

# The Use of Diversity Antennas in High-Speed Wireless Systems: Capacity Gains, Fairness Issues, Multi-User Scheduling

Sem Borst, Phil Whiting  
Bell Laboratories, Lucent Technologies  
P.O. Box 636, Murray Hill, NJ 07974

## Abstract

The use of antenna arrays provides a natural mechanism for enhancing spectral efficiency so as to improve data rates, capacity, or coverage. There are various approaches for operating antenna-array systems, with diversity schemes belonging to the simpler ones to deploy. In the present paper, we specifically examine the capacity improvements from diversity antennas in multi-user scenarios. Although available single-user results yield valuable insight for voice communications, they do not directly translate to multi-user data systems.

First of all, we show that the use of diversity antennas has the effect of dampening the variations in the channel conditions. This reduces the scope for channel-aware scheduling mechanisms such as HDR to obtain capacity gains by exploiting fluctuations in the feasible transmission rates. Conversely, in the presence of scheduling, diversity antennas may produce substantially smaller gains, or even have a negative impact on capacity.

Second, in heterogeneous scenarios, we find that the capacity improvements from diversity antennas may widely vary, depending on what fairness notion is adopted. If the proportion of time slots is used as fairness criterion, then the gains in total throughput tend to be significant, but the edge users only receive a marginal share. In case the fairness measure is defined in terms of throughput ratios, then the gains tend to be limited population-wide.

## 1 Introduction

The use of wireless communications has experienced spectacular growth. While the proliferation of voice services is approaching a natural saturation point, the evolution of wireless data services has only recently started to develop. To accommodate future expansion, next-generation mobile communication systems are being designed to provide Internet access and support high-speed wireless data applications. The continual growth creates increasing pressure to squeeze the most out of the limited amount of wireless spectrum available.

The use of antenna arrays offers a potential technique for improving spectrum efficiency so as to achieve higher data rates, larger capacity, better coverage, or a combination of these. The increase in data rates is of vital importance for enabling high-bandwidth applications in wireless environments. In dense urban areas where cell splitting and sectorization may have reached practical limitations, the gain in capacity (support larger number of users) is particularly relevant. The increase in coverage (install smaller number of base stations) is especially attractive for green field service providers seeking to enter the market at an affordable capital investment.

There are a variety of approaches for operating antenna-array systems, showing a trade-off between complexity, efficiency, and feedback requirements. The various methods may be classified into two broad categories: (i) ‘diversity’ antennas, which are mostly static in nature [4], [16]; and (ii) ‘intelligent’ antennas, which are operated in an adaptive manner (beam-forming, beam-switching, selection diversity) [8], [11], [12], [15], [17]. The various schemes may further be categorized depending on whether they involve the use of multiple transmit antennas, multiple receive antennas, or a combination of both.

For performance reasons, it is desirable that there is enough physical separation within an antenna array to achieve independence across antennas. Obviously, it is more problematic to realize physical separation within an antenna array on a mobile handset than at a base station, although the size of a laptop computer may be adequate to achieve a reasonable degree of independence. Due to the sheer numbers, it is also more cost-effective to equip base stations with multiple antennas than handsets. Thus, on the downlink from base station to mobile, it is easier to have multiple transmit than receive antennas, whereas the opposite is true for the uplink. For data services, the downlink is commonly considered to be the bottleneck due to the traffic asymmetry.

The capacity improvements from diversity antennas in single-user communications have been extensively studied [4], [7], [16], [20]. Information-theoretic results in [4], [16] indicate that the capacity (i) saturates at some finite limit when only the number of transmit antennas is increased, (ii) grows logarithmically when only the number of receive antennas is increased, and (iii) grows roughly linearly when both the number of transmit and receive antennas are increased in proportion.

In the present paper, we investigate the capacity improvements from diversity antennas for downlink data traffic in multi-user scenarios. The capacity gains in multi-user scenarios seem intrinsically harder to quantify, although the single-user results may provide a reasonable estimate for voice services. For multi-user data systems however, it is less straightforward to extrapolate from the single-user gains for the following two reasons.

First of all, in contrast to voice services, there is generally no intrinsic rate requirement for data applications. This raises the issue what throughput shares should be allocated to the various users. Even when the throughput utility is similar for all users, it may not necessarily be optimal to allocate equal throughput shares, since this is not a zero-sum game for spatially heterogeneous users. Weak users, with a low signal-to-noise ratio, require significantly more resources (transmission time, or transmit power) per bit than strong users, with a high signal-to-noise ratio. As a result, the overall throughput strongly depends on the relative throughput shares. In particular, there is a fundamental trade-off between maximizing the total throughput and balancing the individual throughput proportions.

As a result, the capacity improvements from diversity antennas may vary dramatically, depending on what fairness notion is adopted. If the proportion of time slots is used as a fairness criterion, then the gains in total throughput tend to be considerable. However, edge users only receive a negligible portion of the increase in throughput, illustrating the trade-off mentioned above. If the fairness measure is defined in terms of throughput shares, then the gains tend to be modest population-wide.

Second, the relative delay tolerance of data applications opens up the possibility for scheduling downlink transmissions in multi-user scenarios so as to optimize throughput. A particularly attractive approach, in fading environments, is to exploit multi-user diversity by tracking the channel fluctuations, and scheduling transmissions to users when they enjoy (nearly) optimal channel conditions, such as in HDR [2], [3], [6], [13], [14], [18]. Although this scheduling approach also yields throughput gains when combined with diversity antennas, these gains tend to be less substantial than for conventional single-antenna systems. In order to explain this phenomenon, we will show that the use of diversity antennas not only raises the expected feasible rate for a given individual user, but also has the effect of smoothing the stochastic variations. The increase in expected feasible rate directly translates into a higher throughput in a single-user scenario. However, with multi-user diversity, avoiding the downward dips in the transmission rate is less critical, while reducing the upward fluctuations negatively interferes with channel-aware scheduling mechanisms. In that sense, the use of diversity antennas reduces the potential for capacity gains from channel-aware scheduling mechanisms. Conversely, in the presence of scheduling, diversity antennas may yield smaller gains, or even have a negative impact on capacity, especially at low signal-to-noise ratios, exacerbating the fairness issues described above. Similar findings have been reported in [5], [19].

The remainder of the paper is organized as follows. In Section 2, we review some key results for single-user communications. We also perform some numerical experiments to confirm that diversity antennas may produce substantial capacity gains, in particular at high signal-to-

noise ratios. In Section 3, we further elaborate on the single-user results, and provide a formal proof of the fact that adding transmit or receive antennas increases the capacity. Although this property is highly plausible from a physical perspective, the mathematical proof helps identify the effects that contribute to the capacity gains. This insight is particularly relevant in understanding why these gains may diminish or even completely vanish in multi-user scenarios. In Section 4, we focus on a scenario with several stochastically identical users. We demonstrate that adding transmit antennas may produce less significant gains, or even have a negative effect on capacity, especially at relatively low signal-to-noise ratios. In Section 5, we turn to a scenario with heterogeneous users. In Section 6, we make some concluding remarks.

## 2 Preliminary results

In this section we review some key results for single-user communications which will play a crucial role in the later sections. We consider a scenario with  $t$  transmit antennas and  $r$  receive antennas. Denote by  $H$  an  $r \times t$  complex random matrix representing the channel gains between the various transmit and receive antennas. We assume that each entry is a zero-mean Gaussian random variable with independent real and imaginary parts, each with variance  $1/2$ . Or equivalently, each entry has a uniformly distributed phase and Rayleigh distributed magnitude, with expected magnitude square equal to unity. We assume that all entries are independent. The above assumptions model a Rayleigh fading channel, with enough separation within the transmit and receive antennas to achieve independence. The realization of the matrix  $H$  is assumed known at the receiver but not at the transmitter.

Telatar [16] shows that the capacity of the above channel is given by

$$C(r, t, P) = \mathbb{E}[\log \det(I_r + (P/t)HH^\dagger)] = \mathbb{E}[\log \det(I_t + (P/t)H^\dagger H)], \quad (1)$$

with  $I_r$  denoting the  $r$ -dimensional unity matrix,  $H^\dagger$  indicating the transpose conjugate of  $H$ , and  $P$  representing the mean signal-to-noise ratio (or mean received power at a single antenna, with noise normalized to unity), see also Remark 2.1 below. Observe from equation (1) that  $C(r, t, P) = C(t, r, rP/t)$ , which shows a strong symmetry between receive and transmit antennas, except for the increase in power obtained from multiple receive antennas.

The capacity may alternatively be represented as

$$C(r, t, P) = \sum_{k=1}^p \mathbb{E}[\log(1 + P\lambda_k/t)], \quad (2)$$

with  $p = \min\{r, t\}$  and  $\lambda_1, \dots, \lambda_p$  denoting the eigenvalues of the matrix

$$W = \begin{cases} HH^\dagger & r < t \\ H^\dagger H & r \geq t \end{cases}$$

Note that the eigenvalues  $\lambda_1, \dots, \lambda_p$  are real and non-negative, since the matrix  $W$  is non-negative definite.

**Remark 2.1** *The channel capacity as given in the above two formulas provides an information-theoretic upper bound on the transmission rate. In practice, the coding interval must be sufficiently long for the actual transmission rate to approach the channel capacity. In addition, the above two formulas are based on the assumption that the interference may be treated as additive white Gaussian noise, which is appropriate when thermal noise is the dominant source of impairment, or in a scenario with a large population of weak interferers. Extensions to a scenario with a small number of strong interferers may be found in [10].*

For fixed  $r$ , note that  $\frac{1}{t}HH^\dagger \rightarrow I_r$  almost surely as  $t \rightarrow \infty$ , so that

$$C(r, t, P) \rightarrow r \log(1 + P). \quad (3)$$

Thus, for a fixed number of receive antennas, the capacity saturates at some finite limit as the number of transmit antennas is increased.

In case of a single receive antenna ( $r = 1$ ), formulas (1) and (2) reduce to

$$C(1, t, P) = \mathbb{E}[\log(1 + (P/t) \sum_{s=1}^t |h_{kk}|^2)] = \mathbb{E}[\log(1 + PErl_t/t)] = \frac{1}{(t-1)!} \int_0^\infty \log(1 + Pu/t) e^{-u} u^{t-1} du \quad (4)$$

with  $Erl_k$  representing an Erlang distributed random variable with  $k$  phases, each of unit mean. Note that  $Erl_t/t \rightarrow 1$  almost surely as  $t \rightarrow \infty$ , so that

$$C(1, t, P) \rightarrow \log(1 + P), \quad (5)$$

as a special case of the limit result stated above.

Similarly, in case of a single transmit antenna ( $t = 1$ ), formulas (1) and (2) collapse to

$$C(r, 1, P) = \mathbb{E}[\log(1 + P \sum_{s=1}^r |h_{kk}|^2)] = \mathbb{E}[\log(1 + PErl_r)] = \frac{1}{(r-1)!} \int_0^\infty \log(1 + Pu) e^{-u} u^{r-1} du. \quad (6)$$

Since  $Erl_r/r \rightarrow 1$  almost surely as  $r \rightarrow \infty$ , it follows that

$$C(r, 1, P) \approx \log(1 + Pr) \quad (7)$$

as  $r \rightarrow \infty$ . Thus, for a single transmit antenna, the capacity grows logarithmically as the number of receive antennas is increased.

In the general case where  $r > 1$ ,  $t > 1$  the evaluation of the capacity gets more involved. Telatar [16] shows that (1) may then be computed as

$$C(r, t, P) = \int_0^\infty \log(1 + Pu/t) \sum_{k=0}^{p-1} \frac{k!}{(k+n-p)!} [L_k^{n-p}(u)]^2 u^{n-p} du, \quad (8)$$

with  $p = \min\{r, t\}$ ,  $n = \max\{r, t\}$ , and  $L_k^{n-p}(x) = \frac{1}{k!} e^x x^{p-n} \frac{d^k}{dx^k} \{e^{-x} x^{n-p+k}\}$  denoting the associated Laguerre polynomial of order  $k$ .

Observe that in the special cases  $r = 1$  and  $t = 1$ , the above formula indeed reduces to formulas (4) and (6), respectively.

In the case where  $r = t = s$ , the above formula may be simplified somewhat to give

$$C(s, s, P) = \int_0^\infty \log(1 + Pu/s) \sum_{k=0}^{s-1} L_k(u)^2 e^{-u} du, \quad (9)$$

with  $L_k(x) = L_k^0(x) = \frac{1}{k!} e^x \frac{d^k}{dx^k} \{e^{-x} x^k\}$  denoting the Laguerre polynomial of order  $k$ .

Using the change of variables  $y = u/s$ , the above formula may be rewritten as

$$C(s, s, P) = s \int_0^\infty \log(1 + Py) \sum_{k=0}^{s-1} L_k(sy)^2 e^{-sy} dy.$$

It turns out [16] that as  $s \rightarrow \infty$  the term  $\sum_{k=0}^{s-1} L_k(sy)^2 e^{-sy}$  converges to a function

$$F(y) = \begin{cases} \pi^{-1} \sqrt{y^{-1} - 1/4}, & 0 < y < 4 \\ 0 & y > 4 \end{cases}$$

so that

$$C(s, s, P) \sim s \int_0^4 \log(1 + Py) F(y) dy, \quad (10)$$

as  $s \rightarrow \infty$ . Thus, the capacity grows linearly when both the number of transmit and receive antennas is increased.

