Duality, Dirty Paper Coding, and Capacity for Multiuser Wireless Channels

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

ABSTRACT. We introduce techniques that shed new light on the capacity analysis of multiuser channels, and apply these techniques to broadcast and multiple access channels. We first determine a duality between broadcast and multiple access channels that can be used to obtain the capacity region and optimal transmission strategy for one channel based on the capacity-achieving transmission strategy and region for the dual channel. This duality result is applicable to additive Gaussian noise and fading channels for several different notions of fading channel capacity, including ergodic capacity, outage capacity, and minimum rate capacity. Duality provides a powerful connection between the broadcast channel and multiple access channel capacity regions that solves many open capacity problems, and also greatly simplifies the calculation of known regions. We next consider the dirty paper coding strategy of Costa for broadcast channels. Dirty paper coding exploits the known interference between users in a broadcast channel to presubtract out this interference. We apply this coding strategy to broadcast channels with multiple antennas at the transmitter and receiver (the MIMO channel). We show that dirty paper coding provides significant rate gains over the high-data-rate (HDR) strategy used in current cellular systems. We next turn our attention to the MIMO broadcast channel capacity region. Since this channel is in general non-degraded, its capacity region remains an unsolved problem. We first establish a duality between the achievable region of the MIMO broadcast channel using dirty paper coding and the capacity region of the dual MIMO multiple-access channel, which is easy to compute. We then show that dirty paper coding achieves sumrate capacity of the MIMO broadcast channel. The proof exploits duality, dirty paper coding, and clever upper bounding techniques.

1. Introduction

The main focus of this paper is to establish a connection between the capacity regions of broadcast channels (BCs) and multiple access channels (MACs), then exploit this connection to solve new capacity problems. Despite extensive study of the BC and MAC independently, no relationship between them has previously been discovered. We show that the Gaussian MAC and BC are essentially duals of each other. As a result, the capacity regions of the BC and the MAC with the same channel gains and the same noise power at every receiver can be derived from one

Key words and phrases. Broadcast channels, multiple access channels, capacity region, duality, fading channels, MIMO, dirty paper coding, Sato upper bound, sumrate capacity.

Work supported by ONR grants N00014-99-1-0578 and N00014-02-1-0003. The work of S. Vishwanath was supported by a Stanford Graduate Fellowship.

another. In particular, we show that the capacity region of a fading or constant Gaussian BC is equal to the union of the capacity regions of the dual Gaussian MAC where the union is taken over power constraints on the individual users of the MAC that sum up to the BC power constraint. This result allows us to directly compute the capacity region of a Gaussian BC from the capacity region of its dual Gaussian MAC. We also show that the capacity region of the Gaussian MAC equals the intersection of capacity regions of the dual Gaussian BC where the intersection is taken over appropriately scaled channel gains and the total BC power equals the sum of powers of the dual MAC users. Moreover, we obtain an explicit relationship between the optimum power allocation scheme and the decoding order used to achieve points along the boundary of the dual MAC and BC capacity regions. The relationships between the capacity regions and optimal transmission strategies for dual MAC and BCs holds for several different notions of fading channel capacity, including ergodic capacity, outage capacity, and minimum rate capacity.

The duality relationships between the capacity regions of the Gaussian MAC and BC are quite powerful and interesting. First they indicate a direct connection between known capacity results for dual MAC and broadcast channels. More importantly, there are several BCs and MACs where the capacity region and optimal transmission strategy for one channel is either unknown or very hard to compute, whereas the regions are known and easy to compute for the dual channel. In these cases we can use duality to obtain new capacity results. In particular, we will show that the minimum rate capacity region and optimal transmission strategy for a fading MAC can be obtained from the dual BC capacity region, yet no direct method exists to compute this region.

The capacity of a broadcast channel with multiple antennas at the transmitter and receiver (the MIMO BC) is an open problem due to the lack of a general theory on the capacity of non-degraded broadcast channels. Pioneering work in this area by Caire and Shamai [1] developed an achievable set of rates for a broadcast channel with two transmit antennas and one receive antenna at each node based on the dirty paper coding result of Costa [2]. This coding strategy allows a channel with interference known at the transmitter to achieve the same data rate as if the interference did not exist. The original coding strategy was described as writing on dirty paper, where the structure introduced by the "dirt" is exploited in the code design. Computing the corresponding set of achievable rates for the MIMO BC is extremely complex, especially for a larger number of antennas at either the transmitter or the receivers.

We establish a duality between the achievable region of the MIMO BC obtained using dirty paper coding and the capacity region of the dual MIMO MAC for any number of transmit and receive antennas. This result greatly simplifies calculation of the achievable region for the MIMO BC obtained with dirty paper coding, since we can use recently developed iterative waterfilling techniques to obtain the capacity region for the dual MIMO MAC capacity region [4] and then apply duality [3]. We also show that dirty paper coding achieves the sumrate point of the MIMO BC channel capacity region (maximum achievable sum of all users' rates). This proof applies the Sato upper bound [5] to the broadcast channel by allowing joint signal detection at all receivers and assuming worst-case noise, a technique first proposed in [1] that we generalize to multiple transmit and receive antennas.

The remainder of this paper is organized as follows. Section 2 introduces the duality between broadcast and MAC channels and shows how this can be used to compute their dual capacity regions. Section 3 describes the different notions of capacity associated with multiuser fading channels, including the ergodic, outage, and minimum rate capacities. Sections 4 and 5 derive the ergodic and minimum rate capacity regions, respectively, for fading broadcast and MAC channels along with their duality relationship. Section 6 defines the dual MIMO MAC and BCs. Sections 7 outlines the achievable rate region for a broadcast MIMO fading channel with dirty paper coding, and illustrates the gains of dirty paper coding relative to the high-data-rate (HDR) scheme currently used in cellular systems. Section 8 establishes a duality between the dirty paper region for the MIMO BC and the capacity region of the dual MIMO MAC. Section 9 derives an upper bound on the MIMO BC capacity region based on joint detection and worst-case noise, and shows that this bound is achievable at the sumrate point using dirty paper coding. Thus, the sumrate capacity of the MIMO BC is obtained for any number of users, transmit antennas, and receive antennas. Our conclusions are discussed in Section 8.

