2003.

[130] B. Wang, J. Zhang, and A. Host-Madsen, “On the capacity of MIMO relay channels,” *IEEE Trans. Inform. Theory*, vol. 51, no. 1, pp. 29–43, Jan. 2005.

[131] H. Weingarten, Y. Steinberg, and S. Shamai, “Capacity region of the degraded MIMO broadcast channel,” in *Proceedings of Int. Symp. Inform. Theory*, July 2004.

[132] J. Winters, “On the capacity of radio communication systems with diversity in a Rayleigh fading environment,” *IEEE J. Sel. Areas Commun.*, vol. 5, pp. 871–878, June 1987.

[133] J. Winters, J. Salz, and R. Gitlin, “The impact of antenna diversity on the capacity of wireless communication systems,” *IEEE Trans. Commun.*, vol. 42, no. 234, pp. 1740–1751, Feb. 1994.

[134] A. Wyner, “Shannon-theoretic approach to a Gaussian cellular network,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 1713–1727, Nov. 1994.

[135] T. Yoo and A. Goldsmith, “On the optimality of multi-antenna broadcast scheduling using zero-forcing beamforming,” to appear *IEEE J. Select. Areas Commun.*, 2005.

[136] W. Yu, “A dual decomposition approach to the sum power Gaussian vector multiple access channel sum capacity problem,” in *Proceedings of Conference on Information Sciences and Systems (CISS)*, 2003.

[137] W. Yu, “Spatial multiplex in downlink multiuser multiple-antenna wireless environments,” in *Proceedings of IEEE Globecom*, 2003.

[138] W. Yu and J. Cioffi, “Trellis precoding for the broadcast channel,” in *Proceedings of Global Commun. Conf.*, Oct. 2001, pp. 1344–1348.

[139] W. Yu and J. M. Cioffi, “Sum capacity of Gaussian vector broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1875–1892, Sept. 2004.

[140] W. Yu, G. Ginis, and J. Cioffi, “An adaptive multiuser power control algorithm for VDSL,” in *Proceedings of Global Commun. Conf.*, Oct 2001, pp. 394–398.

[141] W. Yu, W. Rhee, S. Boyd, and J. Cioffi, “Iterative water-filling for Gaussian vector multiple access channels,” *IEEE Trans. Inform. Theory*, vol. 50, no. 1, pp. 145–152, Jan. 2004.

[142] W. Yu, W. Rhee, and J. Cioffi, "Optimal power control in multiple access fading channels with multiple antennas," in *Proceedings of Int. Conf. Commun.*, 2001, pp. 575–579.

[143] M. Yuksel and E. Erkip, "Can virtual MIMO mimic a multi-antenna system: Diversity-multiplexing tradeoff for wireless relay networks," in *IEEE Communication Theory Workshop*, May 2005.

[144] 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 2005.

[145] L. Zheng and D. N. Tse, "Packing spheres in the Grassmann manifold: A geometric approach to the non-coherent multi-antenna channel,," *IEEE Trans. Inform. Theory*, vol. 48, no. 2, pp. 359–383, Feb 2002.