

## Antenna Diversity vs. Multiuser Diversity: Quantifying the Tradeoffs

Mari Kobayashi Giuseppe Caire David Gesbert

Mobile Communication Group, Eurecom Institute, Sophia-Antipolis, FRANCE  
E-mail: `firstname.name@eurecom.fr`

### Abstract

We consider the impact on transmit antenna diversity on a TDMA downlink with multiuser diversity scheduling, inspired by the 1xEV-DO scheme in 3G systems. Transmit diversity is examined under the adaptive scheduling policy that achieves the stability region of the transmit queues, and, in the case of infinite backlog, under proportional fair scheduling. We show that, in the realistic case of non-ideal data rate feedback information from the users to the base station, transmit diversity might achieve a larger stability region and is beneficial even for users with symmetric traffic or infinite backlog. This findings attenuate the common belief that “channel hardening” due to transmit diversity is always detrimental for multiuser diversity scheduling systems.

### 1. Motivation

3G wireless cellular networks are expected to support a wide variety of data services. In particular, for applications such as wireless Internet, a high data rate downlink is needed. The downlink of a single cell system<sup>1</sup> is modeled as a fading Gaussian broadcast channel, whose capacity region has been completely characterized under different assumptions in several recent papers (e.g., [13]). In particular, it is known that, under mild fading ergodicity conditions, when the base station is equipped with a single antenna and has perfect *Channel State Information* (CSI), the average throughput (long-term average sum rate) is maximized by transmitting at each instant to the user with the largest fading coefficient (e.g., [1, 5]). Intuitively, serving the best user at each time allows using the channel at the fading peak rather than at the fading average level, thus achieving a power gain. This effect is referred to as *multiuser diversity*. Motivated by this result, downlink scheduling approaches known as High-Data Rate (HDR) system [4] or as 1xEV-DO system in 3G standards [14] have been proposed. Such systems assume that all connected users have infinite

backlog (i.e., all data to be transmitted are already present at the base station). At each time slot, the users measure their channel quality and send a data-rate control (DRC) signal to the base station. This is the data rate that, given the instantaneous channel quality, each user can reliably decode if served. The base station schedules the user to be served in the current slot as the one achieving the maximum of the DRC scaled by an adaptive factor that takes into account *fairness*.

When the base station is equipped with  $M > 1$  antennas, the single-cell downlink falls in the class of *vector Gaussian broadcast channels*, whose capacity region with perfect CSI has been fully characterized in [12] and references therein. In particular, for a system with  $M$  transmit antennas and  $K \geq M$  users, a multiplexing gain of  $M$  can be achieved, i.e., the average throughput scales as  $M \log \text{SNR}$  for high SNR, and  $M$  users can be served at the same time on each slot. A low-complexity alternative usage of multiple transmit antennas for the downlink with scheduling and TDMA consists of the *opportunistic beamforming* scheme proposed in [1], where the multiple antennas are used to generate a rotating beam inducing an artificial fading that varies slowly enough to be measured and fed back by the users but rapidly enough to make the scheduling algorithm share the channel equally among the users avoiding the bottleneck of users permanently in bad channel conditions. A spatial-multiplexing version of the opportunistic beamforming is proposed and analyzed in [10], where  $M$  mutually orthogonal random beams are simultaneously used and the best  $M$  users are selected at each time.

In parallel with the development of multiuser diversity schemes and their multi-antenna opportunistic beamforming versions, the current research and standardization trends in wireless cellular systems have focused on Space-Time Coding (STC). When no DRC feedback signal is available at the transmitter, the event that the transmitted rate falls below the instantaneous mutual information of the fading channel is called *information outage*. This is the event that dominates the decoding error probability for good codes in high SNR conditions [15, 16]. In the most realistic scenario where the base station is equipped with  $M$  antennas and the mobile terminal has a single antenna, STC can achieve  $M$ -

---

This research was supported by France Télécom and by the Institut Eurécom.

<sup>1</sup>Equivalently, we may think of a multi-cell system where no inter-cell processing is allowed because of practical complexity reasons and inter-cell interference is treated as noise.

fold *transmit diversity*, making block error probability decrease as  $O(\text{SNR}^{-M})$  for high SNR, that is,  $M$  times faster than in a single-antenna system.