2. Duality between Broadcast and Multiple Access Channels

In this section we establish a duality between broadcast and multiple access channels and show how this duality can be used to compute the capacity of one channel from the capacity of its dual. A discrete-time BC, where one transmitter sends information to M receivers, is described mathematically by

$$Y_j[i] = \sqrt{h_j[i]}X[i] + n_j[i], \tag{2.1}$$

where $X[i]$ is the transmitted signal, $Y_j[i]$ is the signal received by the jth user, $n_j[i]$ is the receiver noise sample at time i of the jth user, and $h_j[i]$ is the time-varying channel power gain at time i of the jth user. A discrete-time MAC, where many transmitters send information to one receiver, is described mathematically by

$$Y[i] = \sum_{j=1}^{M} \sqrt{h_j[i]}X_j[i] + n[i], \tag{2.2}$$

where $X_j[i]$ is the signal transmitted at time i by the jth transmitter, $h_j[i]$ is the channel power gain between the jth transmitter and the receiver, $Y[i]$ is the received signal at time i, and $n[i]$ is the receiver noise sample at time i. We assume in our system models that the noise power of all the receivers in the BC and the single receiver in the MAC are equal to σ^2. Also, the term $h_j[i]$ is the channel power gain of receiver j in the BC (downlink) and $h_j[i]$ is also the channel power gain of transmitter j in the MAC (uplink). We call this BC the **dual** of the MAC, and vice versa. In our broadcast and MAC models the channel gains can be constant or changing with i (fading). We assume in both cases that all channel gains $h_j[i]$ are known to the transmitter(s) and receiver(s) at time i.

The dual channels have several key differences as well as some key similarities. Specifically, the BC has a single power constraint associated with the transmitter, whereas the MAC has a different individual power constraint for each user. In addition, the interference signal on the BC has the same channel gain as the desired signal, whereas in the MAC these signals are received with different powers (the near-far effect). Despite these differences, a superposition coding strategy is optimal for both channels, and the optimal decoders for each channel exploit successive decoding and interference cancellation. The duality relationship between the two channels is based on exploiting their similar encoding and decoding strategies while bridging their differences by summing the individual MAC power constraints to obtain the BC power constraint and scaling the BC gains to achieve the near-far effect of the MAC.

We now state the explicit duality result, proved in [9], that the capacity region of a Gaussian BC can be characterized in terms of the capacity region of the dual MAC. For simplicity the theorem is stated for a two-user channel, but it applies to any number of users.

THEOREM 2.1. *The capacity region of a constant Gaussian BC with power P is equal to the union of capacity regions of the dual MAC with power P_1 and P_2 such that $P_1 + P_2 = P$:*

$$\mathcal{C}_{BC}(P; h_1, h_2) = \bigcup_{0 \le P_1 \le P} \mathcal{C}_{MAC}(P_1, P - P_1; h_1, h_2). \tag{2.3}$$

In [9] it is also shown that the boundary of each MAC capacity region touches the boundary of the BC capacity region at a *different* point. More generally, each point along the BC capacity region boundary is intersected by a different MAC capacity region boundary. The BC and MAC power schemes corresponding to the intersection at rate point (R_1, R_2) are related by the following duality power transformation [9]:

$$\begin{aligned} R_1^{BC} &= \log\left(1 + \frac{h_1 P_1^{BC}}{\sigma^2}\right) = \log\left(1 + \frac{h_1 P_1^{MAC}}{h_2 P_2^{MAC} + \sigma^2}\right) = R_1^{MAC} \\ R_2^{BC} &= \log\left(1 + \frac{h_2 P_2^{BC}}{h_2 P_1^{BC} + \sigma^2}\right) = \log\left(1 + \frac{h_2 P_2^{MAC}}{\sigma^2}\right) = R_2^{MAC}. \end{aligned} \tag{2.4}$$

Thus, given the power pair (P_1, P_2) that achieves rates (R_1, R_2) for one channel, the corresponding power allocation to achieve these rates on the dual channel must satisfy (2.4). Moreover, the decoding order to achieve a given rate (R_1, R_2) on the MAC is the opposite of the optimal encoding order to achieve this rate on the dual BC: in the BC the user with the best channel gain is decoded last, whereas in the MAC the user with the best channel gain is decoded first. More details on these explicit power and decoding order transformations can be found in [9]. The duality relationship between the Gaussian BC and MAC defined by Theorem 2.1 is illustrated in Figure 1.

We now describe how the capacity region of the MAC can be characterized in terms of the capacity region of the dual BC. This theorem is also proved in [9].

THEOREM 2.2. *The capacity region of a constant Gaussian MAC is equal to the intersection of the capacity regions of the scaled dual BC over all scalings:*

$$\mathcal{C}_{MAC}(P_1, P_2; h_1, h_2) = \bigcap_{\alpha > 0} \mathcal{C}_{BC}\left(\frac{P_1}{\alpha} + P_2; \alpha h_1, h_2\right). \tag{2.5}$$

Moreover, the optimal power allocation and decoding order for the MAC can be obtained by reversing the dual BC encoding order and applying the power transformation in (2.4). Thus, the optimal transmission strategy to achieve any point on the MAC capacity region can be obtained from the dual BC channel. Note that since MAC region is a pentagon, the BC channels characterized by $\alpha = 0$, $\alpha = h_2/h_1$ and $\alpha = \infty$ are sufficient to form the pentagon. If $\alpha = h_2/h_1$, the channel gains of both users are the same and the BC capacity region is bounded

