On the Capacity of Cellular Systems with MIMO

Rick S. Blum, Jack H. Winters and Nelson R. Sollenberger
Wireless Systems Research Department, AT&T Labs - Research
Middletown, NJ 07748-4801

Abstract—**It is shown that the capacity of a single, isolated, multiple transmit and receive antenna array link, with given interference, is maximized by transmitting independent data streams from each antenna for a quasistatic and flat Rayleigh fading channel with independent fading coefficients for each path. However, if such links mutually interfere, in some cases the overall system capacity can be increased by transmitting fewer streams.**

I. Introduction

Transmit and receive antenna arrays used to form multiple-input multiple-output (MIMO) channels have shown great potential in isolated, single link communications without cochannel interference [1]. For a quasistatic and flat Rayleigh fading channel with independent fading, maximum ergodic capacity (average of mutual information) is achieved [1] by sending an independent information stream from each transmit antenna which is the maximum possible number of streams which can be sent. Very recent investigations have shown that cochannel interference can seriously degrade the overall capacity [2] when MIMO channels are used in a cellular system. Here we ask if it is always best to send the maximum possible number of independent information streams in order to achieve maximum capacity. In particular, we investigate the idea of adaptive MIMO, where the number of independent streams transmitted may be fewer than the maximum possible. Note, the number of streams is less than or equal to the number of transmit antennas. For example, we consider a single stream transmitted by multiple antennas in some cases.

Consider a quasistatic and flat Rayleigh fading channel with independent fading coefficients for each path. First we consider the capacity of a single, isolated link with cochannel interference. There are some cases of this type where it seems that it might be possible to achieve higher capacity by reducing the number of MIMO streams. For example, consider a case with n_t transmit antennas, n_r receive antennas and a single MIMO interferer with k streams. If n_t streams are transmitted, array processing theory implies that the interference can be nulled and reception of the data streams enhanced if $n_t + k \leq n_r$. This suggests that if $n_t + k = n_r + 1$ then it might be best to reduce the number of streams transmitted by at least one. We first show that this is not the case using an analytical proof. For simplicity, we focus on systems with two transmit antennas. Our analytical proof shows the capacity of a single, isolated link is not improved by reducing the number of streams transmitted from two to one. For generality, we actually demonstrate that it is always better to send the maximum number of streams possible as opposed to one stream for cases with any n_t. A numerical example illustrates our point.

On the other hand, if one user in a cellular system uses fewer MIMO streams, this will create fewer cochannel interferers for other users. In fact we show that the capacity of each user can be increased if the number of streams transmitted by each user is decreased, for certain signal-to-noise ratios (SNRs) and interference-to-noise ratios (INRs). We demonstrate this by showing that the capacity of a two stream user faced with a two stream interferer can be lower than the capacity of a one stream user faced with a one stream interferer at certain SNRs and INRs.

II. Model of MIMO Channel

First consider a single isolated link with a MIMO user and additive white Gaussian noise only (no interference). To simplify matters assume quasi-static flat Rayleigh fading. The vector of complex baseband samples from the set of n_r receive antennas after matched filtering is

$$\mathbf{y} = (y_1, \ldots, y_{n_r})^T = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where $\mathbf{x} = (x_1, \ldots, x_{n_t})^T$ is the transmitted vector, $\mathbf{H}$ is the channel matrix with independent entries that are each zero-mean complex Gaussian fading coefficients and $\mathbf{n} = (n_1, \ldots, n_{n_r})^T$ is the additive zero-mean complex Gaussian white noise with covariance matrix I_{n_r}. In such cases, it is known [1] that the optimum signaling is Gaussian, even if the noise is not white.

Now consider a cellular system where cochannel interference is present from L other users. In this case the vector of complex baseband samples after matched filtering

This paper is based on work supported by National Science Foundation under Grant No. MIP-9703730 and an ITR grant.