Scheduling and STC are contrasting in both their assumptions and their goals. The former chooses the coding rate such that the selected user is never in outage, i.e., it can always decode its codeword with high probability. In order to do so, it must collect reliably and without delay the DRC signals from the users. On the contrary, the latter is geared to systems dominated by the information outage event. The rate is chosen a priori and no DRC feedback is required. Since in the current 3G standardization both approaches are left as options and might be used at the same time for downlink data transmission, it is natural to ask whether STC and scheduling are compatible. This issue has been investigated in a number of recent papers [2, 3, 6, 11] essentially showing that the *channel hardening* effect of transmit diversity is generally harmful for scheduling.

In this paper we consider the same question under more realistic conditions taking into account finite transmission buffers driven by packet arrival processes, and non-ideal DRC (e.g., due to delay in the DRC feedback loop). Deliberately, we limit our investigation to a TDMA-based system where one user is served at each given time, without any claim of optimality (we do claim practicality since the current system proposals indeed consider TDMA). For a  $K$ -user system, the base station has  $K$  queues. Each queue is characterized by an input arrival process. Following [7–9], we define the system stability region and study the system performance versus the transmit diversity order  $M$  (achieved by ideal ‘‘Gaussian’’ STC) under the adaptive scheduling policy that achieves all points in the stability region. In the case of infinite backlog, we study the impact of transmit diversity under the *Proportional Fair* (PF) scheduling currently considered for 1xEV-DO. We find that, under ideal DRC feedback, transmit diversity is generally detrimental for symmetric traffic users or infinite backlog, and might be beneficial only for finite buffers and very unbalanced traffic. On the other hand we show that, as the quality of the DRC feedback decreases, transmit diversity might provide a significantly larger stability region and may be beneficial also for symmetric traffic or infinite backlog.

## 2. System Model

We consider a base station with  $M$  antennas transmitting to  $K$  user terminals. The channel for user  $k$  in slot  $t$  is defined by

$$\mathbf{y}_k(t) = \mathbf{X}(t)\mathbf{h}_k(t) + \mathbf{z}_k(t) \quad (1)$$

where  $\mathbf{y}_k(t), \mathbf{z}_k(t) \in \mathbb{C}^{N \times 1}$  are the received and the noise signal in slot  $t$ ,  $N$  denotes the number of complex dimensions (channel uses) per slot,  $\mathbf{X}(t) \in \mathbb{C}^{N \times M}$  denotes the transmitted space-time codeword and  $\mathbf{h}_k(t) \in \mathbb{C}^M$  denotes

the  $M$ -input 1-output channel response for the user  $k$  channel in slot  $t$ , assumed time-invariant over each slot. The noise is complex circularly symmetric AWGN with i.i.d.  $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$  components. The base station has transmit peak power  $\gamma$  (energy per symbol), that is, the channel input constraint is  $\frac{1}{N}\text{tr}(\mathbf{X}(t)\mathbf{X}(t)^H) \leq \gamma$  for each codeword. Due to the noise variance normalization,  $\gamma$  takes on the meaning of maximum transmit SNR.

Coding and decoding is performed on a slot-by-slot basis. We assume that  $N$  is large enough such that good Gaussian-like code ensembles exist such that the block error probability for transmitting at rate  $R$  bit/channel use is given by the information outage probability  $\Pr(\log_2(1 + \beta_k(t)\gamma) \leq R)$  where we define  $\beta_k(t) = |\mathbf{h}_k(t)|^2$  and  $\beta(t) = (\beta_1(t), \dots, \beta_K(t))$ .

The base station has  $K$  transmission queues, where queue  $k$  is associated to user  $k$ . Each queue  $k$  is driven by an ergodic stationary input process  $A_k(t)$ , such that  $A_k(t)$  denotes the number of bits input to the queue buffer during slot  $t$ . The input arrival rate is defined by  $\lambda_k = \lim_{t \rightarrow \infty} \frac{1}{tN} \sum_{\tau=1}^t A_k(\tau)$  bit/channel use. The number of bits present in the queue buffer  $k$  in slot  $t$  is denoted by  $s_k(t)$ . At the beginning of each slot, a DRC signal  $\alpha(t) = (\alpha_1(t), \dots, \alpha_K(t))$  is revealed to the transmitter. A TDMA scheduling policy is defined by the sequence  $\{\mathbf{p}(t), \mathbf{R}(t)\}$ , where  $\mathbf{p}(t) = (p_1(t), \dots, p_K(t)) \geq 0$  such that  $\mathbf{p}(t) \in \mathcal{P}_K$  is a sequence of time-sharing parameters<sup>2</sup> and  $\mathbf{R}(t) = (R_1(t), \dots, R_K(t)) \in \mathbb{R}_+^K$  is a sequence of coding rates. In each slot  $t$ , the scheduling policy  $\mathcal{J}$  partitions the slot into disjoint sub-slots of length  $Np_1(t), \dots, Np_K(t)$  and let user  $k$  transmit on the  $k$ -th sub-slot with rate  $R_k(t)$  bit/channel use.