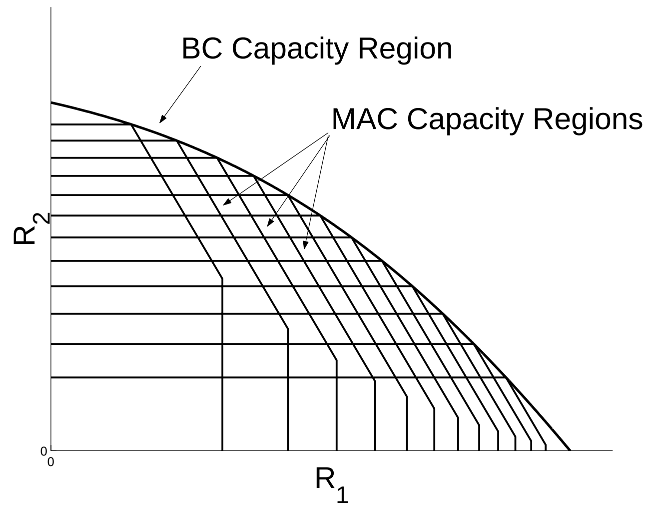

FIGURE 1. Gaussian BC Capacity Region as Union of Dual MAC Regions

by a straight line segment because the capacity region can be achieved by time-sharing. It can be shown [9] that as $\alpha \to 0$, $R_1 \to \log(1 + \frac{h_1 P_1}{\sigma^2})$ and $R_2 \to \infty$. Similarly, as $\alpha \to \infty$, $R_1 \to \infty$ and $R_2 \to \log(1 + \frac{h_2 P_2}{\sigma^2})$. These two regions bound the vertical and horizontal line segments, respectively, of the MAC capacity region. All scaled BC capacity regions except the $\alpha = h_2/h_1$ channel intersect the MAC at exactly one of the two corner points of the MAC region. The $\alpha = h_2/h_1$ channel intersects the MAC region along its time-sharing line. These boundaries along with the relationship defined by Theorem 2.2 is illustrated in Figure 2.

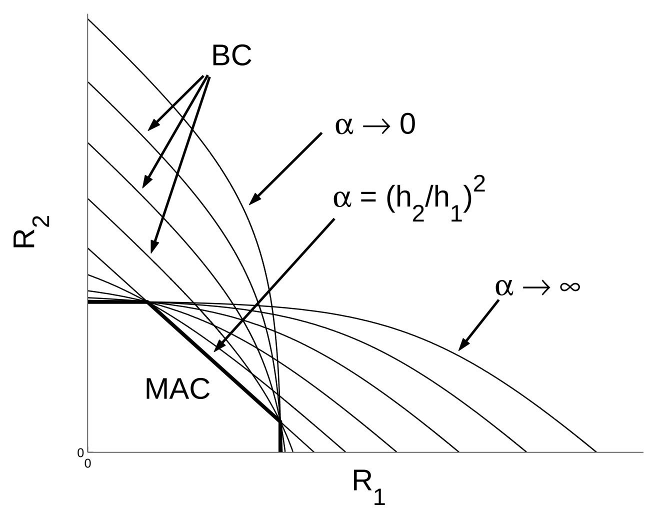

FIGURE 2. Gaussian MAC Capacity Region as Intersection of Dual BC Regions

3. Capacity of Fading Multiuser Channels

Fading channels exhibit random amplitude variations over time. Capacity for such channels depends on what is known about the channel fading at both transmitter and receiver, and whether or not capacity can be averaged over many channel states or must be achieved in each state. We focus on the case where both transmitter and receiver track the channel state perfectly and instantaneously. We also assume a slowly fading channel relative to codeword length, so the channel is constant for the transmission of a single codeword. For this channel there are four different definitions of capacity region for multiuser channels:

- **Ergodic capacity region:** The maximum rate region averaged over all channel states.
- **Outage capacity region:** The maximum rate region that can be achieved in all non-outage fading states subject to some outage probability. Under an outage probability of zero this is called the zero-outage capacity region.
- **Minimum rate capacity region:** The maximum rate region averaged over all fading states subject to a minimum rate for each user required in every fading state.
- **Minimum rate capacity region with outage:** The maximum rate region averaged over all fading states subject to some minimum rate for each user required in all non-outage fading states, with some nonzero outage probability.

The ergodic capacity region for fading broadcast and multiple access channels has been derived in [10] and [11] respectively. The optimal power allocation to achieve this capacity corresponds to a multi-level water-filling over both time (i.e. fading states) and users. As expected, users are allocated the most power when their channels are strong, and little, if any, power when their channels are weak. This results in a channel-dependent delay, which may not work for delay-constrained applications like voice and video.

The outage capacity region for fading broadcast and multiple access channels has been derived in [12] and [13, 14], respectively. For outage capacity each user maintains a constant rate some percentage of time with no data transmitted the rest of the time. The optimal power allocation over nonoutage must effectively invert the fading to eliminate channel variations such that a constant rate can be maintained. Clearly the ergodic capacity region exceeds the outage capacity region, since outage capacity has an additional constraint of constant rate transmission in all fading states. Since the weak channel states will effectively drive the maximum constant rate that can be maintained, outage capacity fails to fully exploit good channel states. However, outage capacity has the advantage that with fixed rate transmission, there is no channel-dependent delay.

Minimum rate capacity combines the concepts of ergodic and outage capacity to maximize the rate averaged over all channel states while maintaining some minimum rate in every fading state. Minimum rate capacity for fading broadcast channels has been derived in [15]. The optimal power allocation for the broadcast channel is a two-step process, where first the minimum power required to achieve the minimum rates in all fading states is allocated, and then the excess power to maximize average rate in excess of the minimum rate is allocated. The optimal allocation of the excess power is a multi-level water-filling based on effective noise that incorporates the minimum rate constraints. We will use these results in Section 5 to obtain the minimum rate capacity, optimal power allocation, and optimal decoding order of the MAC using duality.

The minimum rates associated with the minimum rate capacity region must be within the zero-outage capacity region of the channel. Thus, the boundary of the minimum rate capacity region lies between the ergodic and zero-outage boundaries. The difference between the ergodic and zero-outage capacity regions is a good indicator of the degradation in capacity due to minimum rate constraints. If the zero-outage capacity region is much smaller than the ergodic capacity region, the minimum rate capacity region is generally significantly smaller than the ergodic capacity region. Alternatively, if the zero-outage capacity region is close to the ergodic capacity region, then clearly the minimum rate capacity region will be close as well. More details on these relationships for the fading BC can be found in [15].

4. Fading Channel Capacity and Duality

For fading channels the ergodic capacity region is obtained by averaging over the constant capacity regions. Thus, the duality relationships obtained for BC and MAC Gaussian channels in the previous section extend to the ergodic capacity regions for fading BC and MAC channels. This extension is captured in the following theorems, proved in [9].

