Transmit Diversity vs. Opportunistic Beamforming in Data Packet Mobile Downlink Transmission

Mari Kobayashi, Giuseppe Caire, and David Gesbert *

Institut EURECOM,

2229 Route des Cretes, B.P. 193,

06904 SOPHIA ANTIPOLIS CEDEX, France

E-mail: firstname.name@eurecom.fr

April 22, 2005

*This research was supported by France Télécom and by the Institut Eurécom under the grant 424 87 987. The content of this paper was partially presented in ISITA'2004, Parma, Italy, in 2004.

Abstract

We compare space-time coding (transmit diversity) and random "opportunistic" beamforming in a SDMA/TDMA single-cell downlink system with random packet arrivals, correlated block-fading channels and non-perfect channel state information at the transmitter due to a feedback delay. Our comparison is based on system stability, under the adaptive policy that stabilizes the transmit queues whenever the arrival rates are in the system stability region. We compare the relative merit of transmit diversity with some previously proposed "opportunistic" beamforming schemes. The ability of accurately predicting the channel SNR dominates the performance of opportunistic beamforming. Hence, we propose to exploit synchronous pseudo-random beamforming matrices known a priori to the receivers in order to improve the channel state information quality. Even under this optimistic assumption, it clearly appears that for a given feedback delay the relative merit of opportunistic beamforming versus space-time coding strongly depends on the channel Doppler bandwidth. Therefore, previous naive conclusions on the fact that transmit diversity always hurts the system performance under multiuser-diversity scheduling should be taken with great care.

Key words : transmit diversity, multiuser diversity, opportunistic beamforming, stability, channel state information at transmitter

1 Introduction

We consider the system shown in Fig.1 where a base station with M antennas serves K users, each one equipped with a single antenna. Transmission is slotted and each slot comprises T channel uses (complex dimensions). Information bits arrive randomly at the transmitter and are locally stored into K queues, each associated to one user.

The base station operates in SDMA/TDMA mode: at each slot, $B \leq M$ *streams* of coded signal are generated by encoding packets of information bits from the K queues. Each stream is destined to one user. Hence, the system serves simultaneously B users at any slot. The B streams are transmitted by using some beamforming algorithm, that is generally referred to as the *signaling strategy*. For a given signaling strategy, the resource-allocation policy formed by queue selection (scheduling) and rate allocation is referred to as a SDMA/TDMA policy.

Several results on SDMA/TDMA downlink schemes exist in the literature. Driven by the information theoretic analysis of fading broadcast channels [1, 2], systems that serve the user enjoying the best instantaneous channel conditions have been proposed for high-data rate data packet downlink in evolutionary 3G system standardization [3–5]. When the base station has multiple antennas, random "opportunistic" beamforming has been proposed in [6, 7]. These systems generate $1 \leq B \leq M$ random time-varying beams, such that each user can measure the rate that can be reliably received on each beam and feeds back this information to the transmitter. The scheduling algorithm allocates the best user on each beam at any slot. The ability of a system to serve a user at its peak rate conditions is generally referred to as "multiuser diversity" [6–9].

A rather different use of the M antennas consists of improving the reliability of transmission for each user by space-time coding (STC) (see for example [10–12] and references therein). In this case, we have $B = 1$ (TDMA) since all antennas are used to provide *transmit diversity* to a single user. Scheduling can also be applied on top of STC and it is natural to investigate the relative merit of transmit diversity and multiuser diversity. Partial answers to this question have been provided, for example, in [13–15]. These works, as well as many others, indicate essentially that transmit diversity always *decreases* the throughput of a downlink system under SDMA/TDMA scheduling. In fact, multiuser diversity benefits from highly variable user channel conditions. This effect is amplified, in some sense, by random beamforming. On the contrary, transmit diversity tends to equalize the channel of each user as the number of antennas increases (the "channel hardening" effect [14]). Therefore, while the link of each user becomes individually more reliable, the system throughput is decreased.