Summarizing, we observe from formulas (5), (7), and (10) that the capacity (i) approaches some finite limit when only the number of transmit antennas is increased; (ii) grows slowly but without bound when only the number of receive antennas is increased; and (iii) grows linearly when both the number of transmit and receive antennas is increased.

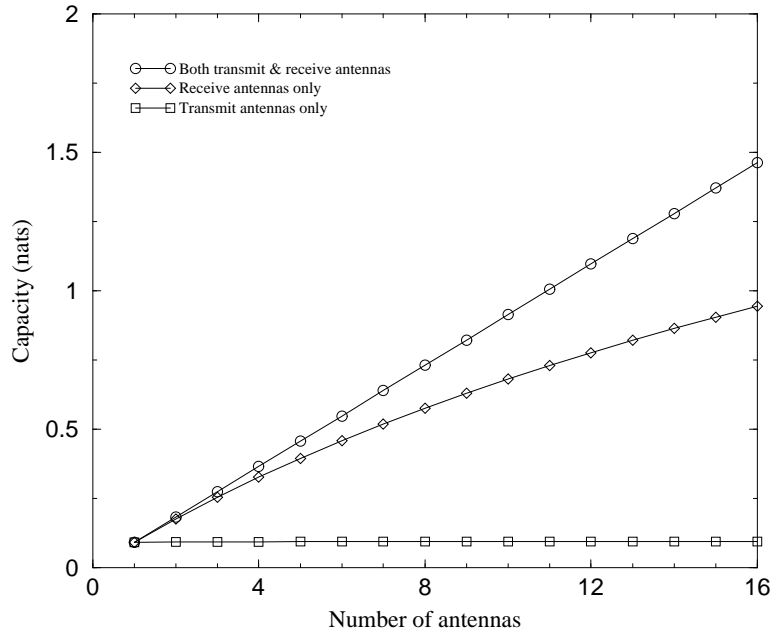


Figure 1: Single-user capacity as function of number of antennas at  $P = -10.0\text{dB}$ .

In order to illustrate the above observations, we conducted some numerical experiments comparing the behavior of formulas (4), (6), and (9) as a function of the number of antennas. The results are displayed in Figures 1-3 for various values of the signal-to-noise ratio  $P$ .

In all cases, the three curves inevitably diverge as the number of antennas grows large, as anticipated from the above observations. However, for a smaller number of antennas, the differences may be quite modest, depending on the value of the signal-to-noise ratio. In particular, at low values of the signal-to-noise ratio, the gains from using multiple transmit antennas in addition to multiple receive antennas tend to be marginal. This is further witnessed by Figure 4 where the capacity is graphed as function of the signal-to-noise ratio for a fixed number of antennas. In the next section, we will examine what physical effects contribute to the gains so as to provide an explanation of this phenomenon.

### 3 Single-user scenario

In this section we further elaborate on some of the results from the previous section for single-user communications. The next proposition shows that adding transmit or receive antennas increases the capacity. Although this property is highly plausible from a physical perspective, and may be implicitly deduced from the derivation in [16], it may not be directly evident

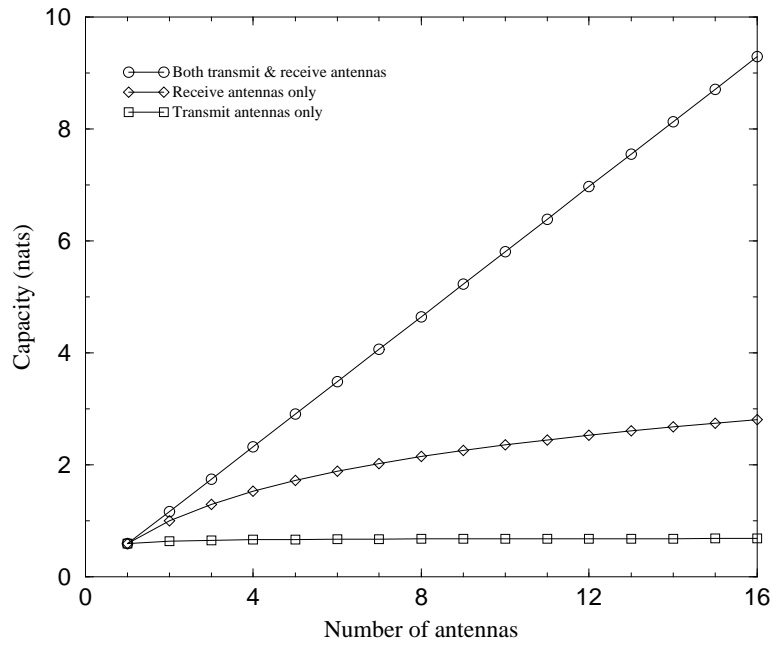


Figure 2: Single-user capacity as function of number of antennas at  $P = 0.0\text{dB}$ .

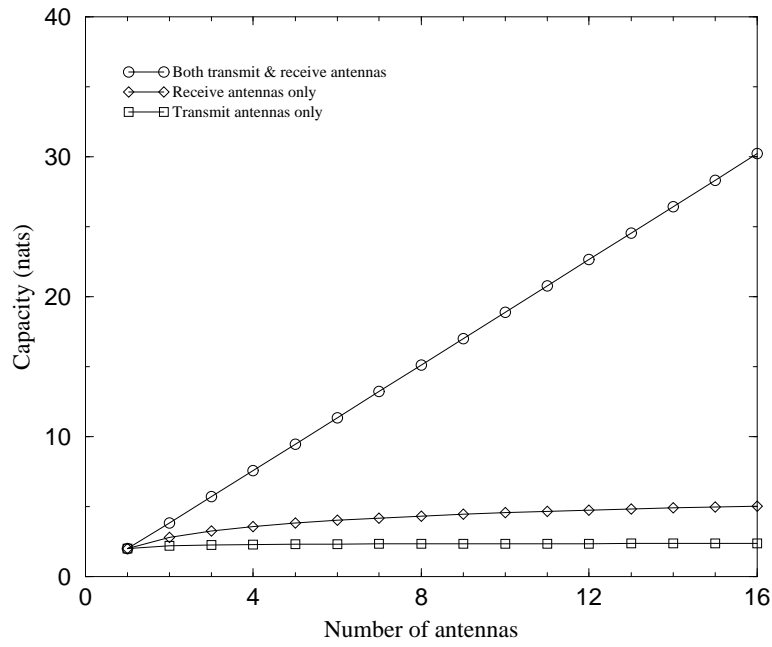


Figure 3: Single-user capacity as function of number of antennas at  $P = 10.0\text{dB}$ .

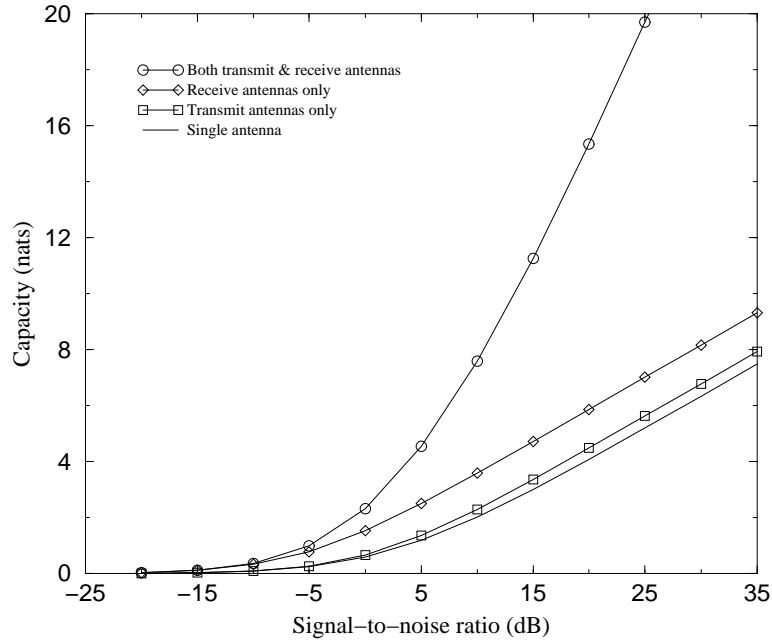


Figure 4: Single-user capacity as function of signal-to-noise ratio for 4 antennas.

from formula (1), which is why we believe a proof is useful. Besides, the proof helps identify the physical effects that contribute to the gains from using multiple antennas. This insight is particularly relevant in understanding why these gains may diminish or even completely vanish in multi-user scenarios as will turn out in later sections.

**Proposition 3.1** *For all  $r \geq q$ ,  $t \geq s$ ,  $C(r, t, P) \geq C(q, s, P)$ .*

### Proof

The statement of the proposition is equivalent to the following two statements: (i) the capacity is increasing in the number of transmit antennas, i.e.,  $C(r, t, P) \geq C(r, s, P)$  for all  $t \geq s \geq 1$ ; (ii) the capacity is increasing in the number of receive antennas, i.e.,  $C(r, t, P) \geq C(q, t, P)$  for all  $r \geq q \geq 1$ .

(i) For clarity of exposition, we first consider the special case  $s = 1$ . For a given  $r \times t$  matrix  $H$ , denote by  $H_k$  the  $r \times 1$  vector consisting of the elements in the  $k$ -th column of  $H$ ,  $k = 1, \dots, t$ . The vector  $H_k$  represents the channel gains between the  $k$ -th transmit antenna and the various receive antennas. By definition, the  $(i, j)$ -th element of the matrix  $HH^\dagger$  is given by  $\sum_{k=1}^t h_{ik} \bar{h}_{jk}$ ,

which means  $HH^\dagger$  may be written as

$$HH^\dagger = \sum_{k=1}^t H_k H_k^\dagger.$$

Telatar [16] proves that the function  $\log \det$  is concave on the set of non-negative definite matrices, which implies that

$$\log \det\left(\sum_{k=1}^K A_k\right) \geq \sum_{k=1}^K \log \det(A_k)$$

for non-negative definite matrices  $A_1, \dots, A_K$ .

In particular, choosing  $K = t$ ,  $A_k = I_r + PH_k H_k^\dagger$ , we obtain

$$\begin{aligned} \log \det(I_r + (P/t)HH^\dagger) &= \log \det\left(I_r + (P/t) \sum_{k=1}^t H_k H_k^\dagger\right) = \log \det\left(1/t \sum_{k=1}^t (I_r + PH_k H_k^\dagger)\right) \geq \\ &1/t \sum_{k=1}^t \log \det(I_r + PH_k H_k^\dagger). \end{aligned}$$

Taking expectations then yields  $C(r, t, P) \geq C(r, 1, P)$  for all  $t \geq 1$ .

We now extend the above proof to the general case  $t \geq s \geq 1$ .

Define  $\Omega_s$  as the collection of all subsets of  $\{1, \dots, t\}$  of size  $s$ . For any subset  $S \in \Omega_s$ , denote by  $H_S$  the  $r \times s$  matrix consisting of the columns of  $H$  indexed by the elements of  $S$ . As before, the matrix  $H_S H_S^\dagger$  may be written as

$$H_S H_S^\dagger = \sum_{k \in S} H_k H_k^\dagger.$$

Now observe that for any collection of matrices  $A_1, \dots, A_t$ , we have the identity relation

$$\sum_{k=1}^t A_k = \left[ \binom{t-1}{s-1} \right]^{-1} \sum_{S \in \Omega_s} \sum_{k \in S} A_k.$$

In particular, taking  $A_k = H_k H_k^\dagger$ , we may write

$$HH^\dagger = \left[ \binom{t-1}{s-1} \right]^{-1} \sum_{S \in \Omega_s} H_S H_S^\dagger.$$

Using the fact that the function  $\log \det$  is concave on the set of non-negative definite matrices, we obtain

$$\log \det(I_r + (P/t)HH^\dagger) = \log \det\left(I_r + (P/t) \left[ \binom{t-1}{s-1} \right]^{-1} \sum_{S \in \Omega_s} H_S H_S^\dagger\right) =$$

$$\begin{aligned} & \log \det \left( \left[ \begin{pmatrix} t \\ s \end{pmatrix} \right]^{-1} \sum_{S \in \Omega_s} I_r + (s/t) \left[ \begin{pmatrix} t-1 \\ s-1 \end{pmatrix} \right]^{-1} \sum_{S \in \Omega_s} (P/s) H_S H_S^\dagger \right) = \\ & \log \det \left( \left[ \begin{pmatrix} t \\ s \end{pmatrix} \right]^{-1} \sum_{S \in \Omega_s} (I_r + (P/s) H_S H_S^\dagger) \right) \geq \left[ \begin{pmatrix} t \\ s \end{pmatrix} \right]^{-1} \sum_{S \in \Omega_s} \log \det (I_r + (P/s) H_S H_S^\dagger). \end{aligned}$$

Taking expectations then yields  $C(r, t, P) \geq C(r, s, P)$  for all  $t \geq s \geq 1$ .

(ii) The statement follows from the following two statements: (i) for fixed total power  $P$ , the capacity is increasing in the number of receive antennas, i.e.,  $C(r, t, P) \geq C(q, t, rP/q)$  for all  $r \geq q \geq 1$ . (ii) the capacity is increasing in the total power i.e.,  $C(r, t, rP/q) \geq C(q, t, P)$  for all  $r \geq q \geq 1$ .