THEOREM 4.1. *The capacity region of a fading Gaussian BC with power constraint $\overline{P}$ is equal to the union of ergodic capacity regions of the dual MAC with power constraints $\overline{P}_1$ and $\overline{P}_2$ such that $\overline{P}_1 + \overline{P}_2 = \overline{P}$:*

$$\mathcal{C}_{BC}(\overline{P}; H_1, H_2) = \bigcup_{0 \le \overline{P}_1 \le \overline{P}} \mathcal{C}_{MAC}(\overline{P}_1, \overline{P} - \overline{P}_1; H_1, H_2). \tag{4.1}$$

The ergodic BC capacity region obtained by a union of ergodic capacity regions of the dual MAC is illustrated in Figure 3.

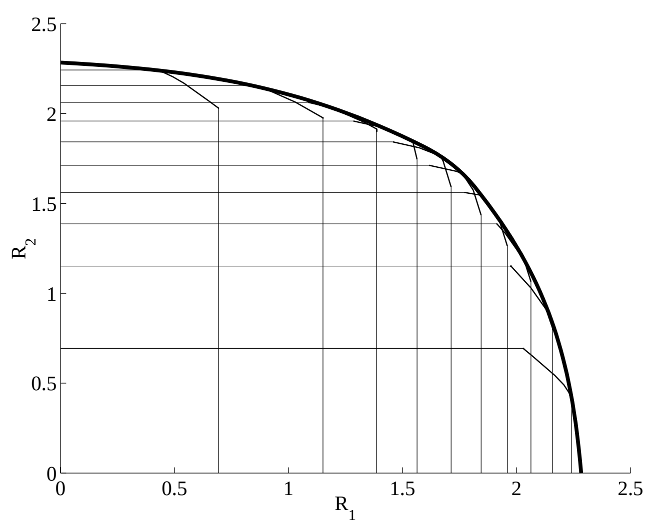

FIGURE 3. Ergodic BC Capacity Region as Union of Dual MAC Regions

The next theorem shows that the ergodic capacity region of the fading MAC can be obtained as the intersection of the regions for the dual BC.

THEOREM 4.2. *The ergodic capacity region of a fading MAC is equal to the intersection of the ergodic capacity regions of the scaled dual BC over all scalings:*

$$\mathcal{C}_{MAC}(\overline{P}_1, \overline{P}_2; H_1, H_2) = \bigcap_{\alpha>0} \mathcal{C}_{BC}(\frac{\overline{P}_1}{\alpha} + \overline{P}_2; \alpha H_1, H_2). \tag{4.2}$$

The ergodic capacity region of the fading MAC obtained by intersecting the ergodic capacity regions of the dual BC is shown in Figure 4.

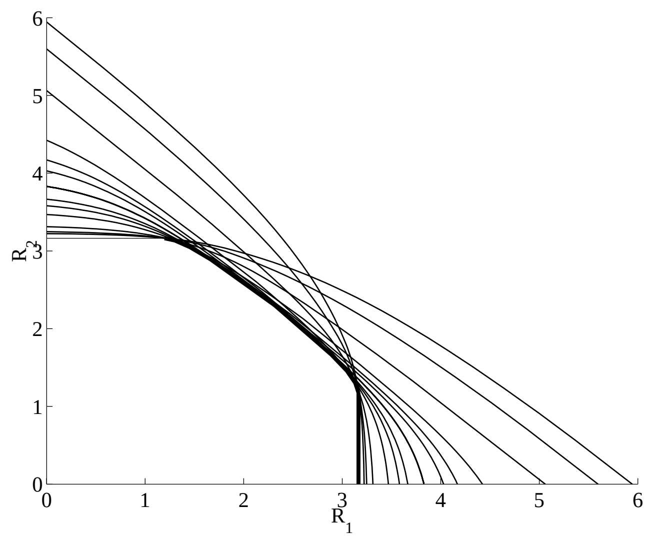

FIGURE 4. Ergodic MAC Capacity Region as Intersection of Dual BC Regions

Note that in addition to the capacity region duality captured in the above two theorems, explicit power and decoding order transformations between the regions are also obtained in [9]. Thus, given a point on the capacity region and optimal transmission strategy (power allocation and decoding order) for one channel, the corresponding optimal transmission strategy for achieving that same rate point on the dual channel can be explicitly obtained.

The duality of the Gaussian MAC and BC is a general result that also holds for all definitions of fading channel capacity given in the previous section. The outage and minimum rate definitions of capacity differ from the ergodic capacity definition since they have restrictions on the *instantaneous* transmission rates (i.e. the state-by-state rates). However, as shown in [9], the duality power transformation for these ergodic regions preserves state-by-state rates (i.e. the same instantaneous rates are achieved in the dual channels). Thus, Theorems 4.1 and 4.2 hold for minimum rate and outage capacity as well as ergodic capacity. In the next two sections we use duality to obtain new capacity results from known results for the dual channels.

5. Minimum Rate Capacity

In this section we show that the minimum rate capacity region of the MAC and BC are duals of each other, so that the capacity region and capacity-achieving transmission strategy for one channel can be found from the other dual channel. We then use this result to find the minimum rate capacity region of the MAC using the known minimum rate capacity region of the BC obtained in [15]. The minimum rate capacity of the MAC and BC, denoted respectively by $\mathcal{C}_{MAC}^{min}(\overline{P}_1, \overline{P}_2, \mathbf{R}^*; H_1, H_2)$ and $\mathcal{C}_{BC}^{min}(\overline{P}_1 + \overline{P}_2, \mathbf{R}^*; H_1, H_2)$, is defined as the maximum ergodic capacity that can be obtained while ensuring that a set of minimum rates $\mathbf{R}^* = (R_1^*, \dots, R_M^*)$ is maintained for all users in *all* fading states. Clearly the minimum rates themselves must be in the zero-outage capacity of the channel for the rates to be achievable in all fading states. Every feasible $\mathbf{R}^*$ defines a different minimum rate capacity region (i.e. the minimum rate capacity region is a function of $\mathbf{R}^*$ in addition to the channel and power constraints).

We now state two theorems, proved in [9], indicating the duality between the minimum rate capacity region of the fading MAC and BC.