Under a given scheduling policy  $\mathcal{J}$ , the buffers evolve according to the stochastic difference equation

$$s_k(t+1) = \begin{cases} \max\{s_k(t) - Np_k(t)R_k(t), 0\} + A_k(t), & \text{if } \log_2(1 + \beta_k(t)\gamma) > R_k(t) \\ s_k(t) + A_k(t), & \text{else} \end{cases} \quad (2)$$

In writing (2), we have assumed that a simple ARQ retransmission protocol is used at the physical layer such that, if a decoding error occurs (information outage event), it is detected with probability 1 and the unsuccessfully decoded data is left in the transmission buffer and will be rescheduled for transmission in a later time.

We are interested in the *causal* scheduling policies  $\mathcal{J}^c$ , for which  $\{\mathbf{p}(t), \mathbf{R}(t)\}$  may depend on all DRC signal sequence and the buffer size history up to time  $t$ . Among

<sup>2</sup>We denote by  $\mathcal{P}_K$  the  $K$ -dimensional probability simplex and by  $V(\mathcal{P}_K)$  the set of its vertices. Therefore  $\mathbf{p} \in \mathcal{P}_K$  means that  $\mathbf{p}$  is a non-negative vector summing to 1, and  $\mathbf{e} \in V(\mathcal{P}_K)$  means that  $\mathbf{e}$  has a single non-zero component equal to 1. We denote by  $\mathbf{e}_k$  the vector in  $V(\mathcal{P}_K)$  with the  $k$ -th component equal to 1.

the class  $\mathcal{T}^c$ , we distinguish the nested subclasses of *memoryless* policies  $\mathcal{T}^m$ , for which  $\{\mathbf{p}(t), \mathbf{R}(t)\}$  is a (possibly random) function of  $\alpha(t), \mathbf{s}(t)$ , of stationary policies  $\mathcal{T}^s$ , that are memoryless policies such that the mapping  $\{\alpha(t), \mathbf{s}(t)\} \mapsto \{\mathbf{p}(t), \mathbf{R}(t)\}$  does not depend on  $t$ , and of 0-1 stationary policies  $\mathcal{T}^{s01}$ , that are stationary policies such that  $\mathbf{p}(t) \in V(\mathcal{P}_K)$ , i.e., the whole slot is assigned to a single user at each time  $t$ .

The average *service rate* of user  $k$  (in bit/channel use), under a given scheduling policy, is given by

$$\underline{\mu}_k = \liminf_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t p_k(\tau) R_k(\tau) \mathbb{1}\{\log_2(1 + \beta_k(\tau)\gamma) > R_k(\tau)\} \quad (3)$$

Notice that this does not coincide in general with the throughput (average spectral efficiency) of user  $k$ , since its buffer might be empty. We follow [7] and define the buffer overflow function

$$g_k(S) = \limsup_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \mathbb{1}\{s_k(\tau) > S\} \quad (4)$$

We say that the system is stable if, for all  $k$ ,  $\lim_{S \rightarrow \infty} g_k(S) = 0$ . We define the system stability region  $\Omega$  as the set of all arrival rates  $K$ -tuples  $\lambda \in \mathbb{R}_+^K$  such that there exists a causal policy for which the system is stable. Clearly, for the system defined above the main goal of a scheduling policy is to stabilize the system whenever  $\lambda \in \Omega$ . In the case of infinite backlog (no transmission buffers), the notion of stability region is meaningless. In this case, the main goal of a scheduling policy is to maximize the throughput subject to some fairness criterion. In this work we shall consider the PF scheduling policy [3], suitably modified in order to handle the case of non-ideal DRC feedback.