R. Blum was on a sabbatical leave from the EECS Dept., Lehigh University, Bethlehem, PA.

N. Sollenberger is now with Mobilink Telecom Inc., Middletown, NJ becomes

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \sum_{j=1}^{L} \mathbf{H}_j \mathbf{x}_j + \mathbf{n} \quad (2)$$

where $\mathbf{H}_j$ and $\mathbf{x}_j$ represent the channel matrix and the transmitted signal of user j respectively. For simplicity, we assume all of the interfering signals $\mathbf{x}_j, j = 1, \ldots, L$ are unknown to the receiver and we model each of them as being Gaussian distributed. Then if we condition on $\mathbf{H}, \mathbf{H}_1, \ldots, \mathbf{H}_L$, the interference-plus-noise from (2), $\sum_{j=1}^{L} \mathbf{H}_j \mathbf{x}_j + \mathbf{n}$, is Gaussian distributed with the covariance matrix $\mathbf{R} = \sum_{j=1}^{L} \mathbf{H}_j^H Cov(\mathbf{x}_j)\mathbf{H}_j + Cov(\mathbf{n})$ where $Cov(\mathbf{x}_j)$ denotes the covariance matrix of $\mathbf{x}_j$. The interference-plus-noise is whitened by multiplying $\mathbf{y}$ by $\mathbf{R}^{-1/2}$ after which the new version of (2) can be represented using (1) if we take $\tilde{\mathbf{H}} = \mathbf{R}^{-1/2}\mathbf{H}$ as the $\mathbf{H}$ which appears in (1), so the optimum signaling for $\mathbf{x}$ is Gaussian. In fact, the Gaussian signaling assumption is exactly the assumption made for the distribution of each $\mathbf{x}_j, j = 1, \ldots, L$ which is reassuring.

III. Isolated Channels

The similarity of (2) to (1) allows us to use existing results to calculate the capacity. First consider the case where it is not possible to send the required channel state information, $\tilde{\mathbf{H}}$, back to the transmitter to choose the capacity-optimizing signaling. Conditioned on $\tilde{\mathbf{H}}$, the capacity obtained using all streams possible $n_m = min(n_t, n_r)$, with no feedback, is [1]

$$C = \sum_{i=1}^{n_m} \log_2 (1 + \frac{\rho}{n_t} \lambda_i) \quad (3)$$

where $\lambda_1 \leq \lambda_2 \cdots \leq \lambda_{n_m}$ are the relevant eigenvalues of $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H$ and ρ is the total signal power from the transmit array. Since noise power is normalized, ρ is also the SNR [1].

First, we want to demonstrate that, for a given interference environment, in terms of ergodic capacity it is always best to send as many streams as possible. To illustrate this point in a simple way, we consider sending one stream as opposed to sending more streams. In particular, consider transmitting only one data stream from the transmit antenna array and then using optimum linear combining at the receiver. For the purpose of this demonstration, initially consider a case with feedback of channel state information to the transmitter, which allows an optimum transmit weight vector to be chosen. Let s denote a complex constellation symbol representing elements from the data stream to be transmitted and assume an optimum unit-length transmit weight vector $\mathbf{w_t}$ will be chosen so that $\mathbf{w_t}s$ is transmitted. The optimum unit-length combining weight vector $\mathbf{w_r}$ is used at the receiver so the capacity obtained using the single stream is

$$\begin{aligned} C &= max_{\mathbf{w_t},\mathbf{w_r}} \log_2 (1 + |\mathbf{w_r}^H \tilde{\mathbf{H}} \mathbf{w_t} s|^2) \\ &= max_{\mathbf{w_t}} \log_2 (1 + \left| \left(\frac{\tilde{\mathbf{H}}\mathbf{w_t}}{|\tilde{\mathbf{H}}\mathbf{w_t}|} \right)^H \tilde{\mathbf{H}}\mathbf{w_t} s \right|^2) \\ &= \log_2 (1 + \rho\, max_{\mathbf{w_t}} |\tilde{\mathbf{H}}\mathbf{w_t}|^2) \\ &= \log_2 (1 + \rho \lambda_{n_m}) \end{aligned} \quad (4)$$

In (4), the simplification to get to the second line is justified by the Cauchy-Schwartz inequality. The final simplifications follow from a well-known theorem of linear algebra.