The first statement is equivalent to the earlier statement that the capacity is increasing in the number of *transmit* antennas, after recalling the symmetry between receive and transmit antennas:  $C(r, t, P) = C(t, r, rP/t) \geq C(t, q, rP/t) = C(q, t, rP/q)$ .

The second statement follows directly from the expression (2) for the capacity in terms of the eigenvalues of the matrix  $W$  which is increasing in  $P$ .

□

Inspection of the above proof shows that there are two physical effects that contribute to the gains from using multiple antennas: (i) it is advantageous to split the power across multiple antennas, due to the concavity of the log det function; (ii) in addition, it is beneficial to increase the total power, which is achieved in the case of multiple receive antennas. In particular, the proof reveals that adding receive antennas would continue to increase the capacity even if the transmit power were reduced in proportion.

In fact, adding receive antennas increases the capacity not just in expectation, but even in a stronger stochastic sense, which will prove to be important later. Specifically, let  $Z(r, t, P)$  be a random variable distributed as  $\log \det(I_r + (P/t) H H^\dagger) = \log \det(I_t + (P/t) H^\dagger H)$ , with  $H$  representing the  $r \times t$  channel matrix as before, so that  $C(r, t, P) = \mathbb{E}Z(r, t, P)$ . Using the fact that the function log det is also increasing on the set of non-negative definite matrices, the above proof may then be extended to show that if  $r \geq q$ , then  $Z(r, t, P) \geq_{\text{st}} Z(q, t, P)$ , with  $\geq_{\text{st}}$  denoting stochastic dominance, i.e.,  $\mathbb{P}\{Z(r, t, P) > z\} \geq \mathbb{P}\{Z(q, t, P) > z\}$  for all  $z \geq 0$ .

The proof further indicates that the gains from splitting the power will be smaller at lower values of the power (or signal-to-noise ratio), since the degree of concavity is then less pronounced. In that regime, the capacity should be roughly linear in the total power, and thus linear in the number of receive antennas, and largely insensitive to the number of transmit antennas, as reflected earlier in Figure 1.

This may be formally verified as follows. Using the representation (2) and the fact that  $\log(1+x) = x + o(x^2)$  for  $x$  near 0, we obtain

$$C_{r,t}/P \rightarrow 1/t \sum_{k=1}^m \lambda_k = \frac{1}{t} \mathbb{E}[\text{trace}(W)/t] = \frac{1}{t} \mathbb{E}[\sum_{i=1}^r \sum_{j=1}^t |h_{ij}|^2] = r$$

as  $P \downarrow 0$ , suggesting that for small values of  $P$ , we have  $C_{r,t} \approx rP$ . In fact, one may prove convergence in distribution,  $\log \det(I_r + (P/t)HH^\dagger)/P \rightarrow Er l_{rt}$  as  $P \downarrow 0$ .

## 4 Homogeneous users

The results in the previous two sections showed that diversity antennas may produce substantial capacity gains in a single-user scenario, especially at high signal-to-noise ratios. In the next few sections we examine the potential gains in multi-user scenarios, which may significantly diminish or even completely vanish as will turn out.

We consider a single base station equipped with  $t$  transmit antennas serving  $M$  data users, each with  $r$  receive antennas. The base station transmits in slots of some fixed duration. In each slot, the base station transmits to exactly one of the users.

We assume that the channel gain matrices for the various users vary over time according to some stochastic process  $\{H_1(n), \dots, H_M(n), n = 1, 2, \dots\}$ , with  $H_m(n)$  representing the  $r \times t$  complex matrix of channel gains for user  $m$  in the  $n$ -th slot. The matrix  $H_m(n)$  determines the maximum rate for user  $m$  in the  $n$ -th slot as described in Section 2,

$$R_m(n) = \log \det(I_r + (P_m/t)H_m(n)H_m^\dagger(n)) = \log \det(I_t + (P_m/t)H_m(n)^\dagger H_m(n)), \quad (11)$$

with  $P_m$  representing the mean signal-to-noise ratio for user  $m$ . Let  $H_1, \dots, H_M$  be random matrices with as distribution the stationary distribution of  $H_1(n), \dots, H_M(n)$ . The base station is supposed to have perfect knowledge of the feasible rate  $R_m(n)$  for user  $m$  at the start of the  $n$ -th slot. In practice, the feasible rate can only be estimated using feedback bits from the mobile, or by inferring the downlink conditions from the uplink statistics. Most of the results below however remain valid if  $R_m(n)$  is interpreted as the *expected* feasible rate, which also allows for modeling transmission errors.

We assume that in each slot the channel gain matrices for the various users are independent and identically distributed, with the same distribution as described in Section 2. Thus, each matrix entry is a zero-mean Gaussian random variable with independent real and imaginary parts, each with variance 1/2. Or equivalently, each matrix entry has a uniformly distributed phase and Rayleigh distributed magnitude, with expected magnitude square equal to unity. In addition, all entries are independent.

**Remark 4.1** *The variations in the channel gain matrices are assumed to capture the fast fading due to multi-path propagation effects. The impact of distance-related path loss and slow fading (shadow fading due to obstacles) is supposed to be reflected in the mean signal-to-noise ratio  $P$ . We assume that the frame length is relatively short, so that the channel conditions remain (nearly) constant over the duration of a frame, and can be reasonably well predicted. Since the coding interval is limited by the frame length, the duration must however be sufficiently long to achieve a transmission rate close to the channel capacity, see also Remark 2.1. In the remainder of the paper, we will use the information-theoretic capacity as given by formula (11) as a proxy for the feasible rate without further mention. Depending on the exact operating conditions, the actual feasible transmission rate may somewhat vary.*

In this section, we focus on a scenario with homogeneous users which all have equal signal-to-noise ratios  $P_1 = \dots = P_M = P$ . In the next section, we will study a scenario with heterogeneous users. We assume that in each slot the base station transmits to the user with the currently maximum feasible rate, i.e., in the  $n$ -th slot the base station selects the user identified as

$$m^*(n) = \arg \max_{m=1, \dots, M} R_m(n).$$

Note that this strategy maximizes the throughput in each slot, and hence the total long-term throughput. By symmetry, this strategy also equalizes the long-term average throughput for all users. The total capacity is given by

$$C(r, t, P, M) = \mathbb{E}[R | R = \max_{m=1, \dots, M} R_m], \quad (12)$$

with  $R_1, \dots, R_M$  independent and identically distributed random variables,

$$R_m = \log \det(I_r + (P/t)H_m H_m^\dagger) = \log \det(I_t + (P/t)H_m^\dagger H_m),$$

with  $H_1, \dots, H_M$  the random matrices defined earlier. Thus, the capacity may be evaluated as

$$\begin{aligned} C(r, t, P, M) &= \int_{z=0}^{\infty} \mathbb{P}\{R > z | R = \max_{m=1, \dots, M} R_m\} dz \\ &= \int_{z=0}^{\infty} [1 - \mathbb{P}\{R \leq z | R = \max_{m=1, \dots, M} R_m\}] dz \\ &= \int_{z=0}^{\infty} [1 - \mathbb{P}\{R_1, \dots, R_M \leq z\}] dz \end{aligned}$$

$$= \int_{z=0}^{\infty} [1 - (\mathbb{P}\{Z(r, t, P) \leq z\})^M] dz,$$

with  $Z(r, t, P)$  a random variable distributed as  $\log \det(I_r + (P/t)HH^\dagger) = \log \det(I_t + (P/t)H^\dagger H)$ , with  $H$  representing the  $r \times t$  channel matrix.

In order to evaluate the capacity, we first consider the case of just a single receive antenna ( $r = 1$ ). As in (4), we have  $Z(1, t, P) = \log(1 + PErl_t/t)$ , with  $Erl_t$  an Erlang distributed random variable with  $t$  phases, each of unit mean, so that

$$\mathbb{P}\{Z(1, t, P) \leq \log(1 + Pu/t)\} = \mathbb{P}\{Erl_t \leq u\} = 1 - \sum_{k=0}^{t-1} e^{-u} \frac{u^k}{k!}.$$

Making the substitution  $z = \log(1 + Pu/t)$  and using partial integration, we find

$$\begin{aligned} C(1, t, P, M) &= \int_{u=0}^{\infty} \frac{P/t}{1 + Pu/t} [1 - (1 - \sum_{k=0}^{t-1} e^{-u} \frac{u^k}{k!})^M] du \\ &= \frac{M}{(t-1)!} \int_{u=0}^{\infty} \log(1 + Pu/t) \left(1 - \sum_{k=0}^{t-1} e^{-u} \frac{u^k}{k!}\right)^{M-1} e^{-u} u^{t-1} du. \end{aligned} \quad (13)$$

Similarly, in the case of just a single transmit antenna ( $t = 1$ ), we obtain

$$C(r, 1, P, M) = \frac{M}{(r-1)!} \int_{u=0}^{\infty} \log(1 + Pu) \left(1 - \sum_{k=0}^{t-1} e^{-u} \frac{u^k}{k!}\right)^{M-1} e^{-u} u^{t-1} du. \quad (14)$$

The evaluation of the capacity in the general case where  $r > 1$  and  $t > 1$  appears rather complicated. In Appendix A we describe how the capacity may be evaluated for the case of two receive and two transmit antennas. Unfortunately, the approach does not seem to readily extend to larger values of  $r$  and  $t$ .

We performed some numerical experiments to compare the behavior of formulas (12), (13), and (14) as a function of the number of antennas for a scenario with  $M = 4$  identical users. To evaluate formula (12) we generated random matrices using Monte Carlo simulation.

We also compared the capacity results with the corresponding numbers when the transmission slots are not dynamically allocated to the user with the maximum current feasible rate, but simply divided among the users in a static fashion. Note that in the latter case the total capacity is identical to the single-user capacity as considered before.

The results are graphed in Figures 5-7 for various values of the signal-to-noise ratio  $P$ . The solid lines correspond to the dynamic scheduling strategy which always assigns the transmission slot

to the user with the currently maximum feasible rate. The dashed lines refer to round robin strategy which simply allocates equal fractions of time slots to all users in a static fashion. In all cases, the scheduling strategy achieves higher capacity than the round robin scheme. This is consistent with the earlier observation that scheduling in fact maximizes the throughput in each slot, and hence the total long-term throughput. However, the throughput gains from scheduling are not spectacular, and seem additive in nature rather than multiplicative. In particular, the gains seem to diminish as the number of transmit antennas increases.

This phenomenon may be explained by recalling that increasing the number of transmit antennas amounts to splitting the power. Spreading the power raises the expected feasible rate for a given individual user due to the concavity of the log det function. However, averaging the power has the effect of reducing the stochastic variations that channel-aware scheduling mechanisms aim to exploit. As a result, adding transmit antennas reduces the potential for capacity gains from multi-user scheduling. (Note that increasing the number of receive antennas has a similar effect of reducing the (relative) fluctuations, but in that case the negative impact is off-set by the larger improvement in the expected rate due the increase in total received power. In fact, we observed before that adding receive antennas increases the feasible rate in a stochastic sense, so that it continues to improve throughput in case of multi-user scheduling.) In particular, the gains from scheduling may be larger in situations with multiple receive antennas only than with both transmit and receive antennas. In the previous section, we found that the additional gains from transmit antennas over using receive antennas only could be marginal for certain parameter regimes, for example a relatively small number of antennas and low values of the signal-to-noise ratio. Thus, in these circumstances, the larger gains from scheduling could close the small gap between receive antennas only and both transmit and receive antennas. This is confirmed by Figure 5, where receive diversity only yields higher throughput than both transmit and receive diversity, unless the number of antennas is sufficiently large. In other words, in these scenarios transmit diversity antennas may negatively affect capacity. This is further illustrated in Figure 8 where the throughput is plotted as function of the signal-to-noise ratio for a fixed number of antennas. Similar observations have been made in [5], [19] for the special case of just a single receive antenna. It is shown in [19] that in the limiting regime with a large number of transmit antennas and a large number of users, the throughput with diversity may in fact be infinitely worse than with no diversity. In the case of a large number of users and a large number of receive antennas however, the throughput with diversity may be higher than with no diversity by any arbitrary factor.

We repeated the above set of experiments for a scenario with  $M = 8$  identical users. The results are plotted in Figures 5-7. The conclusions are basically similar, except that the gains

from scheduling over round robin are slightly larger than before. This may be explained by observing that the capacity is now determined by the maximum of twice the number of independent random variables from before.

In order to further explore the above issues, we also examined the total throughput as function of the number of users for a fixed number of antennas. The results are presented in Figures 12-14 for various values of the signal-to-noise ratio  $P$ . Observe that the total throughput grows more slowly in the case of four transmit antennas than in the case of just a single transmit antenna. This is a further manifestation of the fact that multiple transmit antennas have the effect of reducing the variations in the feasible rates without significantly increasing the mean rates. As a result, the throughput for four transmit antennas and a single receive antenna is smaller than for a single transmit and receive antenna, unless the number of users is extremely small. Similarly, at low values of the signal-to-noise, the throughput for four transmit and receive antennas is smaller than for four receive antennas only.