THEOREM 5.1. *The minimum rate capacity region of a fading Gaussian BC with perfect channel information at the transmitter and receivers is given by:*

$$\mathcal{C}_{BC}^{min}(\overline{P}, \mathbf{R}^*; H_1, H_2) = \bigcup_{0 \le \overline{P}_1 \le \overline{P}} \mathcal{C}_{MAC}^{min}(\overline{P}_1, \overline{P} - \overline{P}_1, \mathbf{R}^*; H_1, H_2), \tag{5.1}$$

for all $\mathbf{R}^* \in \mathcal{C}_{BC}^0(\overline{P}; H_1, H_2)$.

THEOREM 5.2. *The minimum rate capacity region of a fading Gaussian MAC with perfect channel information at the transmitters and receiver is given by:*

$$\mathcal{C}_{MAC}^{min}(\overline{P}_1, \overline{P}_2, \mathbf{R}^*; H_1, H_2) = \bigcap_{\alpha \ge 0} \mathcal{C}_{BC}^{min}(\frac{\overline{P}_1}{\alpha} + \overline{P}_2, \mathbf{R}^*; \alpha H_1, H_2) \tag{5.2}$$

for all $\mathbf{R}^* \in \mathcal{C}_{MAC}^0(\overline{P}_1, \overline{P}_2; H_1, H_2)$.

While these theorems characterize the relationship between the dual capacity regions, we also wish to obtain an explicit transformation between the optimal transmission strategies. The power and decoding order transformations to achieve any point on the minimum rate capacity region of the fading MAC from the corresponding region of the dual BC that intersects that point is obtained in [15]. Thus we can explicitly characterize the MAC minimum rate capacity region and its corresponding capacity-achieving transmission strategy without having to directly solve for it. Specifically, these results are obtained by applying the power and decoding transformations to the known optimal transmission strategies of the dual BC obtained in [15].

6. Multi-Antenna Multiple Access and Broadcast Channels

Adding multiple antennas to the transmitter and receiver has recently been shown to greatly increase the capacity of wireless channels [16, 17]. As a result, the capacity and capacity region for single and multiuser multiple input multiple output (MIMO) systems is of great interest. While the capacity region of the MIMO MAC has been fully characterized [7, 17, 18], the capacity of the MIMO BC is difficult to obtain due to a lack of general theory on the capacity of non-degraded broadcast channels. However, in the next few sections we show that duality combined with a novel encoding technique called dirty paper coding can shed much light on the achievable rates of MIMO BCs and in fact determine their sumrate capacity point.

Consider a K-user MIMO Gaussian BC in which receiver k has r_k receive antennas and the transmitter has t transmit antennas. The $r_k \times t$ matrix $\mathbf{H_k}$ defines the channel gains from the transmitter to the r_k antennas of receiver k. Each receiver has additive white Gaussian noise with unit variance. The dual MIMO MAC channel is arrived at by converting the receivers in the BC into transmitters in the MAC and converting the t-antenna transmitter into a t-antenna receiver with additive noise of unit variance. The channel gains of the dual MAC are the same as that of the broadcast channel. Specifically, if $\mathbf{H_k}$ is the $t \times r_k$ matrix defining the channel gains from t-antenna transmitter to the kth receiver with r_k antennas in the BC, then $\mathbf{H_k}^\dagger$ is the $r_k \times t$ matrix defining the channel gains from transmitter k with r_k antennas to the t-antenna receiver in the dual MAC.

The MIMO BC is a nondegraded broadcast channel due to the multiple receive antennas. Thus, receivers are not necessarily "better" or "worse" than one another. The capacity region of a nondegraded broadcast channel is an open problem in general. However, the sumrate capacity of the MIMO BC for two transmit antennas ($t = 2$) and two users with one receive antenna each ($r_1 = r_2 = 1$) was obtained in pioneering work by Caire and Shamai [1]. In that work a set of achievable rates (the achievable region) was obtained by using the dirty paper coding technique of Costa [2] (also known as coding for non-causally known interference). It was also shown that the dirty paper coding technique at the sumrate point equals the Sato upper bound with joint receiver decoding and worst-case noise. Thus, dirty paper coding achieves the sumrate capacity of the MIMO BC for the two-user case with $t = 2$ and $r_1 = r_2 = 1$. However, computing the dirty paper coding region is extremely complex and this approach does not appear to work for the more general class of channels (any number of users with any number of transmit and receive antennas). However, we will see shortly that exploiting duality can greatly simplify this calculation. First, however, we must describe dirty paper coding in more detail.

7. Dirty Paper Coding

The basic premise of dirty paper coding is that if interference to a given user is known in advance, the encoding strategy can exploit the structure of the interference such that the capacity is the same as if there was no interference at all. The encoding strategy cleverly distributes the codewords based on the interference, and the decoder must know how to read these codewords. Dirty paper coding is a natural technique to use on the broadcast channel since the interference between all users is known. An achievable region for the MIMO BC based on dirty paper coding was first proposed in [1]. In [19], the region was extended to the more general multiple-user, multiple-antenna case using the following extension of the "dirty paper result" [2] to the vector case:

LEMMA 7.1. *[Yu, Cioffi] Consider a channel with* $\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{s}_k + \mathbf{n}_k$*, where* $\mathbf{y}_k$ *is the received vector,* $\mathbf{x}_k$ *the transmitted vector,* $\mathbf{s}_k$ *the vector Gaussian interference, and* $\mathbf{n}_k$ *the vector white Gaussian noise. If* $\mathbf{s}_k$ *and* $\mathbf{n}_k$ *are independent and non-causal knowledge of* $\mathbf{s}_k$ *is available at the transmitter but not at the receiver, then the capacity of the channel is the same as if* $\mathbf{s}_k$ *is not present.*