In cases without feedback, it would not be possible to pick the transmit vectors based on $\tilde{\mathbf{H}}$. Instead, we pick fixed vectors with equal magnitude complex components. For example, in the case of one stream, we pick $\mathbf{w_{t1}} = (1/\sqrt{n_t}, \ldots, 1/\sqrt{n_t})^T$. Selecting the weight vectors in this way leads to the same ergodic capacity as in the expected value of (4) except that there is a "coherence" loss that reduces the SNR by at least a factor of n_t so that

$$E\{C\} \leq E\left\{ \log_2 (1 + \frac{\rho}{n_t} \lambda_{n_m}) \right\}. \quad (5)$$

The inequality in (5) follows from Jensen's inequality. By comparing the expected value of (3) with (5), it is easy to see that the expected value of (3) will always be as large as or larger than (5) since (5) is exactly equal to the expected value of the last term in (3) and the other terms in (3) are nonnegative. Thus it always makes sense to transmit the maximum number of streams provided the interference is fixed. These conclusions can also be verified numerically. Here we focus on a case with $n_r = n_t = 2$. Numerical results illustrating the superiority of using two streams are provided in the solid curves in Figure 1 which show the ratio of the two-stream capacity to the one-stream capacity versus SNR for a common (0, 10, 20, or 30 dB) INR due to a single stream interferer. Since the ratio is always larger than unity, transmitting two streams is always better for the same single stream interferer. From Figure 1, it is clear that the increased capacity from transmitting two streams gets smaller as INR increases as expected.

IV. MIMO Stream Control

A key issue that has not yet been addressed is that interference from one user will hurt another. Thus in fact, it may be better to have all users use fewer than the maximum number of possible streams in order to increase the

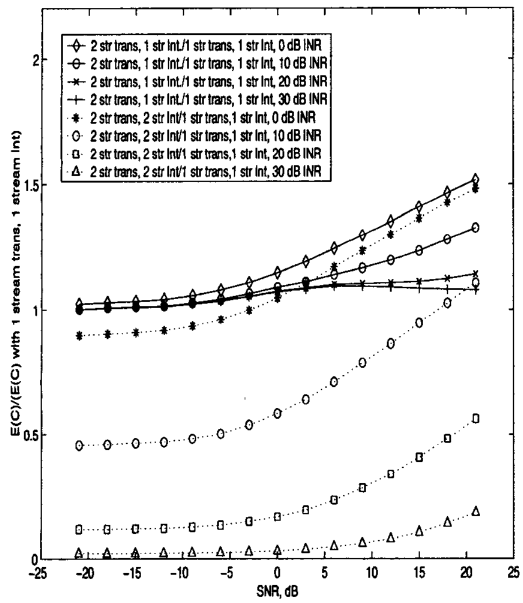

Fig. 1. Solid curves are ergodic capacity $E\{C\}$ for 2 transmitted streams divided by the ergodic capacity for 1 stream transmitted with the same 1 stream interference. The dashed curves are ergodic capacity for 1 transmitted stream and 1 interference stream divided by the ergodic capacity for 2 transmitted streams and 2 interference streams. All assume flat fading with either 0, 10, 20, or 30 dB INR.

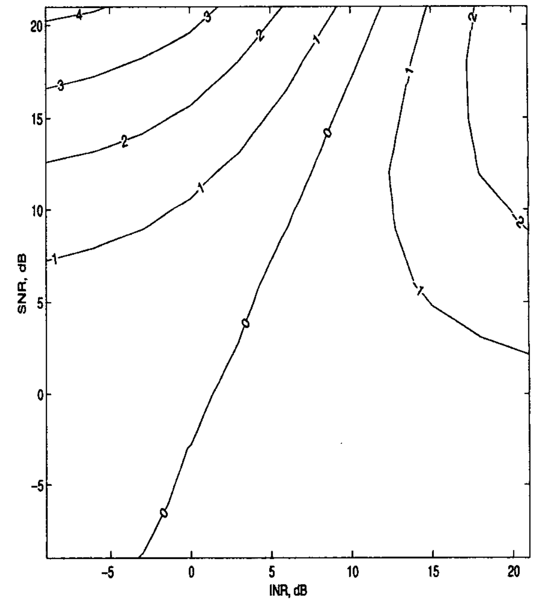

Fig. 2. Ergodic capacity (2 streams for desired and interference signals) - ergodic capacity (1 stream for desired and interference signals) versus SNR and INR.