**Remark 4.2** *An alternative method for operating antenna-array systems is selection transmit diversity, where all the power is allocated to the currently ‘best’ transmit antenna, rather than statically split across all antennas, see for instance [8]. In combination with multi-user diversity scheduling, this approach results in the selection of the ‘best’ user-antenna pair. Note that this requires additional feedback information and pilot signals associated with the various transmit antennas.*

*In contrast to static transmit diversity, it is immediate that under no circumstance selection diversity can cause a degradation in throughput as compared to a single transmit antenna. In fact, it is not difficult to see that the total throughput with  $t$  transmit antennas to choose from is identical to that in a situation with a single transmit antenna and  $t$  times as many users. However, since the total throughput only grows slowly with the number of users, as indicated by Figures 12-14, the gains from selection transmit diversity seem to be intrinsically limited.*

**Remark 4.3** *As a final comment, it is worth pointing out that the numerical experiments concentrated on the long-term average throughput. Although the long-term average is a crucial factor in determining end-to-end throughput performance for data applications, transient throughput measures are important too. In particular, throughput stabilization is of vital importance, since even temporary starvation may cause a severe degradation in TCP throughput performance. In that sense, the reduction in the fluctuations of the transmission rate due to the use of diversity antennas may offer similar benefits as for voice communications, despite the adverse interaction with channel-aware scheduling mechanisms.*

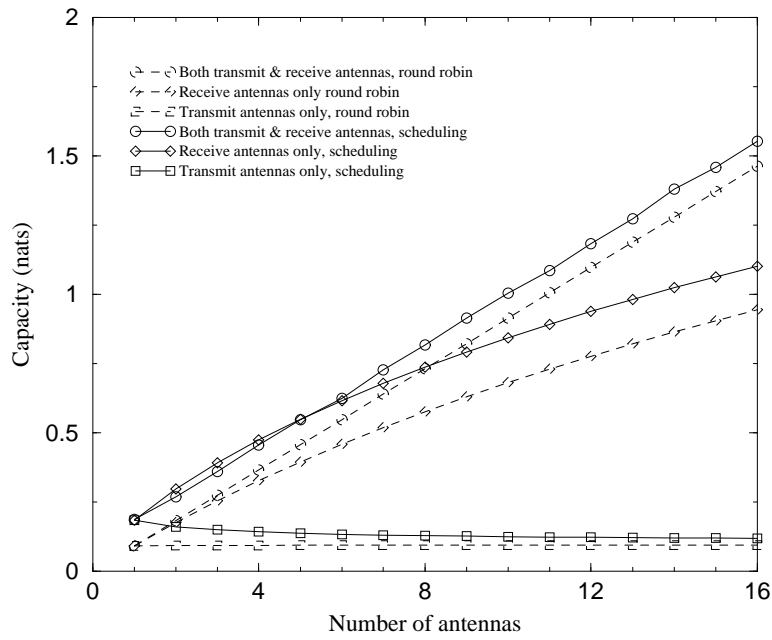


Figure 5: Total capacity as function of number of antennas; 4 identical users at  $P = -10.0\text{dB}$ .

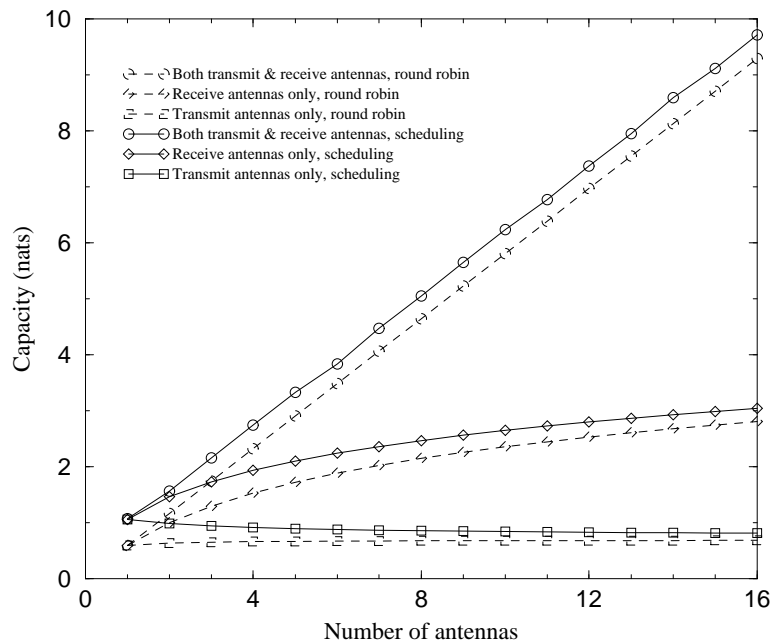


Figure 6: Total capacity as function of number of antennas; 4 identical users at  $P = 0.0\text{dB}$ .

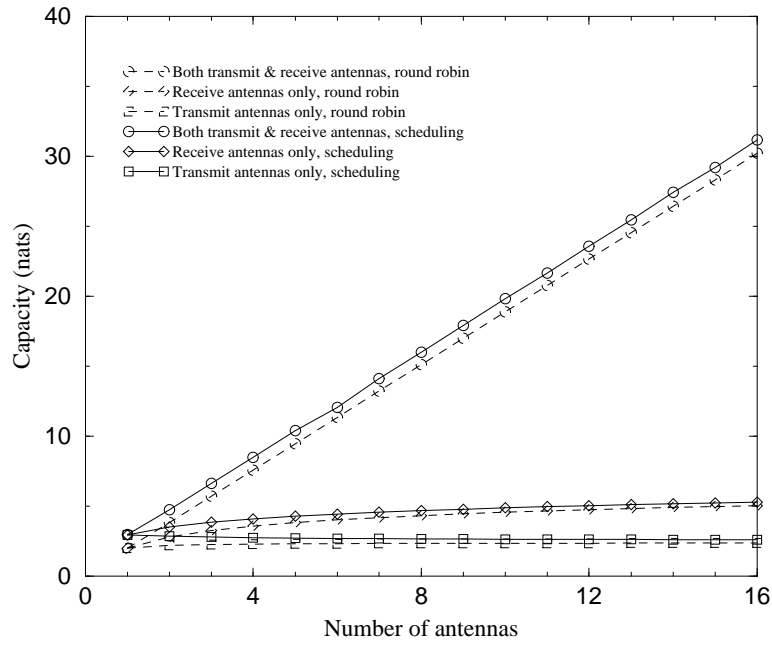


Figure 7: Total capacity as function of number of antennas; 4 identical users at  $P = 10.0$  dB.

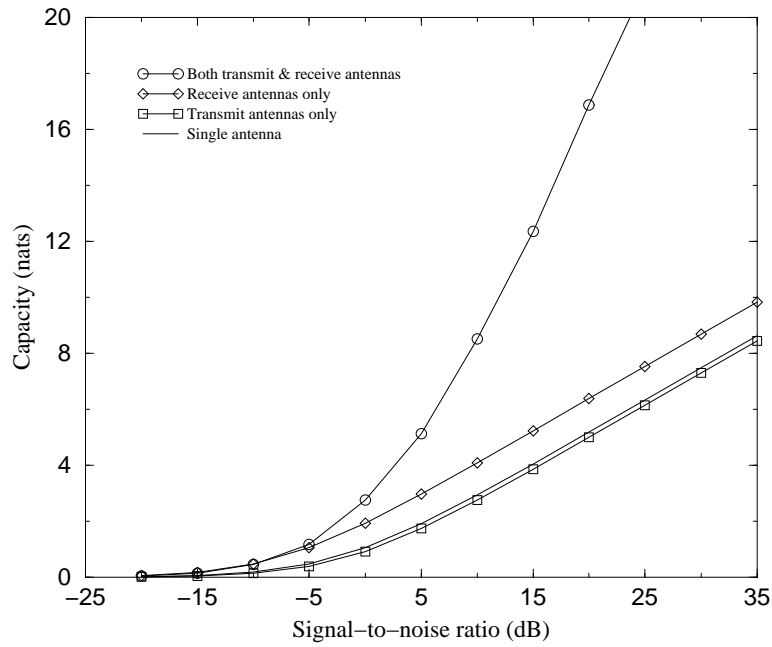


Figure 8: Total capacity as function of signal-to-noise ratio for 4 antennas; 4 identical users.

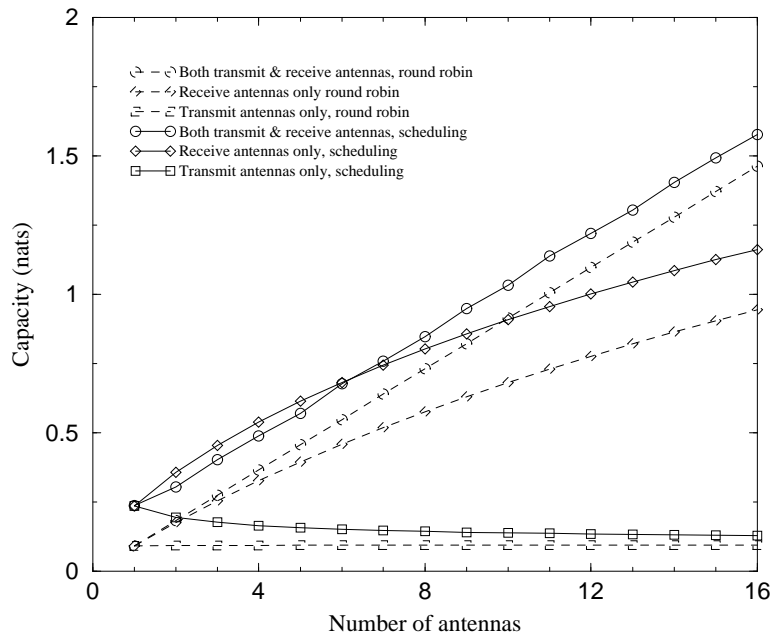


Figure 9: Total capacity as function of number of antennas; 8 identical users at  $P = -10.0\text{dB}$ .

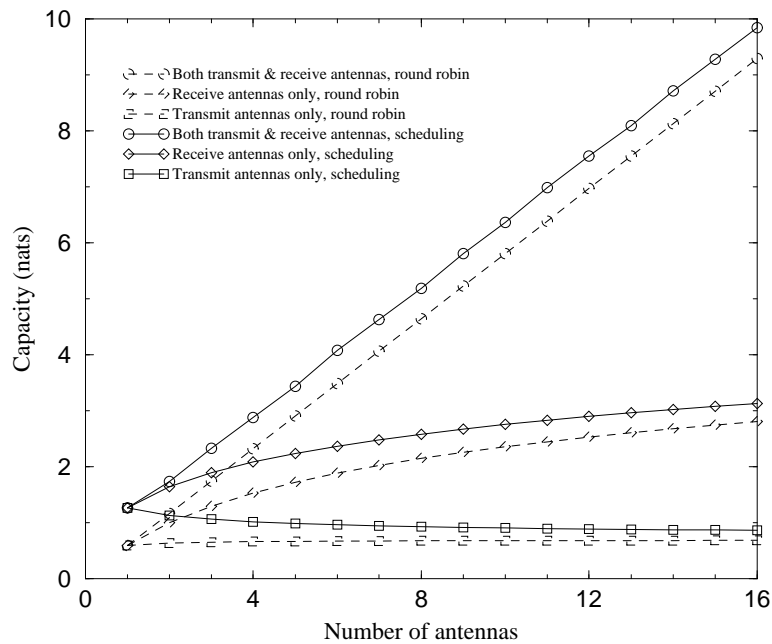


Figure 10: Total capacity as function of number of antennas; 8 identical users at  $P = 0.0\text{dB}$ .

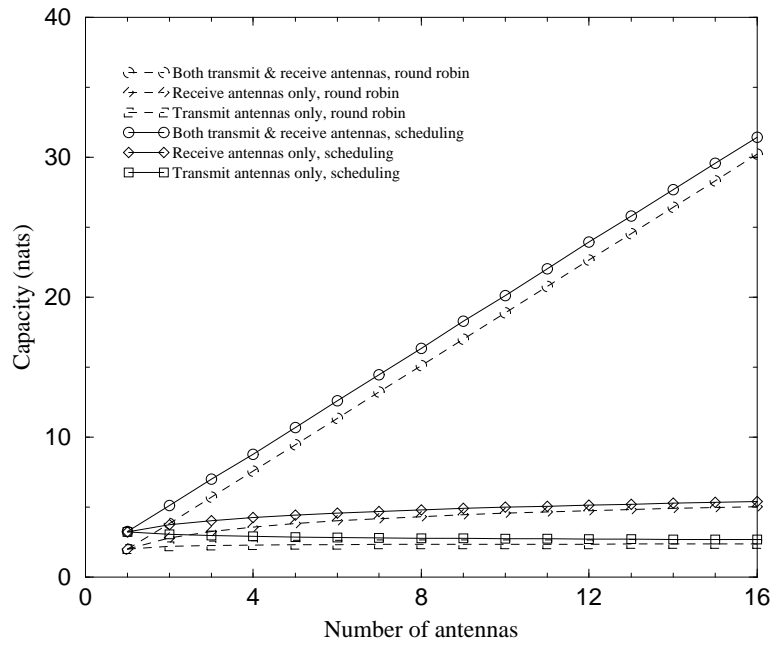


Figure 11: Total capacity as function of number of antennas; 8 identical users at  $P = 10.0\text{dB}$ .