The transmission strategy under dirty paper coding first picks a codeword for receiver 1. The transmitter then chooses a codeword for receiver 2 with full (non-causal) knowledge of the codeword intended for receiver 1. Therefore 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 as interference. This process continues for all K receivers. Since the ordering of the users clearly matters in such a procedure, the following is an achievable set of rates

$$R_{\pi(i)} = \frac{1}{2}\log\frac{|\mathbf{I} + \mathbf{H}_{\pi(i)}(\sum_{j\geq i}\mathbf{\Sigma}_{\pi(j)})\mathbf{H}^{\dagger}_{\pi(i)}|}{|\mathbf{I} + \mathbf{H}_{\pi(i)}(\sum_{j>i}\mathbf{\Sigma}_{\pi(j)})\mathbf{H}^{\dagger}_{\pi(i)}|}, \quad i = 1, \ldots, K,$$

where $\pi(i)$ is a permutation that depends on the encoding order. The dirty-paper region $\mathcal{C}_{\mathrm{dirtypaper}}(P, \mathbf{H})$ is defined as the union of all such rates vectors over all covariance matrices $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K$ such that $\mathrm{Tr}(\mathbf{\Sigma}_1 + \ldots \mathbf{\Sigma}_K) = \mathrm{Tr}(\mathbf{\Sigma}_x) \leq P$ and over all permutations $(\pi(1), \ldots, \pi(K))$. The transmitted signal is $\mathbf{x} = \mathbf{x_1} + \ldots + \mathbf{x_K}$ and the input covariance matrices are of the form $\mathbf{\Sigma}_i = \mathbb{E}[\mathbf{x_i}\mathbf{x_i}^{\dagger}]$.

Many cellular systems use a technique called high data rate (HDR) on the downlink for high rate data transmission. HDR is an adaptive technique whereby the user with the best channel gain is allocated the total bandwidth and power. When the channel changes, the bandwidth and power are reallocated, and fairness is achieved over time if all users have the same channel statistics. Dirty paper coding is a more efficient technique since all users are active at the same time, and the power allocation and encoding/decoding order is optimized relative to the channel. We have recently computed the gains of dirty paper coding relative to HDR. Note that computing the dirty paper coding region directly is intractable, however the duality method described below allows this computation to be made based on the dual MIMO MAC, whose capacity is easy to compute. The capacity gains of dirty paper coding relative to HDR are illustrated in Figure 5, where we see up to a factor of seven capacity increase.

While Figure 5 indicates that dirty paper coding exhibits large gains over other techniques, we would like to see if it actually achieves Shannon capacity. In order to address this question, we will first establish a duality between the dirty paper achievable region for the MIMO BC and the MIMO MAC capacity region.

8. MIMO MAC Capacity Region

The capacity region of a general MIMO MAC was obtained in [7, 17, 18]. We now describe this capacity region for the dual MIMO MAC of our BC model. For any set of powers $(P_1, \ldots, P_K)$, the capacity of the MIMO MAC is

$$\mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K; \mathbf{H}^{\dagger}) \triangleq \bigcup_{\{\mathrm{Tr}(\mathbf{P}_1)\leq P_i \ \forall i\}} \Big\{ (R_1, \ldots, R_K) : \tag{8.1}$$

$$\left.\sum_{i\in S} R_i \leq \frac{1}{2}\log|\mathbf{I} + \sum_{i\in S}\mathbf{H}_i^{\dagger} P_i \mathbf{H}_i| \ \forall S \subseteq \{1, \ldots, M\}\right\} \tag{8.2}$$

For $P > 0$, we denote by $\mathcal{C}_{\mathrm{union}}(P, \mathbf{H}^{\dagger})$

$$\mathcal{C}_{\mathrm{union}}(P, \mathbf{H}^{\dagger}) = \bigcup_{\sum_{i=0}^{K} P_i \leq P} \mathcal{C}_{\mathrm{MAC}}(P_1, \ldots, P_K; \mathbf{H}^{\dagger}). \tag{8.3}$$

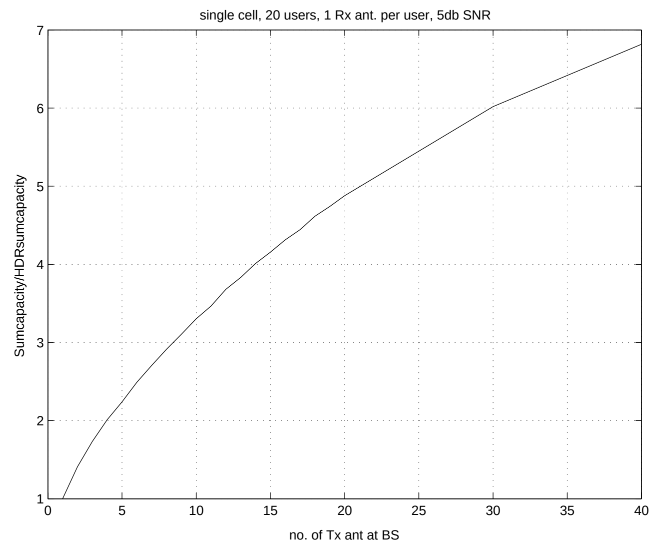

FIGURE 5. Capacity Gain of Dirty Paper Coding over HDR

It can be easily shown that this region is the capacity region of a MAC when the transmitters have a sum power constraint instead of individual power constraints but are not allowed to cooperate.

Our next theorem, proved in [20], indicates that the capacity region of the MIMO MAC with a total power constraint of P for the K transmitters is the same as the dirty paper region of the dual MIMO BC with power constraint P. In other words, any rate vector that is achievable in the dual MAC with power constraints $(P_1, \dots, P_K)$ is in the dirty paper region of the BC with power constraint $\sum_{i=1}^{K} P_i$. Conversely, any rate vector that is in the dirty paper region of the BC is also in the dual MIMO MAC region with the same total power constraint.