capacity of each user. This is illustrated by the dashed curves in Figure 1 for the particular case of $n_t = n_r = 2$ where each user experiences interference from one other user who uses the same number of streams as they do. Here we find that if each user uses a single stream, the capacity is often higher than if each user uses the maximum possible number of streams, which is two in this case. Reducing the number of streams transmitted is similar to power control. Thus other users see more favorable interference environments when you control the number of streams you employ. In particular, using fewer streams may enable the other users to null extra interferers with the extra degrees of freedom. One might wonder exactly how important it is to select the correct number of streams. Figure 2 shows the difference in capacity using 2 streams versus 1 stream desired and interference signals as a function of SNR and INR using a contour plot. If SNR and INR can be estimated, such a plot could be used to select if one or two streams should be used by two users that are mutually interfering. The contour plot shows clearly the loss in capacity that would result by not making the selection correctly for a given SNR and INR. In fact, one can imagine more complicated schemes similar to the power control algorithms typically used in cellular systems that would control the number of streams used by each user. One observation is that for SNR and INR between 0 and 10 dB, which holds strong interest in cellular systems, there is a modest difference between the capacities of 1 and 2 stream systems.

V. Optimum Link Capacity with Feedback

Now consider the case where it is possible to send the required channel state information, $\mathbf{H}$, back to the transmitter to choose the capacity-optimizing signaling. Under the constraint of fixed total transmit power, capacity with feedback can be found using Lagrange multipliers [3], [4], [5] and a well-known water filling solution results. The water filling will choose the number of streams to transmit, which can be n_t at most. We call this link-optimum signaling. One can also consider a scheme constrained to use at most $n_s < n_t$ as the maximum number of streams that could be transmitted. The best approach of this type can be shown to follow the water filling solution but only transmitting the streams corresponding to the n_s largest eigenvalues. Of course, for the same interference environment, the link-optimum approach must be better than limiting the maximum number of streams since the link-optimum approach yields best performance for this case by construction. Viewed another way, if the extra streams are not helpful, the link-optimum approach can simply shut them off by giving them no power.

Consider a case with $n_t = n_r = 2$ and compare the link-optimum signaling approach just described to one with $n_s = 1$ which always sends one stream. The performance of these schemes is compared in the solid curves in Figure 3 for cases with fixed one stream interference. In Figure 3, all curves are normalized by the capacity achieved by the scheme which always transmits only one stream and sees an interferer which also always transmits only one stream. The solid curves in Figure 3 are always above unity, indicating that it is always better to use the link-optimum scheme. Since the two schemes compared see exactly the same interference environments, this follows directly from the arguments producing the link-optimum approach for this case. Notice that the optimum approach does not provide significant improvement for cases with weak signal strengths (SNRs below 0 dB) or very strong interference (INRs greater than 10 dB for the range of SNRs shown).

Now consider a system with two users which interfere with one another ($n_t = n_r = 2$ again). The dotted curves in Figure 3, which are defined similarly to those in Figure 1, illustrate that it is again sometimes useful to reduce the number of streams transmitted by each user to provide increased capacity for each user. Except for the curve for 0 dB INR for the case of $SNR > 0$ dB, all the other dotted curves are below unity. Thus for sufficiently weak signals or sufficiently large interference, limiting to one stream improves performance over the link-optimum scheme. The key is that we are comparing one scheme which is always transmitting 1 stream and always receiving 1 stream interference to another scheme which is sometimes transmitting 2 streams and sometimes receiving 2 stream interference. Note that the interference environment is different in the two schemes we compare.