### 3. Main Results

We assume that  $\{\alpha(t), \beta(t)\}$  are jointly stationary and ergodic, with joint first-order pdf  $f_{\alpha, \beta}(\mathbf{a}, \mathbf{b})$ . We assume also the Markov chain condition  $\{\alpha(\tau) : \tau = 1, \dots, t-1\} \rightarrow \alpha(t) \rightarrow \beta(t)$ . Under these assumptions we have

**Proposition 1.** For any time-sharing sequence  $\mathbf{p}(t)$  that depends causally on the DRC signal  $\alpha(t)$ , the rate allocation policy maximizing the service rate (3) is given by

$$R_k^*(\mathbf{a}) = \sup_{R \geq 0} R \Pr(\log_2(1 + \beta_k \gamma) > R | \alpha = \mathbf{a}) \quad (5)$$

□

For later use, we define the conditional outage rate of user  $k$  given  $\alpha = \mathbf{a}$  as

$$R_{\text{out},k}(\mathbf{a}) = R_k^*(\mathbf{a}) \Pr(\log_2(1 + \beta_k \gamma) \geq R_k^*(\mathbf{a}) | \alpha = \mathbf{a}) \quad (6)$$

The function  $R^*$  is determined uniquely by the joint first-order pdf of  $\{\alpha(t), \beta_k(t)\}$  which is independent of  $t$  because of stationarity. In the case  $\alpha_k(t)$  is a one-to-one function of  $\beta_k(t)$  (perfect CSI), no outage event occurs and Proposition 1 yields

$$R_k^*(a) = R_{\text{out},k}(a) = \log_2(1 + \gamma a) \quad (7)$$

i.e., the usual AWGN capacity with channel gain  $a$ . Our second result gives an explicit expression for the system stability region.

**Proposition 2.** If  $\{\alpha(t), \beta(t)\}$  and  $\{A_1(t), \dots, A_K(t)\}$  are jointly stationary ergodic Markov processes, the system stability region is given by

$$\Omega = \bigcup_{\mathbf{p} \in \mathcal{T}^{s01}} \left\{ \lambda \in \mathbb{R}_+^K : \lambda_k \leq \mathbb{E}[p_k(\alpha) R_{\text{out},k}(\alpha)], \right. \\ \left. k = 1, \dots, K \right\} \quad (8)$$

It follows immediately from the convexity of  $\Omega$  that its boundary surface, denoted by  $\partial\Omega$ , is given by

**Proposition 3.**  $\partial\Omega$  is the convex closure of the points  $\lambda$  solution of

$$\max_{\lambda \in \Omega} \sum_{k=1}^K \theta_k \lambda_k \quad (9)$$

for all non-negative weight vectors  $\theta \in \mathcal{P}_K$ . □

The boundary surface can be obtained by maximizing over  $\mathbf{p} \in \mathcal{T}^{s01}$ , for all  $\theta \in \mathcal{P}_K$ , the functional  $\sum_{k=1}^K \theta_k \mathbb{E}[p_k(\alpha) R_{\text{out},k}(\alpha)]$ . The solution is readily given by

$$\mathbf{p}(\mathbf{a}) = \arg \max_{\mathbf{q} \in \mathcal{P}_K} \sum_{k=1}^K q_k \theta_k R_{\text{out},k}(\mathbf{a}) \quad (10)$$

which yields the solution  $\mathbf{p}(\mathbf{a}) = \mathbf{e}_{\hat{k}}$  where  $\hat{k} = \arg \max_k \theta_k R_{\text{out},k}(\mathbf{a})$ . We notice that the solution of (10) is a deterministic 0-1 policy. However,  $\partial\Omega$  might require time-sharing between such deterministic policies, i.e., it is generally achieved by random 0-1 policies in  $\mathcal{T}^{s01}$ , as it is expected from Proposition 2.

In practice, it is important to have a single adaptive policy that learns implicitly the arrival rates and guarantees stability whenever  $\lambda \in \Omega$ . Following the approach of [7] (see also [8, 9]), based on Liapunov stability, we can show:

**Proposition 4 (Max-Stability scheduling).** Under the joint stationary and ergodic Markov assumption of the channel state, DRC signal and arrival processes, the adaptive policy given by

$$\mathbf{p}(t) = \arg \max_{\mathbf{q} \in \mathcal{P}_K} \sum_{k=1}^K q_k \theta_k s_k(t) R_{\text{out},k}(\alpha(t)) \quad (11)$$

for any positive weight vector  $\theta \in \mathcal{P}_K$  achieves stability for all  $\lambda \in \Omega$ . Explicitly, (11) is given by  $\mathbf{p}(t) = \mathbf{e}_{\hat{k}}$  where  $\hat{k} = \arg \max_k \theta_k s_k(t) R_{\text{out},k}(\alpha(t))$ . □

The weighting vector  $\theta$  introduce priorities among the users and can be used to induce lower average delay for certain users while guaranteeing the system stability.