THEOREM 8.1. *The dirty paper region of a MIMO BC channel with power constraint P is equal to the the capacity region of the dual MIMO MAC with sum power constraint P.*

$$\mathcal{C}_{\text{dirtypaper}}(P, \mathbf{H}) = \mathcal{C}_{\text{union}}(P, \mathbf{H}^{\dagger}).$$

In addition to the capacity region relationships, we also wish to obtain an explicit transformation between the transmission strategies of dirty paper coding for the MIMO BC and the corresponding capacity region of the MIMO MAC. These explicit transformations will be briefly summarized here, details can be found in [20]. Suppose we are given a rate point in the MIMO MAC, and the corresponding input covariance matrix and decoding order that achieves this rate point. We need to find the corresponding input covariance matrix and encoding order to achieve the same rate point on the MIMO BC using dirty paper coding. It is quite difficult to find the input covariance matrix that yields the same rates and also satisfies the power constraint. It turns out that the correct transformation defines an effective channel and then flips it to obtain the covariance matrix. The encoding order for the MIMO BC under dirty paper coding is the reverse of the decoding order for the MIMO MAC. Explicit formulas for these transformations can be found in [20]. Interestingly, this transformation also shows that beamforming (transmit covariance matrix of rank one) is optimal for dirty paper coding on the broadcast channel with multiple transmit antennas but only one receive antenna at each user. That is because the dual MIMO MAC channel has one transmit antenna and thus a rank one covariance matrix, and the transformation defined in [20] preserves this rank.

Similarly, suppose a rate point achievable using dirty paper coding on the MIMO BC and its corresponding input covariance matrix and encoding order is known. The covariance matrix for the dual MIMO MAC is also obtained by a "flipping of the effective channel" and decoding based on the reverse order of the BC dirty paper encoding. In general the MIMO MAC capacity region is easy to compute, since it is a convex region. In fact, an iterative waterpouring formula was recently obtained that computes this region very efficiently [4]. However, the dirty paper coding achievable region is quite hard to compute since it is nonconvex [1]. Thus, while the dirty paper coding region for the MIMO BC cannot typically be computed directly for a general number of users and antennas, this region is easily computed by exploiting duality and the transformations between the MIMO MAC and this dirty paper coding region [3].

9. Sato Upper Bound and Sumrate Capacity

In the previous sections we showed that the dirty paper region of the BC and the union of the dual MAC capacity regions are equal. Now we show that dirty paper coding is the sumrate capacity achieving strategy for the MIMO BC. We do this based on Sato's upper bound for the capacity region of general BCs [5]. This bound utilizes the capacity of the *cooperative system* whereby the different receivers cooperate to decode the transmitted signal. Since the cooperative system is the same as the BC, but with receiver coordination, the capacity of the cooperative system ($\mathcal{C}_{\text{coop}}(P, \mathbf{H})$) is an upper bound on the BC capacity region ($\mathcal{C}_{\text{BC}}(P, \mathbf{H})$). We now show that the bound can be tightened by introducing noise correlation.

Since the capacity region of a general BC depends only on the marginal transition probabilities (i.e. $p(y_i|x)$) and not on the entire joint distribution $p(y_1, \ldots, y_K|x)$, we can introduce correlation between the noise vectors at different receivers of the BC without affecting the BC capacity region. This correlation does, however, affect the capacity of the cooperative system, which is still an upper bound on the sumrate of the BC. By searching over all feasible positive definite noise covariance matrices, we get the following bound for sum rate capacity, based on the capacity of the cooperative system with the worst case noise:

$$\mathcal{C}_{\text{BC}}^{\text{sumrate}}(P, \mathbf{H}) \leq \inf_{\Sigma_z} \max_{\Sigma_x} \frac{1}{2} \log |I + \Sigma_z^{-1/2} H \Sigma_x H^T \Sigma_z^{-1/2}|.$$

An explicit formula for the resulting worst case noise is found in [20]. By using the fact that the capacity region of the dual MIMO MAC equals the capacity region of the dirty paper region of the MIMO BC (and therefore the maximum sumrate of the MAC and the dirty paper region are equal), we are able to show using Lagrangian duality that this bound is tight for the MIMO BC. The proof that dirty paper coding achieves the Sato upper bound, and therefore equals the sumrate capacity of the MIMO BC, is depicted pictorially in Figure 6. Note that a special case of this proof for a single antenna at each receiver was obtained independently in [21]. This

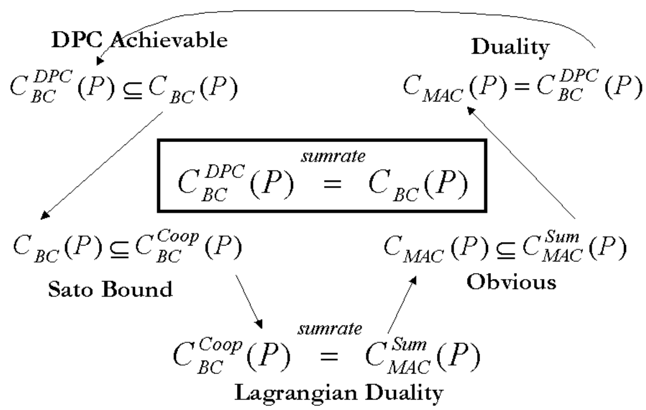

FIGURE 6. MIMO BC Sumrate Capacity Proof

proof is somewhat simpler since beamforming is optimal in this case, and therefore the covariance matrix transformations described in Section 8 are not needed.

Figure 7 shows the dirty paper region, Sato upper bound, and sumrate point where the two meet. We conjecture that dirty paper coding achieves the full capacity region of the MIMO BC and not just the sum rate point. We have obtained a proof of this result under the assumption that Gaussian inputs are optimal for the MIMO BC, but we have yet to prove this assumption. More details about this conjecture, our proof, and the one missing link can be found in [22]