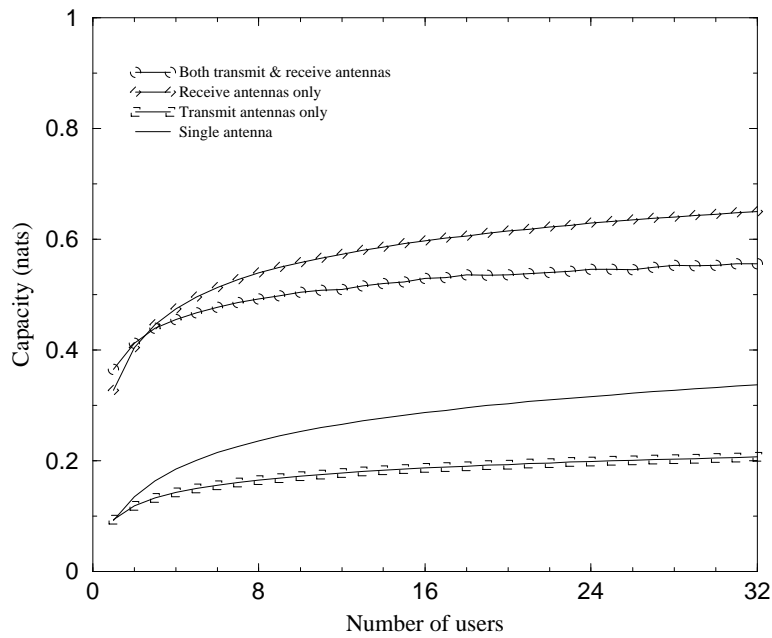


Figure 12: Total capacity as function of number of users at  $P = -10.0\text{dB}$ ; 4 antennas.

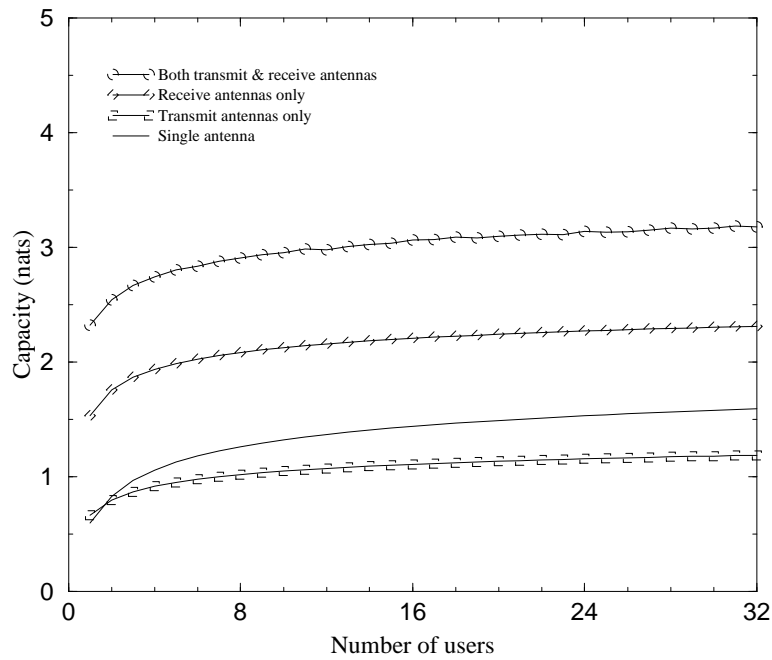


Figure 13: Total capacity as function of number of users at  $P = 0.0\text{dB}$ ; 4 antennas.

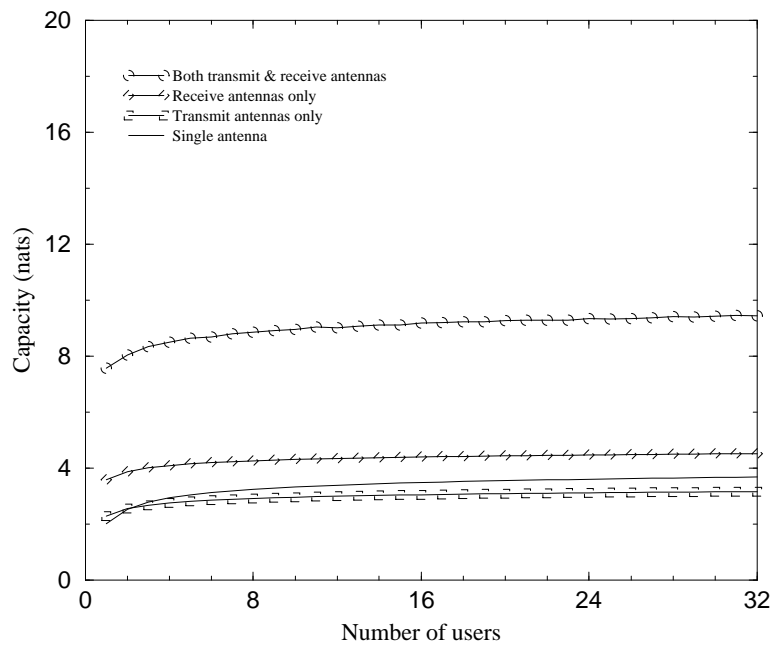


Figure 14: Total capacity as function of number of users at  $P = 10.0\text{dB}$ ; 4 antennas.

## 5 Heterogeneous users

In the previous section we focused on a scenario with identical users. We considered a scheduling strategy which always assigns the transmission slot to the user with the currently maximum feasible rate, and hence maximizes the total throughput. By symmetry, this strategy also equalizes the long-term average throughput among all users as well as the long-term fraction of time slots allocated.

In this section we turn the attention to a scenario with heterogeneous users. The above-described strategy then still maximizes the total throughput. However, it tends to assign a larger fraction of the time slots and thus provide higher throughput to users with better signal-to-noise ratios, possibly causing starvation of users with low signal-to-noise ratios. Equalizing the throughput among all users, in contrast, would require a larger fraction of the time slots to be allocated to users with low signal-to-noise ratios, which would reduce total throughput. The above observations reflect the trade-off that typically exists between maximizing total throughput and balancing the throughput among all users. The ‘optimal’ trade-off between these two conflicting objectives is not unique, and depends on the specific fairness criterion used. In that sense, one usually cannot identify a unique optimal throughput vector. In order to side-step the fairness issue, it is useful to adopt the notion of throughput vectors that are ‘Pareto-optimal’ in the sense that they are not dominated by any other achievable throughput vector.

In this section we consider scheduling strategies which are defined by a set of weights  $w_1, \dots, w_M$  for each of the users. The strategy defined by weight vector  $w$  assigns the  $n$ -th transmission slot to the user  $m^*(n)$  identified as

$$m^*(n) = \arg \max_{m=1, \dots, M} w_m R_m(n).$$

Note that these strategies may be viewed as a natural extension of the best-user scheduling strategy considered in the previous section. It is shown in [1], [3] that the above class of scheduling strategies are Pareto-optimal in the sense that the realized throughput vector cannot be dominated by any other achievable throughput vector.

In order to further explore the above issues, we conducted some numerical experiments for a scenario with  $M = 8$  heterogeneous users. For conceptual transparency, we assume that the users may be categorized into two distinct classes,  $M_1$  ‘weak’ users and  $M_2$  ‘strong’ users. However, most of the qualitative arguments developed below pertain to general scenarios with an arbitrary population of heterogeneous users.

In the first set of experiments, we consider a situation where all users receive the same (long-term) fraction of time slots. As before, we further distinguish between two strategies. The

first one is a static round-robin scheme which equally divides the time slots among the users in a cyclic fashion, oblivious of the state of channel. Note that the throughput received by each individual user is identical to that in a homogeneous scenario with the same total number of users as studied in the previous section. Also, the total throughput is simply equal to the average value of the total throughputs for the two user classes served in isolation. The second strategy does take the state of the channel into account, and dynamically assigns time slots using weights  $w_1^*$  for the weak users and  $w_2^*$  for the strong users as indicated above. The relative weights are set such that all users receive the same long-term fraction of time slots. In order to determine these weights, we used a slightly modified version of the control algorithm in [3], which balances the number of time slots among users rather than the throughputs.

The results are displayed in Figures 15-17. Each of data points is obtained using a run of 100,000 frames. The irregularities in the curves are a further manifestation of the stiffening of the feasible rates mentioned earlier. Due to the concentration of the rates around the mean values, the per-class throughputs are extremely sensitive to the values of the weights. As a result, the convergence of the weights to the correct values is exceedingly slow.

Note that Figure 16 with  $M_1 = 4$  weak users at  $P_1 = -10.0\text{dB}$  and  $M_2 = 4$  strong users at  $P_2 = 10.0\text{dB}$  looks quite similar to Figure 7 for the case of  $M = 4$  homogeneous users at  $P = 10.0\text{dB}$ , scaled down by a factor 2. This may be explained by observing that lion share in the total throughput as plotted in Figure 16 is constituted by the throughput of the 4 strong users, which however only receive half of the time slots.

The severe imbalance in throughput between weak and strong users mentioned above may be viewed as ‘unfair’. Next, we therefore consider a situation where all users receive the same long-term throughput. As in the previous case, we further distinguish between two strategies. The first one always allocates the time slot to the user with the currently minimum cumulative throughput, independent of the state of the channel. The second strategy does take the state of the channel into account, and dynamically assigns time slots using weights  $w_1^*$  for the weak users and  $w_2^*$  for the strong users as indicated above. The weights are now set in such a way that all users receive the same long-term throughput. In order to calculate these weights, we used the control algorithm in [3], exploiting the symmetry between users within the two classes. The results are graphed in Figures 18-20. Observe that Figure 19 with  $M_1 = 4$  weak users at  $P_1 = -10.0\text{dB}$  and  $M_2 = 4$  strong users at  $P_2 = 10.0\text{dB}$  now somewhat resembles Figure 5 for the case of  $M = 4$  homogeneous users at  $P = -10.0\text{dB}$ . This may be understood as follows. For the throughput to be balanced between the two classes, the vast majority of the time slots need to be allocated to the weak users. As a result, the capacity gains from the various diversity schemes basically coincide with those in a scenario with weak users only.

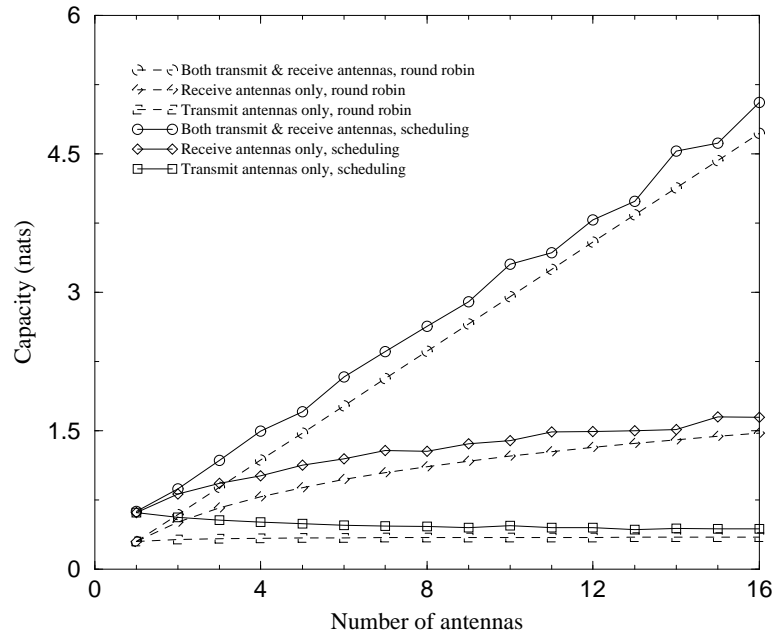


Figure 15: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -20.0\text{dB}$ ; 4 users at  $P_2 = 0.0\text{dB}$ ; equal fraction of time slots.

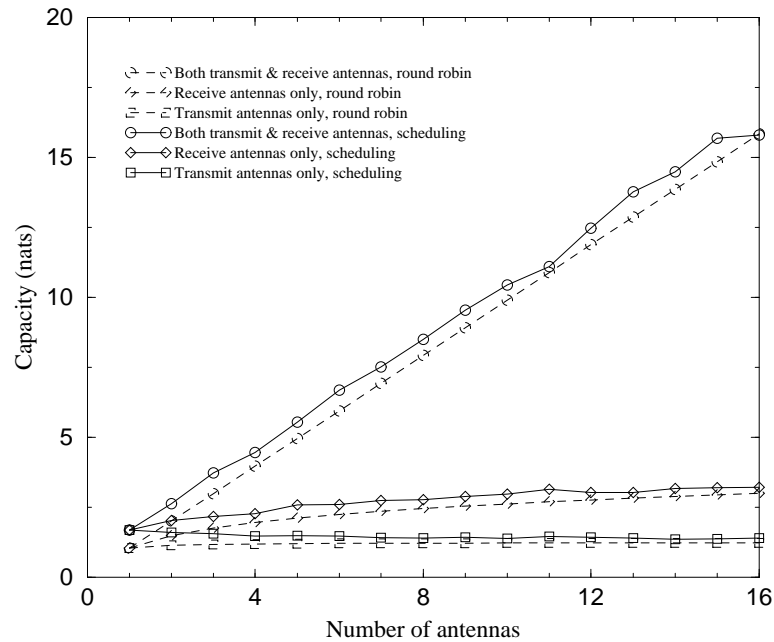


Figure 16: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ ; equal fraction of time slots.