The assumptions underlying these results are: 1) channel errors never occur, i.e., once a user is scheduled and is allocated a given (channel dependent) rate, the message will be successfully received with probability 1; 2) transmission queues have infinite buffers and all users are backlogged. Assumption 1) implies that the rate that can be supported by the channel can be perfectly known to the transmitter. Hence, it does not take into account the possibility of channel measurement errors and of feedback delay which causes the transmitter to allocate rates based on outdated information. It is obviously clear that, under this assumption, STC is of no help. In fact, STC is geared to cope with bad fading channel conditions unknown to the transmitter [16]. Assumption 2) implies that there are always bits available for transmission for any user at any slot. This, in turns, implies that the scheduling policy should try to maximize the system throughput subject to some *fairness* criterion. Unfortunately, "fairness" is not a well-defined concept and many criteria exist. Indeed, several different scheduling algorithms have been proposed that are optimal subject to different fairness criteria (see for example [6, 17–20] and references therein). The lack of a uniquely defined criterion to rank the different SDMA/TDMA policies makes this comparison very difficult.

Under random bits arrival, the notion of fairness is replaced by the notion of *stability* [21, 22]. In this setting, achieving any point in the system stability region (defined formally later) subsumes any reasonable fairness criterion and may be considered as the single most important goal of a downlink resource-allocation policy.

In this work, we build on the framework of [21] and find the SDMA/TDMA stability region under non-perfect channel state information at the transmitter (CSIT), that yields non-zero decoding error probability. In our case, non-perfect CSIT is essentially due to a delay in the feedback and to the fact that channels are time-varying. Then, we find the adaptive transmission policy that stabilizes the system for all arrival rates inside the system *stability region*. Finally, we apply the stability framework to provide a fair comparison between STC (transmit diversity) and random beamforming. The realistic assumption of non-perfect CSIT yields a non-trivial tradeoff between the transmit diversity achieved by STC and the multiuser diversity achieved by opportunistic beamforming. This tradeoff is missed by the analysis of [6, 7, 13–15, 17–20] that neglects the fact that decoding errors may occur with non-zero probability. In our more refined and realistic setting, the ability of accurately predicting the channel SNR clearly emerges as one of the main limiting factors of multiuser diversity-based schemes. Therefore, we also propose an improvement of the random beamforming where each user is synchronized with a common random number generator that produces beamforming matrices. Hence, the matrices can be considered *a priori* known[1] and can be used for channel estimation. This makes the rate of variation of the random beams and the ability of estimating the channels unrelated, whereas the latter depends only on the physical channel rate of variation.

Even with the proposed scheme, that represents a "best case" for opportunistic beamforming, our results show that for a given feedback delay the relative merit of opportunistic beamforming versus STC strongly depends on the channel Doppler bandwidth. In particular, for slowly-varying channels the opportunistic beamforming with $B = M$ achieves the best average delay, while for faster channels STC is best. In light of these results, the utility of random beamforming with $B = 1$ is fully questionable.

[1]Notice that in practical CDMA systems each user is synchronized with the random spreading/scrambling code of the base station. Therefore, this assumption is fully realistic.

2 System model and definition of stability

We assume a frequency non-selective block-fading channel where the signal received at user k terminal in slot t is given by

$$\mathbf{y}_k(t) = \mathbf{X}(t)\mathbf{h}_k(t) + \mathbf{w}_k(t) \tag{1}$$

where $\mathbf{X}(t) \in \mathbb{C}^{T\times M}$ is the transmitted codeword, $\mathbf{h}_k(t) \in \mathbb{C}^{M\times 1}$ denotes the M-input 1-output channel response, assumed constant in time and frequency over each slot, and $\mathbf{w}_k(t) \in \mathbb{C}^{T\times 1}$ is a complex circularly symmetric AWGN with components $\sim \mathcal{CN}(0,1)$. The base station has fixed transmit energy per channel use denoted by γ, that is, $\mathrm{tr}(\mathbf{X}(t)\mathbf{X}(t)^H) \leq \gamma T$, for all t. Due to the noise variance normalization, γ takes on the meaning of *transmit* SNR.