The complimentary cumulative distribution functions (ccdfs) of the capacity are also of interest. Figure 4 illustrates the improvements that result in the capacity ccdf when receiving one stream desired and one stream interference as opposed to receiving two streams desired with two stream interference.

We note that for cases without feedback, generally our ccdf results (not shown) are similar to those in Figure 4. It is worth noting that, for cases without feedback, reasonably large improvements in the capacity ccdfs can be achieved by incorporating delay diversity [6]. On the other hand, the ergodic capacity curves, like those in Figure 1, show little difference regardless of whether delay

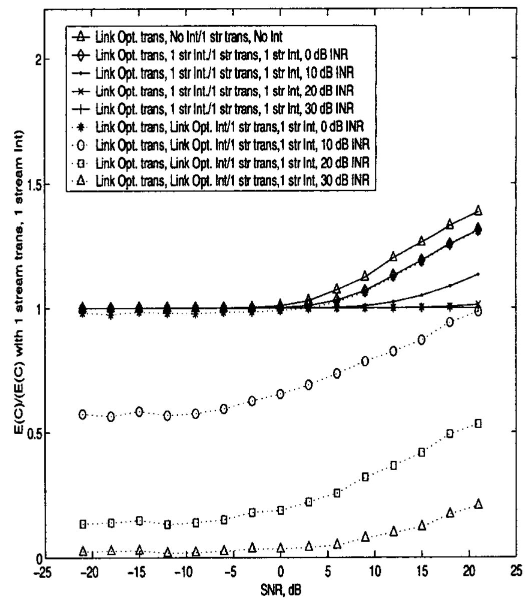

Fig. 3. Solid curves are ergodic capacity $E\{C\}$ for optimum transmission (2 streams or fewer) divided by the average capacity for 1 stream with the same 1 stream interference. The dashed curves are the optimum average capacity (2 transmitted streams or fewer) with an interferer of the same type (2 interference streams or fewer) divided by the average capacity for 1 transmitted stream and 1 interference stream. All assume flat fading with either 0, 10, 20, or 30 dB INR.

diversity is used. For example, the ergodic capacities for sending a single stream without delay diversity are barely distinguishable from the same curves for the case where delay diversity is employed.

VI. Cases with a Larger Number of Antennas

The general characteristics of the results for cases with $n_t > 2$, $n_r > 2$ and for a different number of streams transmitted are similar to what we have shown for $n_t = n_r = 2$. This is true both for cases with and without feedback. There are some minor differences which are easy to predict. For example, consider a case with $n_t = n_r = 4$ where feedback is available. Consider comparing the performance of a case where a maximum of four streams are transmitted to a case where one stream is transmitted as illustrated in Figure 5. The major differences between Figure 3 and Figure 5 are for the solid curves with larger INRs. These curves rise rapidly in Figure 5, while they are flat and close to unity in Figure 3. The results in Figure 3 are for a case where the maximum number of streams sent is either two or one when there is a one stream interferer. If a maximum of two streams can be

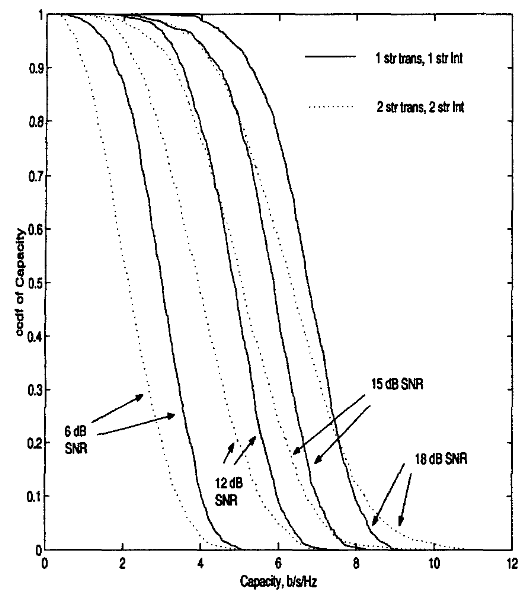

Fig. 4. Capacity ccdfs for various $SNRs$ with $INR = 10dB$ comparing two streams desired with two stream interference to one stream desired and one stream interference.