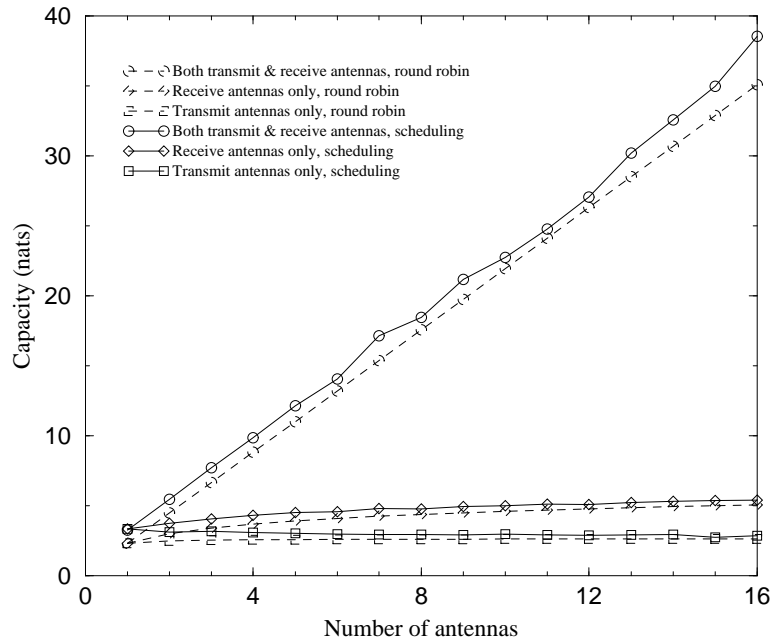


Figure 17: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = 0.0\text{dB}$ ; 4 users at  $P_2 = 20.0\text{dB}$ ; equal fraction of time slots.

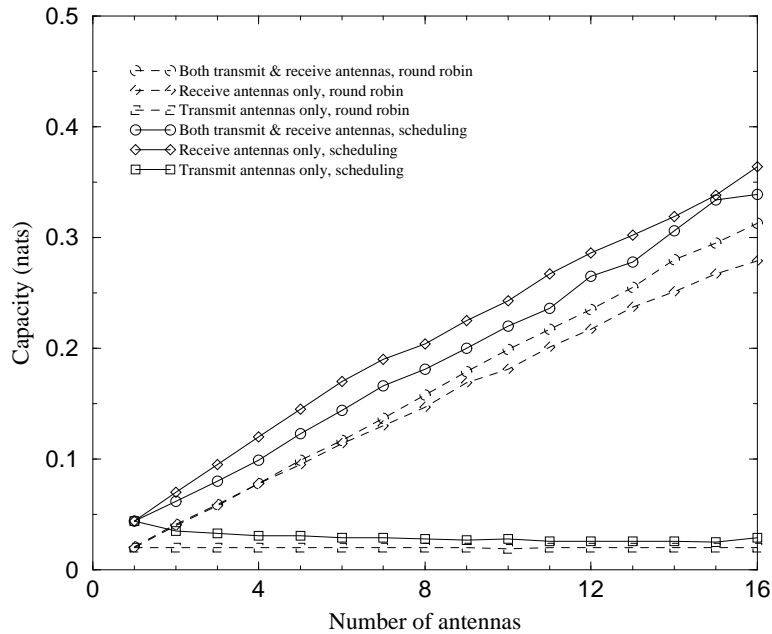


Figure 18: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -20.0\text{dB}$ ; 4 users at  $P_2 = 0.0\text{dB}$ ; equal throughputs.

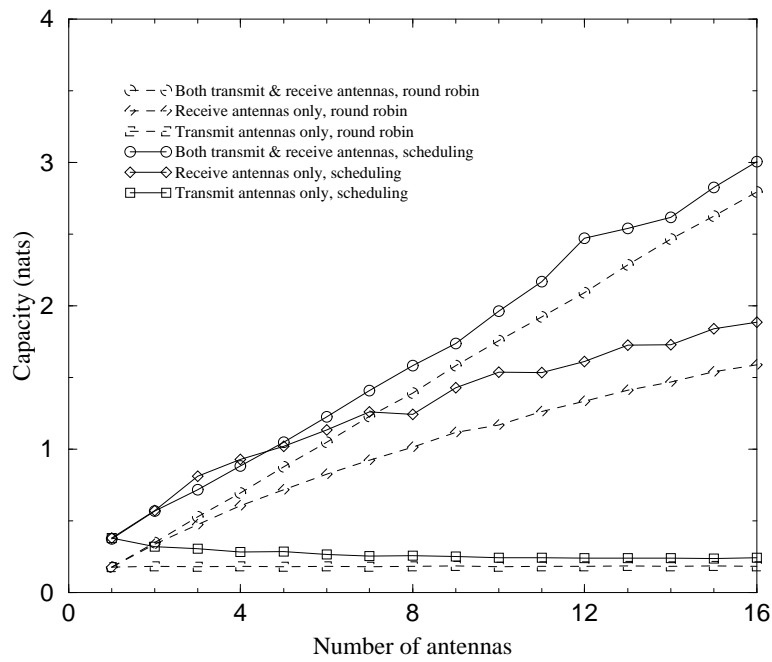


Figure 19: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ ; equal throughputs.

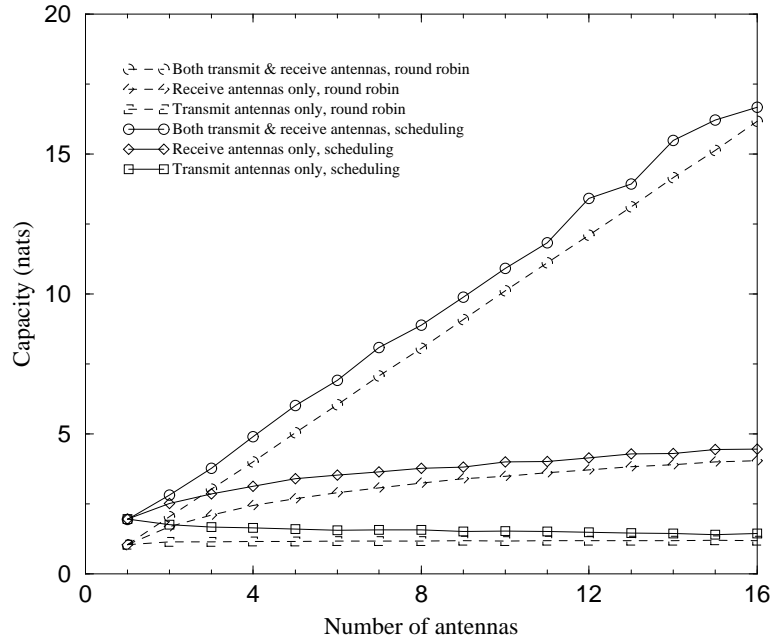


Figure 20: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = 0.0\text{dB}$ ; 4 users at  $P_2 = 20.0\text{dB}$ ; equal throughputs.

It may be argued unreasonable to insist on balanced throughput when users have radically different characteristics. Granting equal throughput to the weak leaves only a small fraction of the time slots to the strong users and degrades total throughput dramatically.

Hence, we next consider a situation where we allow the throughput of the strong users to be 9 times as large as that of the weak users. Although the ratio of 9 may seem rather extreme, it turns out that it still results in the majority of the time slots being allocated to the weak users. As in the previous case, we further distinguish between two strategies. The first one always allocates the time slot to the user with the currently minimum cumulative throughput, but now with the throughputs of the strong users normalized by a factor 9. The second strategy assigns time slots using weights  $w_1^*$  for the weak users and  $w_2^*$  for the strong users. The weights are now set in such a way that the strong users receive the 9-fold long-term throughput of the weak users. In order to compute these weights, we used again the control algorithm in [3].

The results are plotted in Figures 21-23. Comparing the results with the corresponding numbers in Figures 18-20 for equal throughput proportions, we see that allowing a degree of disparity in the throughputs of individual users helps to significantly raise total throughput.

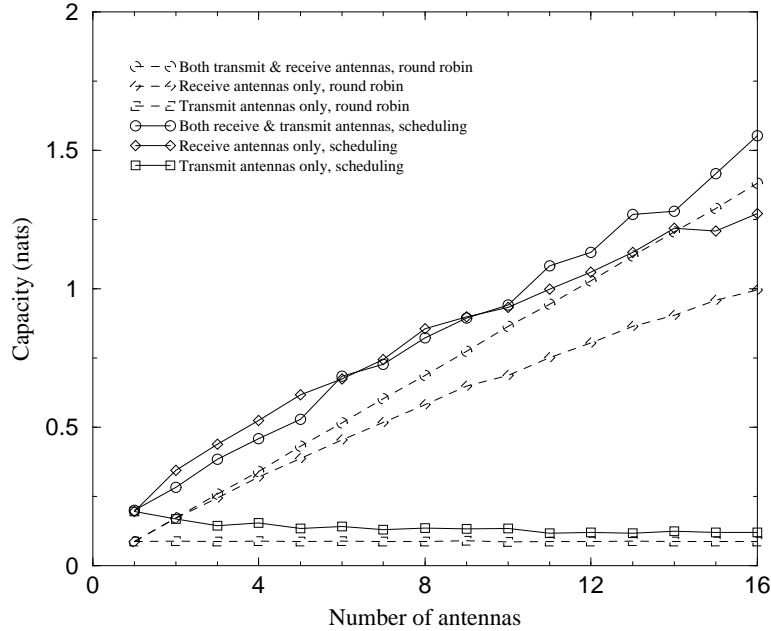


Figure 21: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -20.0\text{dB}$ ; 4 users at  $P_2 = 0.0\text{dB}$ ; unequal throughputs.

However, the capacity gains from the various diversity schemes are still largely reflective of those in a scenario with poor users only.

Similarly, it may be reasoned that dividing time slots equally is not entirely fair when users have extremely diverse characteristics. With an equal share, the weak users may receive hardly any throughput, while the strong users may still enjoy an excessive amount of throughput. Therefore, we also consider a situation where we grant 9 times as many time slots to the weak users as to the strong users. Although the ratio 9 may seem quite generous, it turns out that it does not prevent the weak users from still receiving far less throughput than the strong users. As before, we further distinguish between two strategies. The first one divides the time slot in the target proportions in a static fashion. The second strategy assigns time slots using weights  $w_1^*$  for the weak users and  $w_2^*$  for the strong users. The weights are now set in such a way that the weak users receive 9 times as many time slots as the weak users. In order to determine these weights, we used again a slightly modified version of the control algorithm in [3], which balances the normalized number of time slots among users rather than the throughputs.

The results are presented in Figures 24-26. As in a previous case, Figure 25 looks like a scaled version of Figure 7, although to a somewhat lesser extent. This may be explained by observing

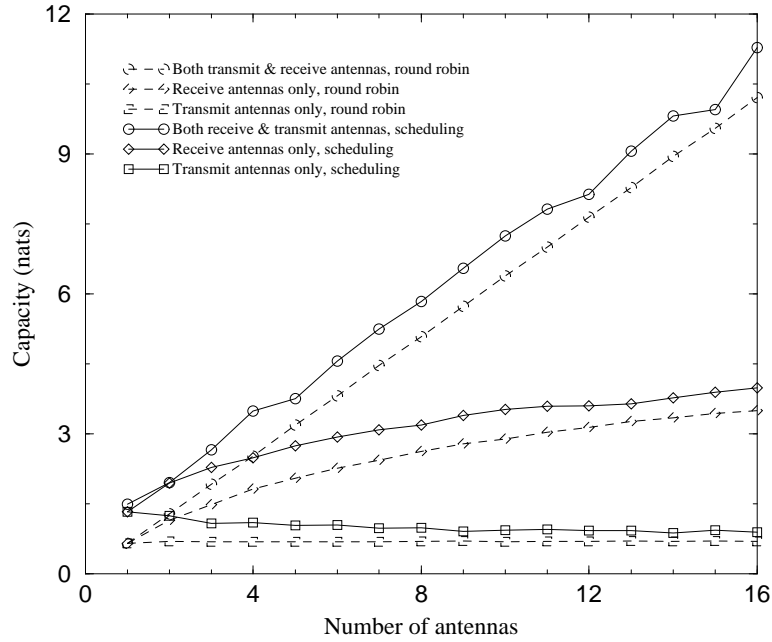


Figure 22: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ ; unequal throughputs.

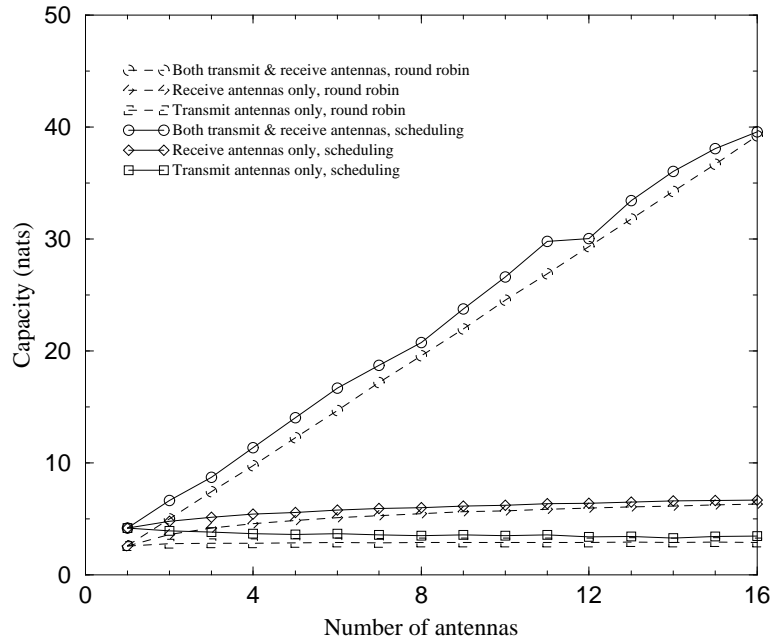


Figure 23: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = 0.0\text{dB}$ ; 4 users at  $P_2 = 20.0\text{dB}$ ; unequal throughputs.