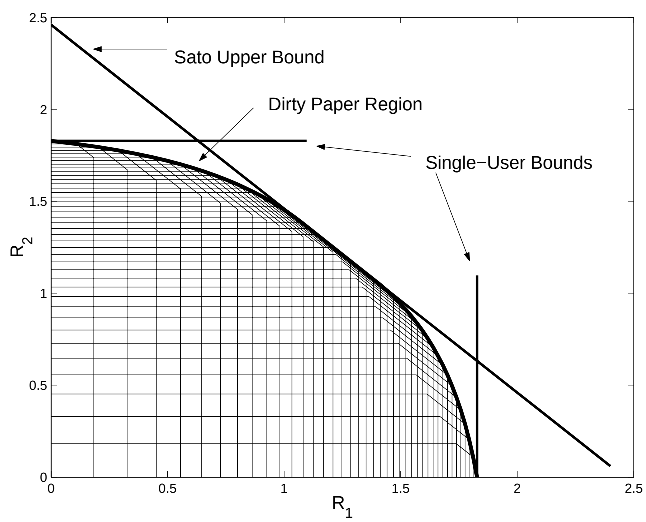

FIGURE 7. Dirty Paper Region and Sato Upper Bound for MIMO BC

10. Conclusions

Duality is a power technique to relate capacity regions for broadcast and MAC channels. Duality holds for many notions of capacity, including ergodic, outage, and minimum rate capacity. Duality can be used to show the connection between capacity regions and optimal transmission strategies for BC and MAC channels. More importantly, it can be used to obtain new capacity results previously thought intractable for the MAC or BC when results are known and/or easily computable for the dual channel. This concept is used to obtain the minimum rate capacity region of a fading MAC channel. We then turn our attention to the MIMO BC. We introduce the notion of dirty paper coding for this channel, which subtracts out the effect of known interference and thus achieves much higher rates than other techniques such as HDR. Unfortunately, the rates achievable using dirty paper coding are hard to compute. However, we show that a duality exists between the achievable rates of the MIMO BC under dirty paper coding and the capacity region of the dual MIMO MAC under a sum power constraint. Since this MAC region is easy to compute, we can obtain the rates of the MIMO BC under dirty paper coding quite easily using duality. Moreover, using duality and Sato's upper bound with joint detection and worst case noise, we show that dirty paper coding achieves the sumrate point of the MIMO BC capacity region. We conjecture that dirty paper coding achieves the entire region, but this proof requires Gaussian inputs to be optimal for this channel, which we have not yet shown.

References

[1] G. Caire and S. Shamai, "On Achievable Rates in a Multi-antenna Broadcast Downl ink", 38th Annual Allerton Conference on Commun., Control and Computing, Monticello, I L, Oct. 4 - 6, 2000.

[2] M. Costa, "Writing on Dirty Paper", IEEE. Trans. Inf. Theory, v. 29, pp. 439-4 41, May 1983.

[3] N. Jindal, S. Jafar, S. Vishwanath, and A. Goldsmith, "Sum power iterative water-filling for vector broadcast channels," *Proc. Asilomar Conf. on Signls., Syst., Compt.*, Oct. 2002.

[4] W. Yu, W. Rhee, S. Boyd, J. Cioffi, "Iterative Water-filling for Vector Multipl e Access Channels", pp. 322, Proc. IEEE Int. Symp. Inf. Theory, (ISIT), Washington DC, June 24-29, 2001.

[5] H. Sato, "An Outer Bound on the Capacity Region of Broadcast Channel", *IEEE. Trans. Inf. Theory*, v. 24, pp. 374-377, May 1978.

[6] G. Caire and S. Shamai, "On the Achievable Throughput of a Multi-antenna Gaussi an Broadcast Channel", Submitted to *IEEE Trans. on Inform. Theory*, July 2001.

[7] S. Verdú, "Multiple-access channels with memory with and without frame synchronism," *IEEE Trans. Info. Theory,* pp. 605-619, May 1989.

[8] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Info. Theory,* pp. 733-751, July 1989.

[9] N. Jindal, S. Vishwanath, and A. J. Goldsmith, "On the duality of multiple-access and broadcast channels," *Proc. IEEE Allerton Conf. Commun. Cont. Comput.*, Oct. 2001. Also submitted to *IEEE Trans. Inform. Theory.*

[10] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part I: Ergodic capacity", IEEE Trans. Inform. Theory, vol. 47, pp 1083-1102, March 2001.

[11] D.N. Tse and S. Hanly, "Multi-Access Fading Channels: Part I: Polymatroid Struct ure, Optimal Resource Allocation and Throughput Capacities", IEEE Trans. Inform. Theory, v. 44, pp 2796-2815, Nov. 1998.

[12] L. Li and A. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part II: Outage capacity", IEEE Trans. Inform. Theory, vol. 47, pp 1103-1127, March 2001.

[13] S.V. Hanly and D. N. Tse, "Multiaccess fading channels- Part II:Delay-limited capacities", IEEE Trans. Inform. Theory, vol. 44, pp. 2816-2831, Nov. 1998.

[14] L. Li and A. Goldsmith, "Outage capacities and optimal power allocation for fading multiple-access channels", in revision for IEEE Trans. Inform. Theory. See also *Proc. IEEE Intl. Symp. Inform. Theory,* pg. 305, June 2000.

[15] N. Jindal and A. Goldsmith, "Capacity and optimal power allocation for fading broadcast channels with minimum rates", *Proc. IEEE Globecom,* pp. 1292-1296, Nov. 2001. Also in revision for *IEEE Trans. Inform. Theory.*

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

[17] E. Telatar, "Capacity of Multi-antenna Gaussian Channels," European Trans. on Telecomm. ETT, 10(6):585-596, November 1999.

[18] W. Yu, W. Rhee, S. Boyd, J. Cioffi, "Iterative Water-filling for Vector Multipl e Access Channels", pp. 322, Proc. IEEE Int. Symp. Inf. Theory, (ISIT), Washington DC, June 24-29, 2001.

[19] W. Yu and J. Cioffi, "Trellis Precoding for the Broadcast Channel", Proc. IEEE Globecom, pp. 1344-1338, Nov., 2001.

[20] S. Vishwanath, N. Jindal, and A. J. Goldsmith, "On the capacity of multiple input multiple output broadcast channels," To appear: *Proc. IEEE Intl. Conf. Commun.*, April 2002. Also submitted to *IEEE Trans. Inform. Theory.*

[21] P. Viswanath and D. Tse, "Sum Capacity of the Multi-Antenna Broadcast Channel", *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 497, Lausanne, Switzerland, Jun. 30- Jul. 5, 2002. Also submitted to *IEEE Trans. on Inform. Theory.*

[22] S. Vishwanath, G. Kramer, S. Shamai, S. A. Jafar, and A. J. Goldsmith, "Capacity bounds for Gaussian vector broadcast channels," *Proc. DIMACS Workshop Signal. Proc. Wireless Trans.*, Oct. 7-9, 2002.

Department of Electrical Engineering, Mail Code 9515, Stanford University, Stanford, CA 94305-9515

E-mail address: {`njindal,sriram,syed,andrea`}`@wsl.stanford.edu`