For the case of infinite backlog, we modify the PF scheduling policy in order to take into account the fact that DRC is not ideal, so that the requested rate might not be decodable. The modified PF policy is given as follows: for given  $\nu > 0$ , the time-averaged throughput of user  $k$  is given by the recursive equation

$$T_k(t+1) = \begin{cases} (1-\nu)T_k(t) + \nu p_k^{pf}(t) R_k^*(\alpha(t)), \\ \quad \text{if } \log_2(1 + \beta_k(t)\gamma) > R_k^*(\alpha(t)) \\ (1-\nu)T_k(t), \\ \quad \text{else} \end{cases} \quad (12)$$

where  $R_k^*(\mathbf{a})$  is given in Proposition 1 and  $\mathbf{p}^{pf}(t) = \mathbf{e}_{\hat{k}}$  with  $\hat{k} = \arg \max_k \frac{R_{out,k}(\alpha(t))}{T_k(t)}$ . Notice that  $R_{out,k}(\alpha(t))$  rather than  $R_k^*(\alpha(t))$  is used to assign the user. This corresponds to maximizing the *expected* instantaneous throughput normalized by the long-term average throughput.

#### 4. Examples

We evaluate the impact of transmit diversity on the TDMA downlink with scheduling under the max-stability scheduling (for finite backlog and given packet arrival processes) and under the modified PF scheduling (for infinite backlog).

**Setting.** We assume mutually independent arrival processes such that  $A_k(t) = \sum_{j=1}^{M_k(t)} b_{k,j}(t)$ , where  $M_k(t)$  is an i.i.d. Poisson distributed sequence that counts the number of packets arrived to the  $k$ -th buffer during slot  $t$  and  $b_{k,j}(t)$  are i.i.d. exponentially distributed random variables expressing the number of bits per packet. We take  $\mathbb{E}[b_{k,j}(t)] = N$ , so that  $\lambda_k = \mathbb{E}[M_k(t)]$ .

The fading  $\mathbf{h}_k(t)$  is a stationary ergodic Gaussian complex circularly symmetric vector process with i.i.d. components, such that  $\mathbf{h}_k(t) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \frac{1}{M}\mathbf{I})$ . We assume that the DRC feedback signal  $\alpha_k(t)$  is obtained from an estimation  $\mathbf{g}_k(t)$  of  $\mathbf{h}_k(t)$ , jointly Gaussian with  $\mathbf{h}_k(t)$ . Moreover, we assume that the pairs  $\{\mathbf{h}_k(t), \mathbf{g}_k(t)\}$  are mutually i.i.d. for different users. In particular, let  $\{\mathbf{h}_k(t)\}$  evolve in time according to Jake's autocorrelation model, where  $\mathbb{E}[\mathbf{h}_k(t)\mathbf{h}_k(t-\ell)^H] = \frac{r_\ell}{M}\mathbf{I}$ , with  $r_\ell = J_0(2\pi f_D T \ell)$  and  $f_D$  and  $T$  denoting the one-sided Doppler bandwidth (in Hz) and the slot duration (in seconds), respectively. We assume that the receivers can compute the DRC signal from exact past measurements of the channel  $\{\mathbf{h}_k(t-d), \dots, \mathbf{h}_k(t-d-n+1)\}$ , where  $d$  is the delay of the DRC feedback loop and  $n$  is the size of the observation window.

We define  $\mathbf{g}_k$  as the MMSE estimate of  $\mathbf{h}_k$  from  $\mathbf{h}_k(t-d), \dots, \mathbf{h}_k(t-d-n+1)$ . It is immediate to show that the

covariance of  $[\mathbf{h}_k^T, \mathbf{g}_k^T]^T$  is given by

$$\frac{1}{M} \begin{bmatrix} \mathbf{I} & (1-\sigma_e^2)\mathbf{I} \\ (1-\sigma_e^2)\mathbf{I} & (1-\sigma_e^2)\mathbf{I} \end{bmatrix}$$

where  $\sigma_e^2$  is the MMSE estimation error variance. It follows that  $\beta_k$  given  $\mathbf{g}_k$  is non-central chi-squared distributed with  $2M$  degrees of freedom, and its distribution depends only on  $\sigma_e^2$  (assumed to be known) and on  $|\mathbf{g}_k|^2$ . Hence, the DRC signal to be  $\alpha_k(t) = |\mathbf{g}_k(t)|^2$ . Notice that the Markov chain assumption  $\{\alpha(\tau) : \tau = 1, \dots, t-1\} \rightarrow \alpha(t) \rightarrow \beta(t)$  holds asymptotically for large  $n$ .