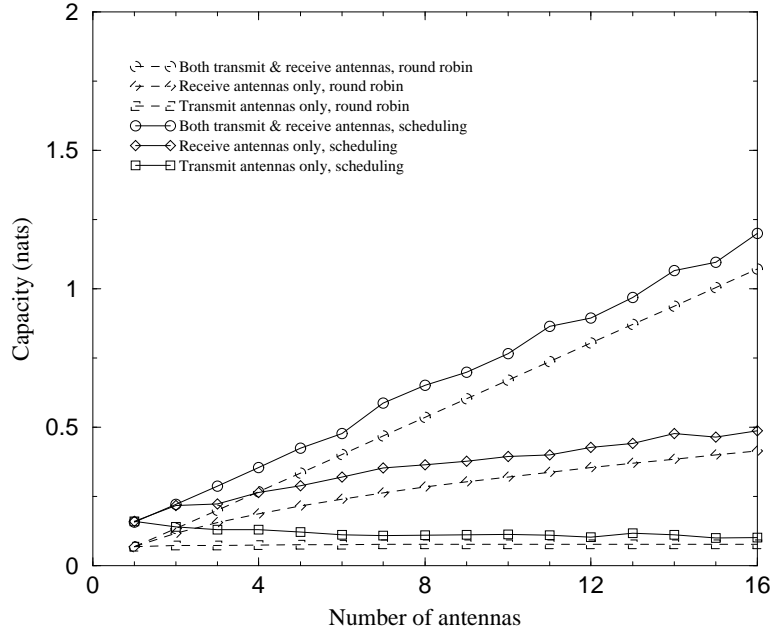


Figure 24: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -20.0\text{dB}$ ; 4 users at  $P_2 = 0.0\text{dB}$ ; unequal fraction of time slots.

that the total throughput as graphed in Figure 22 is still heavily dominated by the throughput of the strong users, which now however receive only one tenth of the time slots.

In the final set of experiments, we consider a scheduling strategy proposed in [6], [18], which also relies on user-specific weights to assign time slots. The weights are essentially set reciprocal to the long-term throughputs received by the various users. This strategy achieves ‘proportional’ fairness, which means that the throughput of no user can be improved without a greater percentage reduction in the throughputs of the other users.

The results are shown in Figures 27-29. Note that the results strongly resemble the corresponding numbers in Figures 15-17 where all users are allocated the same fraction of time slots. It turns out that the above-described scheduling rule indeed results in all users receiving roughly the same number of time slots.

For conceptual simplicity, we concentrated in the above experiments on a somewhat idealized scenario with just two distinct user classes. However, most of the qualitative conclusions also apply in general scenarios with an arbitrary configuration of heterogeneous users. In particular, we conclude that in heterogeneous scenarios the capacity improvements from diversity antennas may widely vary, depending on what fairness notion is adopted. If the proportion of time slots is used as fairness criterion, then the gains in total throughput tend to be significant, but the

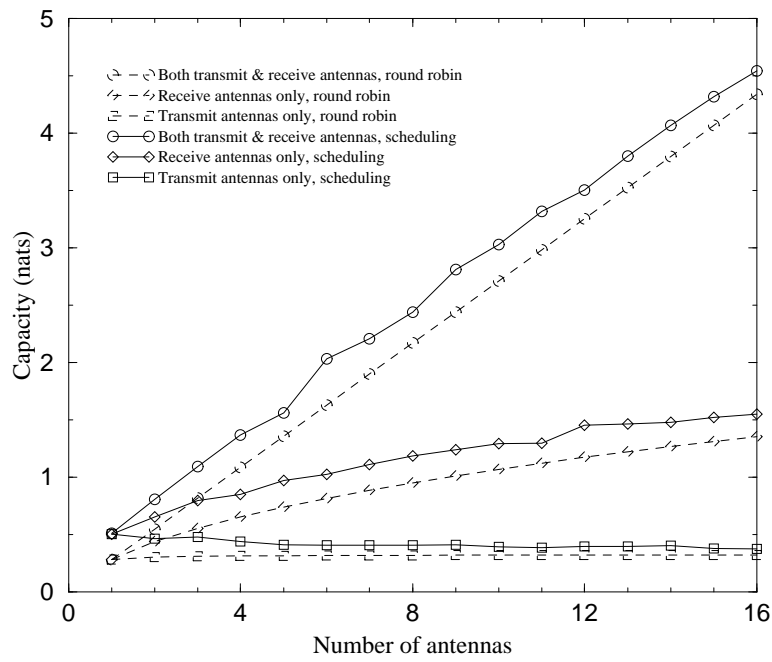


Figure 25: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ ; unequal fraction of time slots.

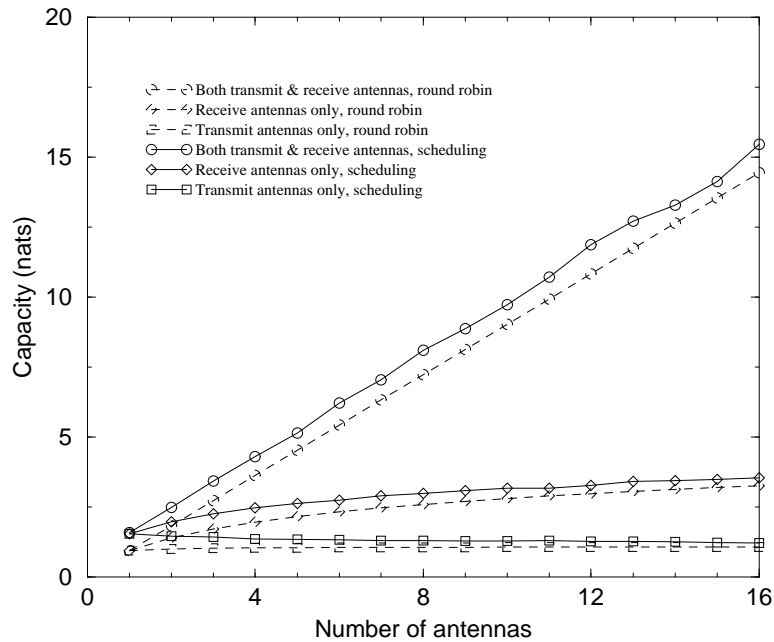


Figure 26: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = 0.0\text{dB}$ ; 4 users at  $P_2 = 20.0\text{dB}$ ; unequal fraction of time slots.

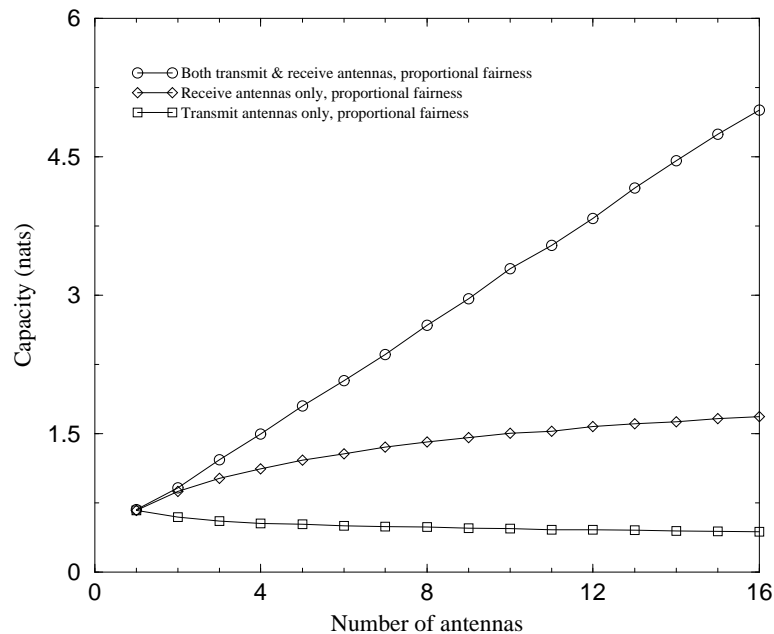


Figure 27: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -20.0\text{dB}$ ; 4 users at  $P_2 = 0.0\text{dB}$ ; proportional fairness.

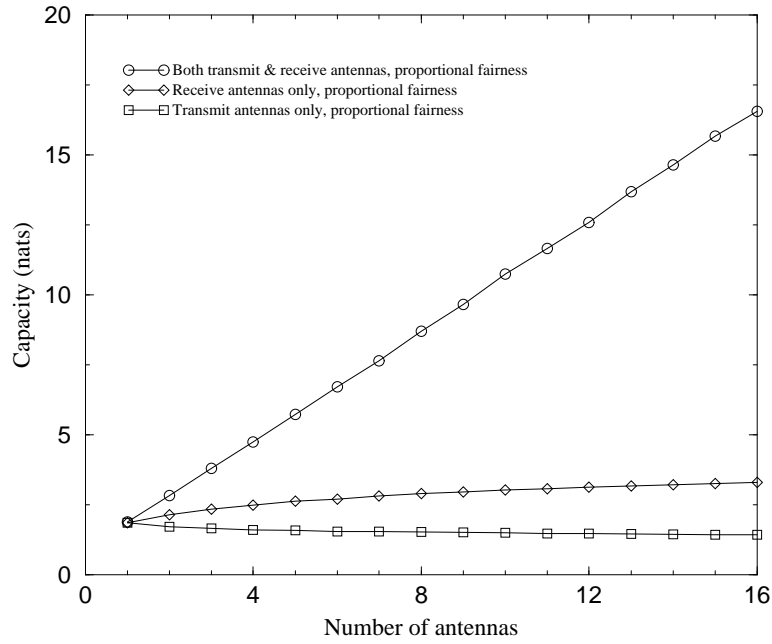


Figure 28: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ ; proportional fairness.

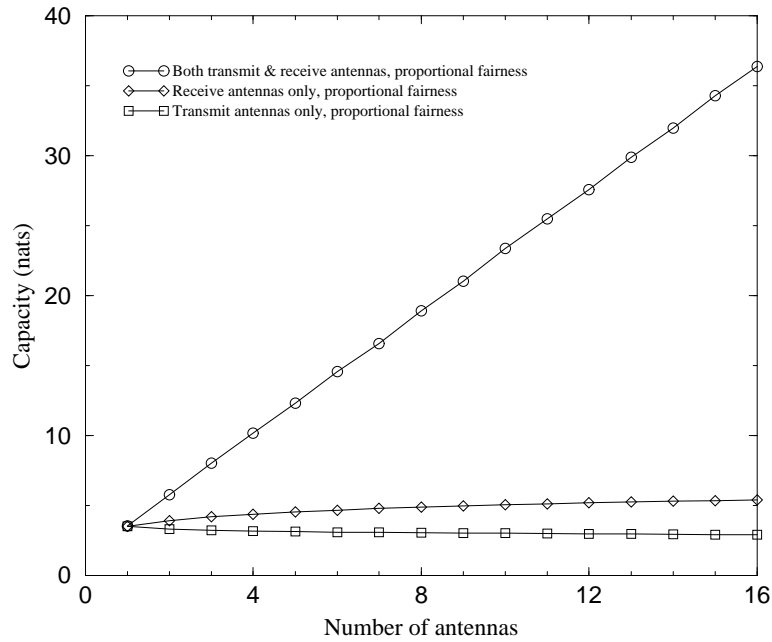


Figure 29: Total capacity as function of number of antennas; 8 users; 4 users at  $P_1 = 0.0\text{dB}$ ; 4 users at  $P_2 = 20.0\text{dB}$ ; proportional fairness.

edge users only receive a marginal share. In case the fairness measure is defined in terms of throughput ratios, then the gains tend to be limited population-wide.

In order to illustrate the above points from a slightly different perspective, we also considered the achievable throughput region. We assume equal throughputs within each of the two user classes. The regions are indicated in Figures 30 and 31 for scenarios with two and four antennas, respectively. Note that the points on the vertical axis correspond to a scenario with  $M_1 = 4$  stochastically identical users in isolation, and agree with the numbers in Figure 5 for the case of two and four antennas, respectively. Similarly, the points on the horizontal axis coincide with the results in Figure 7.

In order to maximize the total throughput, it is crucial to operate close to the horizontal axis. At these operating points however, there is a severe imbalance in the throughputs received by the two user classes. In order to reduce the disparity in throughput, it is necessary to operate closer to the vertical axis, which however causes a dramatic reduction in total throughput, reflecting the trade-off mentioned earlier. Further observe the cross-over point between the boundaries for receive diversity only and both receive and transmit diversity. This indicates that when the majority of time slots are allocated to the weak users, receive diversity only performs better than both receive and transmit diversity, as suggested by the results in Figure 5.