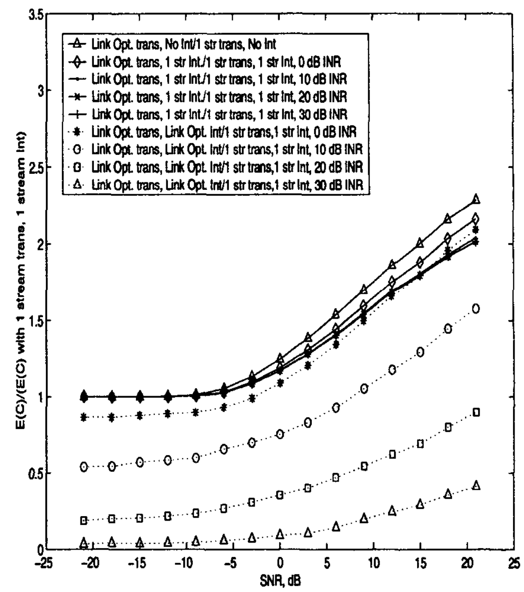

Fig. 5. Solid curves are ergodic capacity $E\{C\}$ with $n_t = n_r = 4$ for optimum transmission (4 streams or fewer) divided by the average capacity for 1 stream with the same 1 stream interference. The dashed curves are the optimum average capacity (4 transmitted streams or fewer) with an interferer of the same type (4 interference streams or fewer) divided by the average capacity for 1 transmitted stream and 1 interference stream. All assume flat fading with either 0, 10, 20, or 30 dB INR.

sent with a very strong one stream interferer, then only one stream will be sent by a link-optimum approach. Thus the ratio of the capacity with two and one streams is close to unity. In the case of Figure 5, the maximum number of streams that can be sent is either four or one. Thus, even when there is one very strong one stream interferer, in the case where a maximum of four streams can be transmitted, there are still three interference-free streams. This is the reason for the rapid increase in Figure 5. In fact, if we consider a case with $n_t = n_r = 4$ and two streams of interference and we compare a case where a maximum of two streams can be transmitted to a case where four streams can be transmitted, then the behavior for large INR is very similar to that in Figure 3.

VII. Conclusions

We have analyzed MIMO capacity with interference. We have introduced the interesting idea of stream control. It is clear that the results given can be easily extended to OFDM communication systems with nonflat fading [7], and an arbitrary number of cochannel users. We note that the results presented here hold only for infinite size constellations and infinitely complex coding since they are based on capacity. For a discussion of cases without coding, the interested reader is referred to [8].

References

[1] G. J. Foschini and M. J. Gans, "On Limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, Vol. 6, pp. 311-335, 1998.

[2] S. Catreux, L. J. Greenstein, and P. F. Dressen, "Simulation results for an interference-limited multiple-input multiple-output cellular System," *Communication Letters*, Dec. 2000.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian Channels," *European Transactions on Telecommunications*, 10(6): 585-595, Nov.-Dec. 1999.

[4] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Transactions on Communications*, COM-46, pp. 357-366, March 1998.

[5] T. M. Cover and J. A. Thomas, *Elements of Information Theory* John Wiley & Sons, New York, 1991 (especially pp. 250-256 and pp. 12-23).

[6] N. Seshadri and J. H. Winters, "Two signaling schemes for improving error performance of frequency division duplex (FDD) transmission systems using transmitter antenna," *International Journal of Wireless Information Networks*, Vol. 1, No. 1, 1994.

[7] R. S. Blum, Y. (Geoffrey) Li, J. H. Winters, and Q. Yan, "Improved Space-time Coding for MIMO-OFDM Wireless Communications," to appear in *IEEE Trans. on Communications.*

[8] R. W. Heath Jr. and A. Paulraj, "Switching between spatial multiplexing and transmit diversity based on constellation distance," Allerton Conf. on Comm. Cont. and Comp., Oct. 2000.