In our experiments we have considered mobile speeds 25km/h and 60km/h. By letting  $T = 1.67$  msec [4], the prediction order  $n = 8$ , the feedback delay  $d = 2$  slot, we obtain  $\sigma_e^2 = 0.05$  and  $0.40$ , respectively.

**Sum rate for infinite backlog.** The maximum sum-rate is given by explicitly by

$$\begin{aligned} \overline{R}_{out}(K, M) &= \mathbb{E}_{\alpha} \left[ \max_{k=1, \dots, K} R_{out,k}(\alpha_k) \right] \quad (13) \\ &= \int_0^\infty R_{out}(x) K \cdot \left( \frac{\left(\frac{Mx}{1-\sigma_e^2}\right)^M e^{-\frac{Mx}{1-\sigma_e^2}} x^{M-1}}{(M-1)!} \right) \\ &\quad \cdot \left( 1 - e^{-\frac{Mx}{1-\sigma_e^2}} \sum_{k=0}^{M-1} \frac{\left(\frac{Mx}{1-\sigma_e^2}\right)^k}{k!} \right)^{K-1} dx \end{aligned}$$

Fig. 1 shows the maximum sum-rate vs. a number of users for estimation error  $\sigma_e^2 = 0$  (ideal DRC) and  $\sigma_e^2 = 0.05, 0.40$  (non-ideal DRC). We observe that there exists a threshold  $K_{th}$  of the number of users above which transmit diversity becomes harmful. This depends heavily on the DRC quality. We have  $K_{th} = 2$  for ideal DRC, and  $K_{th} = 5, K_{th} > 20$  for non-ideal DRC with  $\sigma_e^2 = 0.05, 0.40$ , respectively.

Fig. 2 shows the sum-rate as a function of the number of users under the modified PF scheduling with  $\nu = 0.005$ . By introducing the fairness to balance the average throughput between users, multiuser diversity cannot be fully exploited. As a result, transmit diversity becomes slightly more useful or at least less harmful than with best-user scheduling. However, for practical values of  $\nu$ , the difference between best-user and PF scheduling is almost negligible.

**Two-user stability region.** Fig. 3 shows two-user stability region for  $\sigma_e^2 = 0.0, 0.40$ . For non-ideal DRC, transmit diversity enlarges the whole stability region including the maximal sum-rate point since  $\overline{R}_{out}(2, 1) < \overline{R}_{out}(2, M > 1)$ . On the other hand, for perfect CSI transmit diversity improves only the vertices of the stability region. This implies that transmit diversity is beneficial especially in asymmetric traffic conditions.

**Average buffer size.** We evaluated the impact of transmit diversity under max-stability scheduling in terms of the

average buffer size under both symmetric and asymmetric arrival processes. Here, the average buffer size denotes the time- and user-averaged buffer size expressed in bit. We consider a  $K = 20$  user system with channel estimation error  $\sigma_e^2 = 0.40$  (60km/h). Fig. 4 shows the average buffer size as a function of the sum arrival rate under symmetric arrivals, i.e.,  $\lambda_1 = \dots = \lambda_{20}$ . The buffers overflow when the sum arrival rate achieves the boundary of the stability region, which corresponds to the max sum-rate for the symmetric arrivals. Transmit diversity reduces the buffer size over the whole arrival rate range.

In the case of asymmetric arrivals, we partitioned the 20 users into four different classes with 5 users each, with arrival rates such that  $\lambda_2 = 8\lambda_1, \lambda_3 = 32\lambda_1, \lambda_4 = 64\lambda_1$ . Fig. 5 shows the average buffer size as a function of the sum arrival rate. We observe that the buffers diverge earlier than in the symmetric arrival case and that the gain due to transmit diversity is more significant for asymmetric arrivals especially close to at the boundary point.