## 6 Conclusion

We have examined the capacity improvements from diversity antennas in multi-user scenarios. We showed that the use of diversity antennas has the effect of dampening the variations in the channel conditions. This reduces the scope for channel-aware scheduling mechanisms to obtain capacity gains by exploiting fluctuations in the feasible transmission rates. Conversely, in the presence of scheduling, diversity antennas may yield smaller gains, or even have a negative impact on capacity.

Second, in heterogeneous scenarios, we demonstrated that the capacity improvements from diversity antennas may dramatically vary, depending on what fairness notion is adopted. If the proportion of time slots is used as a fairness criterion, then the gains in total throughput tend to be considerable. However, edge users only receive a negligible portion of the increase in throughput, illustrating the trade-off mentioned above. If the fairness measure is defined in terms of throughput shares, then the gains tend to be modest population-wide.

To overcome some of the inherent limitations of (transmit) diversity antennas, it may be worth considering alternative schemes for operating antenna-array systems, such as dynamic receive or transmit beam-forming, possibly in combination with transmitting to several users simulta-

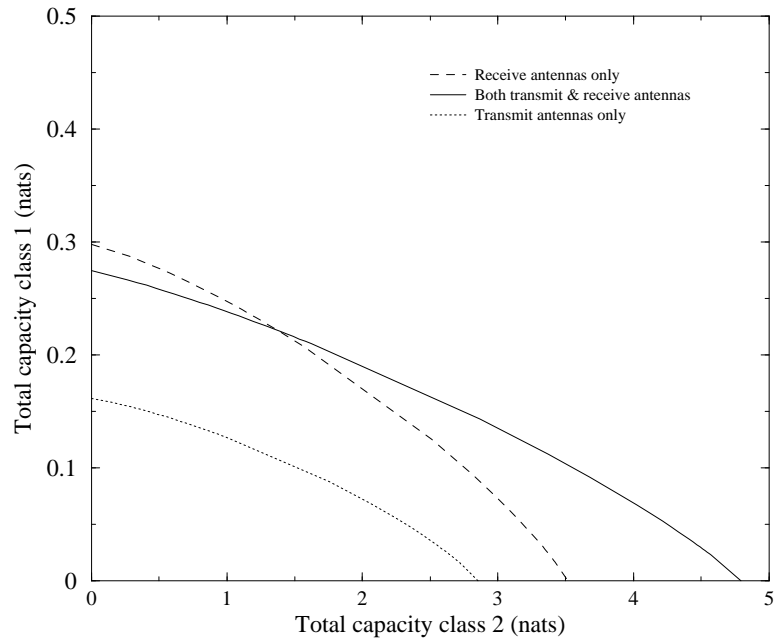


Figure 30: Achievable throughput region for 2 antennas; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ .

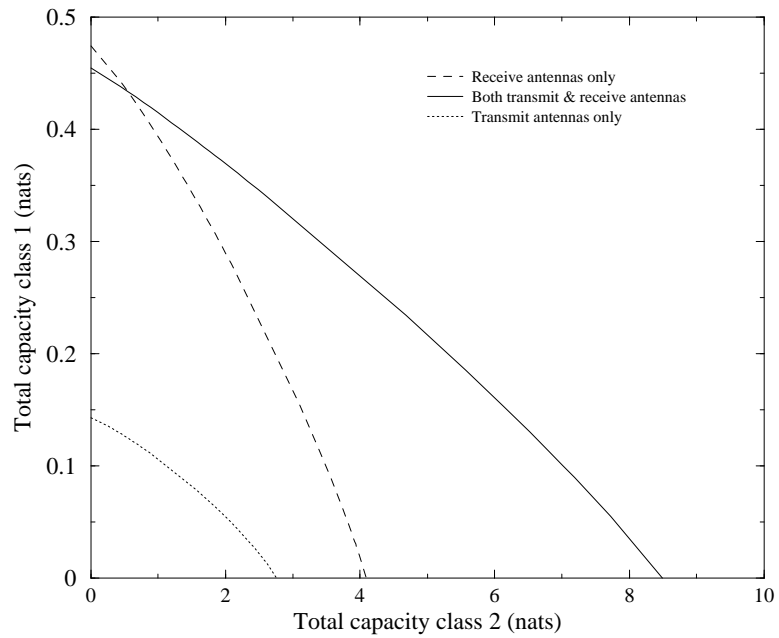


Figure 31: Achievable throughput region for 4 antennas; 4 users at  $P_1 = -10.0\text{dB}$ ; 4 users at  $P_2 = 10.0\text{dB}$ .

neously as proposed in [9]. Under the beam-forming approach, channel information is used to set complex weights for the antenna elements which in turn determine the antenna gain. We now contrast the use of transmit and receive beam-forming techniques as applied to scheduled systems such as Qualcomm HDR. A key feature of this approach is the reliance on the *user's rate being predictable*. Transmit beam-forming debilitates this predictability, since there is a “knock-on” change in out-of-cell interference. This difficulty is not shared by receive beam-forming at the user's handset where the synthesized beam impacts the signal-to-noise ratio of that user only, and has no implications for other users at all. Predictability is thus preserved. Receive beam-forming also allows the nulling or suppression of adjacent cell interference which tends to affect edge users more severely. Note that the diversity approach described here allows interference suppression as well, if for example the gain matrix to the dominant interfering cell can be estimated.

**Acknowledgment** The authors are grateful to J.A. Morrison for some helpful observations.

## A The evaluation of $C(2, 2, P, M)$

In order to evaluate the capacity  $C(2, 2, P, M)$ , we need to calculate the distribution of the random variable  $Z \equiv Z(2, 2, P)$ . As in (2),  $Z$  may be represented as

$$Z = \log(1 + P'\lambda_1) + \log(1 + P'\lambda_2),$$

with  $P' = P/2$  and  $\lambda_1, \lambda_2$  denoting the real and non-negative eigenvalues of the matrix  $W$  whose joint density is that of the Wishart distribution with parameters 2, 2 [16]:

$$\frac{1}{2}e^{-u_1-u_2}(u_1 - u_2)^2.$$

In order to compute the density  $\frac{d}{dw}\mathbb{P}\{Z \leq w\}$ , we first consider the Laplace transform

$$\begin{aligned} H(s) &= \mathbb{E}[e^{-sZ}] \\ &= \frac{1}{2} \int_{u_1=0}^{\infty} \int_{u_2=0}^{\infty} e^{-u_1-u_2}(u_1 - u_2)^2 \frac{1}{(1 + P'u_1)^s} \frac{1}{(1 + P'u_2)^s} du_1 du_2 \\ &= \int_{u_1=0}^{\infty} \frac{e^{-u_1} u_1^2}{(1 + P'u_1)^s} du_1 \int_{u_2=0}^{\infty} \frac{e^{-u_2}}{(1 + P'u_2)^s} du_2 - \left[ \int_{u=0}^{\infty} e^{-u} u \frac{1}{(1 + P'u)^s} du \right]^2 \\ &= F_2(s)F_0(s) - (F_1(s))^2, \end{aligned}$$

with

$$F_n(s) = \int_{u=0}^{\infty} \frac{u^n e^{-u}}{(1 + P'u)^s} du.$$

Now observe that  $F_n(s) = \int_{w=0}^{\infty} e^{-sw} f_n(w) dw$ , with

$$f_n(w) = \int_{u=0}^{\infty} u^n e^{-u} \delta(w - \log(1 + P'u)) du$$

where  $\delta$  denotes the Dirac delta function.

Thus,

$$\frac{d}{dw} \mathbb{P}\{Z > w\} = \int_{v=0}^w f_2(v) f_0(w-v) dv - \int_{v=0}^w f_1(v) f_1(w-v) dv.$$

Making the substitution  $1 + P'u = e^v$ , we obtain

$$f_n(w) = \left(\frac{1}{P'}\right)^{n+1} (e^w - 1) e^{-\frac{1}{P'}(e^w - 1)} e^w.$$

Hence,

$$\begin{aligned} \int_{v=0}^w f_1(v) f_1(w-v) dv &= \frac{1}{P'^4} \int_{v=0}^w (e^v - 1)(e^{w-v} - 1) e^{-\frac{1}{P'}(e^v + e^{w-v} - 2)} e^w dv \\ &= \frac{x}{P'^4} \int_{y=1}^x (y-1) \left(\frac{x}{y} - 1\right) e^{-\frac{1}{P'}(y+x/y-2)} \frac{dy}{y}, \end{aligned}$$

with  $x = e^w$ ,  $y = e^v$ , and,

$$\int_{v=0}^w f_2(v) f_0(w-v) dv = \frac{x}{P'^4} \int_{y=1}^x (y-1)^2 e^{-\frac{1}{P'}(y+x/y-2)} \frac{dy}{y}.$$

Taking the difference, we finally obtain

$$\mathbb{P}\{Z > \log x\} = \frac{2}{P'^4} e^{2/P'} \int_{y=1}^x (y-1) \left(1 - \frac{x}{y^2}\right) e^{-\frac{1}{P'}(y+x/y)} dy \quad (15)$$

$$= \frac{2}{P'^3} \left[ e^{2/P'} \sqrt{x} \int_{z=\frac{1}{\sqrt{x}}}^{\sqrt{x}} e^{-\frac{\sqrt{x}}{P'}(z+\frac{1}{z})} dz - (x-1) e^{-1(x-1)/P'} \right]. \quad (16)$$

## References

- [1] Andrews, D.M., Kumaran, K., Ramanan, K., Stolyar, A.L., Vijayakumar, R., Whiting, P.A. (2000). CDMA data QoS scheduling on the forward link with variable channel conditions. Technical Memorandum 10009626-000404-05TM, Bell Laboratories, Lucent Technologies.
- [2] Bender, P., Black, P., Grob, M., Padovani, R., Sindhushayana, N., Viterbi, A. (2000). CDMA/HDR: a bandwidth-efficient high-speed wireless data service for nomadic users. *IEEE Commun. Mag.* **38** (7), 70–77.
- [3] Borst, S.C., Whiting, S.C. (2001). Dynamic rate control algorithms for HDR throughput optimization. Technical Memorandum, Bell Laboratories, Lucent Technologies. Shortened version in: *Proc. Infocom 2001*, 976–985.
- [4] Foschini, G.J., Gans, M.J. (1998). On limits of wireless communications in a fading environment when using multiple-element antennas. *Wireless Pers. Commun.* **6**, 311–335.
- [5] Huang, H.C., Viswanathan, H., Blanksby, A., Haleem, M. (2000). Multiple antenna enhancements for a high rate CDMA packet data system. *J. VLSI Proc.*, to appear.
- [6] Jalali, A., Padovani, R., Pankaj, R. (2000). Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system. In: *Proc. 50th IEEE Veh. Techn. Conf.*, 1854–1858.
- [7] Knopp, R., Caire, G. (1999). Power control schemes for TDD systems with multiple transmit and receive antennas. In: *Proc. IEEE Globecom '99*, 2326–2330.
- [8] Kogiantis, A.G., Ozarow, L. (2001). Distributed multi-antenna scheduling for downlink and uplink packet data. Technical Memorandum JW9130000-001001-01TM, Bell Laboratories, Lucent Technologies.
- [9] Kumaran, K, Viswanathan, H. (2001). Scheduling in multiple antenna downlink wireless systems. Technical Memorandum 10009626-010720-01TM, Bell Laboratories, Lucent Technologies.
- [10] Lozano, A., Tulino, A.M. (2001). Capacity of multiple-transmit multiple-receive antennas architectures. Internal document, Bell Laboratories, Lucent Technologies.
- [11] Rashid-Farrokhi, F., Liu, K.J.R., Tassiulas, L. (1998). Transmit beamforming and power control for cellular wireless systems. *IEEE J. Sel. Areas Commun.* **16**, 1437–1450.

- [12] Rashid-Farrokhi, F., Tassiulas, L., K.J.R. Liu (1998). Joint optimal power control and beamforming in wireless networks using antenna arrays. *IEEE Trans. Commun.* **46**, 1313–1324.
- [13] Shakkottai, S., Stolyar, A.L. (2000). Throughput-optimal scheduling for time-varying channels: the exponential rule. Technical Memorandum 10009626-010102-01TM, Bell Laboratories, Lucent Technologies.
- [14] Shamai, S., Verdú, S. (2001). The effect of frequency-flat fading on the spectral efficiency of CDMA. *IEEE Trans. Inf. Theory*, to appear.
- [15] Swales, S.C., Beach, M.A., Edwards, D.J., McGeehan, J.P. (1990). The performance enhancement of multibeam adaptive base-station antennas for cellular land mobile radio systems. *IEEE Trans. Veh. Techn.*, 56–67.
- [16] Telatar, Í.E. (1999). Capacity of multi-antenna Gaussian channels. *European Trans. Telecom* **10**, 851–866.
- [17] Thompson, J.S., Grant, P.M., Mulgrew, B. (1996). Smart antenna arrays for CDMA systems. *IEEE Pers. Commun. Mag.* **3** (5), 16–25.
- [18] Tse, D. (1999). Forward link multiuser diversity through rate adaptation and scheduling. Presentation at Bell Laboratories, Lucent Technologies.
- [19] Viswanathan, H., Venkatesan, S. (2001). The impact of antenna diversity in packet data systems with scheduling. Technical Memorandum 10009651-010611-01TM, Bell Laboratories, Lucent Technologies.
- [20] Winters, J.H., Salz, J., Gitlin, R.D. (1994). The impact of antenna diversity on the capacity of wireless communication systems. *IEEE Trans. Commun.* **42**, 1740–1751.