## References

- [1] P.Viswanath, D.N.C.Tse, and R.Laroia, "Opportunistic Beamforming Using Dumb Antennas", *IEEE Trans. on Inform. Theory*, vol.48, no. 6, pp. 1277-1294, June 2002.
- [2] R.Gozali, R.M.Buehrer, and B.D.Woerner, "The Impact of Multiuser Diversity on Space-Time Block Coding", *IEEE Communications Letters*, vol.7, no. 5, pp. 213-215, May 2003.
- [3] J.Jiang, R.M.Buehrer, and W.H.Tranter, "Antenna Diversity in Multiuser Data Networks", *IEEE Trans. on Commun.*, vol.52, no.3, pp. 490-497, March 2004.
- [4] P.Bender, P.Black, M.Grob, R.Padovani, N.Sindhushayana, and A.Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users", *IEEE Commun. Mag.*, vol.38, pp. 70-77, July 2000.
- [5] R.Knopp and P.A.Humblet, "Information capacity and power control in single-cell multiuser communications", *Proc. IEEE ICC*, Seattle, WA, June 1995.
- [6] B.M.Hochwald, T.L.Marzetta, and V.Tarokh, "Multi-Antenna Channel Hardening and its Implications for Rate Feedback and Scheduling", *Submitted to IEEE Trans. on Inform. Theory*, May 2002.
- [7] M.J.Neely, E.Modiano, and C.E.Rohrs, "Power Allocation and Routing in Multibeam Satellites With Time-Varying Channels", *IEEE/ACM Transaction on Networking*, vol.11, no. 1, pp. 138-152, February 2003.
- [8] H.Boche and M.Wiczanski, "Stability Region of Arrival Rates and Optimal Scheduling for MIMO-MAC - A Cross-Layer Approach", *Proceeding of IZS*, Zurich, February 2004.
- [9] E.M.Yeh and A.S.Cohen, "Information Theory, Queueing, and Resource Allocation in Multi-user Fading Communications", *Proceedings of the 2004 Conference on Information Sciences and Systems* Princeton, NJ, March 2004.
- [10] M.Sharif and B.Hassibi, "On the Capacity of MIMO Broadcast Channel with Partial Side Information", *Submitted to IEEE Trans. on Inform. Theory*, 2003.
- [11] S.Borst and P.Whiting, "The Use of Diversity Antennas in High-Speed Wireless Systems: Capacity Gains, Fairness Issues", *Bell Laboratories Technical Memorandum*, 2001.
- [12] H. Weingarten, Y. Steinberg and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel", *Proceeding of ISIT 2004*, Chicago IL, USA, July 2004.
- [13] L.Li and A.J.Goldsmith, "Capacity and Optimal Resource Allocation for Fading Broadcast Channels-Part I: Ergodic Capacity", *IEEE Trans. on Inform. Theory*, vol.47, no. 3, pp. 1083-1102, March 2001.
- [14] "CDMA2000 High Rate Packet Data Air Interface Specification", TIA/EIA/3GPP2 Standard IS-856/3GPP2 C.S.0024, v3.0, Dec.2001.
- [15] L.H.Ozarow, S.Shamai, A.D.Wyner, "Information theoretic considerations for cellular mobile radio", *IEEE Transactions on Vehicular Technology*, vol.43, no. 2, pp. 359 - 378, May 1994.
- [16] L.Zheng and D.N.C.Tse, "Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels", *IEEE Trans. on Inform. Theory*, vol.49, no. 5, pp. 1073 - 1096, May 2003.

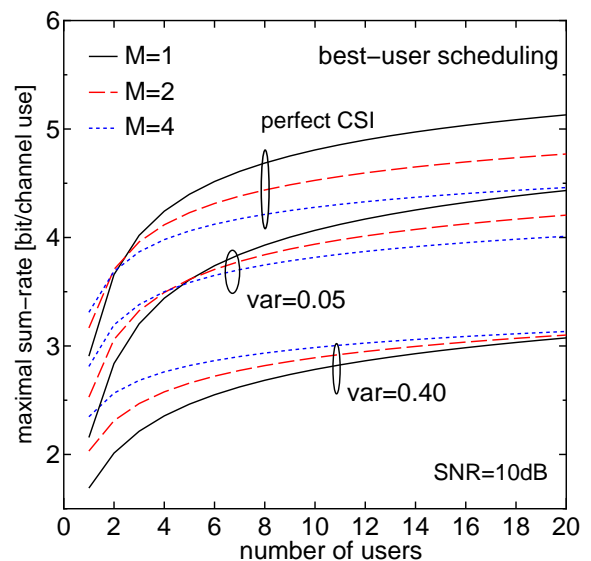


Figure 1: Max. sum-rate vs. number of users (best-user selection).

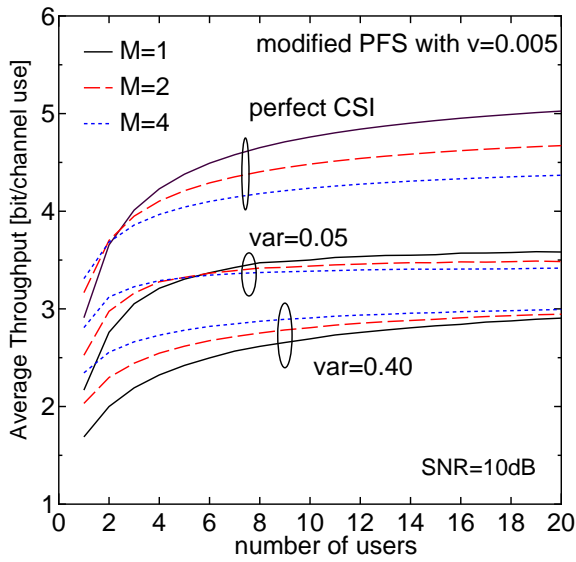


Figure 2: Max. sum-rate vs. number of users (PF scheduling).

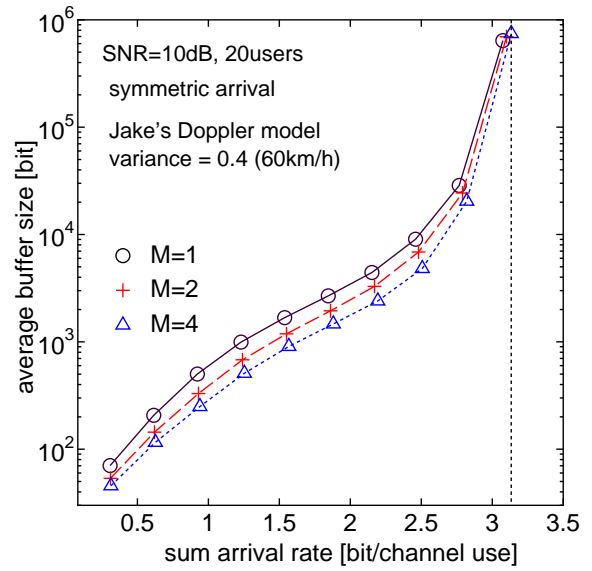


Figure 4: Average buffer size for symmetric arrivals,  $\sigma_e^2 = 0.40$ .

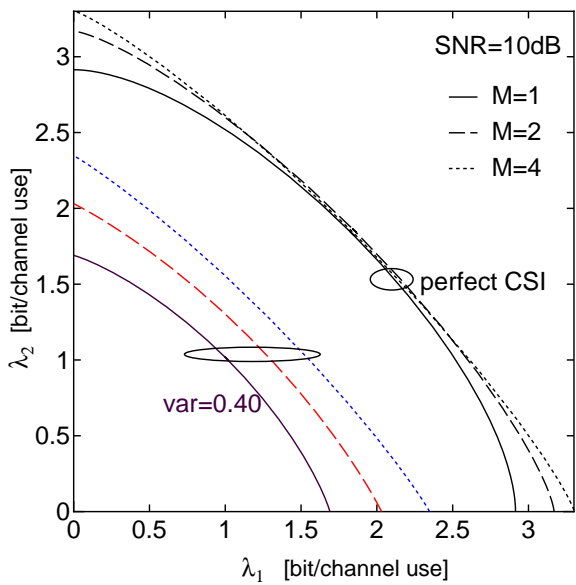


Figure 3: Stability region for  $\sigma_e^2 = 0.40$ .

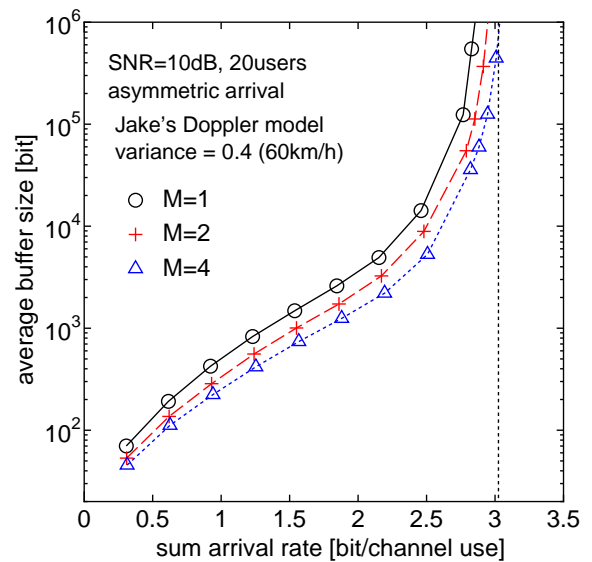


Figure 5: Average buffer size for asymmetric arrivals,  $\sigma_e^2 = 0.40$ .