Coding and decoding is performed on a slot-by-slot basis. We assume that T is large enough such that powerful codes exist whose error probability is characterized by a threshold behavior: letting $R(t)$ denote the transmission rate on a given slot t and $\overline{R}(t)$ denote the supremum of the coding rates supported by the channel, which is a random variable because of fading, the decoding error probability in slot t is equal to $\mathbf{1}\{\overline{R}(t) \leq R(t)\}$, i.e., the decoder makes an error with very large probability if the transmission rate is above the maximum achievable rate of the channel, while error probability is negligibly small if it is below.[2] We shall refer to the probability $\mathbb{P}(\overline{R}(t) \leq R(t))$ as the *outage probability.* In order to handle decoding errors, we assume an ideal ARQ protocol such that the unsuccessfully decoded information bits are left in the transmission buffer and shall be re-scheduled for transmission at a later time.

A SDMA/TDMA policy is generally a function of the Channel State Information at the Transmitter (CSIT) and of the state of the transmitter queues. CSIT can be obtained in

[2] This threshold behavior of block error probability can be approached in practice by concatenated coding schemes such as inner trellis codes with Reed-Solomon outer coding, turbo codes and low-density parity-check codes, and it is closely related to the concept of information outage probability [23] and of ϵ-capacity, or outage capacity [24], which are widely used in the information-theoretic analysis of block-fading channels.

several ways depending on the system. For the systems considered in this work, specific idealized models for CSIT and for the signaling strategies will be specified in Section 4. In general, we assume that at the beginning of each slot t, both a CSIT signal $\boldsymbol{\alpha}(t) \in \mathcal{A}$ and the queue buffer states are revealed to the transmitter, where $\mathcal{A}$ denotes the CSIT signal alphabet. The arrival process of queue k is denoted by $A_k(t)$, an ergodic process with *arrival rate* $\lambda_k \triangleq \frac{1}{T}\mathbb{E}[A_k(t)]$ (bit/channel use). The buffer size of queue k is denoted by $S_k(t)$ (bit). We let $\boldsymbol{\alpha}_1^t = \{\boldsymbol{\alpha}(\tau) : \tau = 1, \ldots, t\} \in \mathcal{A}^t$ denote the sequence of CSIT signals up to time t, and $\mathbf{S}_1^t = \{S_1(\tau), \ldots, S_K(\tau) : \tau = 1, \ldots, t\} \in \mathbb{R}_+^{Kt}$ denote the queue buffer state sequence up to time t. An SDMA/TDMA resource allocation policy is formally defined by two sequences of functions

$$\begin{aligned} \mathbf{P}^{(t)} &: \mathcal{A}^t \times \mathbb{R}_+^{Kt} \rightarrow [0,1]^{K \times B} \\ \mathbf{R}^{(t)} &: \mathcal{A}^t \times \mathbb{R}_+^{Kt} \rightarrow \mathbb{R}_+^{K \times B} \end{aligned} \tag{2}$$

for $t = 1, 2, \ldots$, such that each user k on each slot t is given a fraction $p_{k,j}^{(t)}(\boldsymbol{\alpha}_1^t, \mathbf{S}_1^t)$ of dimensions of stream $j = 1, \ldots, B$, and transmits at rate $R_{k,j}^{(t)}(\boldsymbol{\alpha}_1^t, \mathbf{S}_1^t)$ bit/channel use. The function $\mathbf{P}^{(t)}$ must satisfy the SDMA/TDMA feasibility constraint

$$\sum_{k=1}^{K} p_{k,j}^{(t)} \leq 1 \tag{3}$$

for each $j = 1, \ldots, B$ and for all t. Notice also that the functions $\mathbf{P}^{(t)}$ and $\mathbf{R}^{(t)}$ depend *causally* on the CSIT and queue state sequences.

For a given signaling strategy and SDMA/TDMA policy $\{\mathbf{P}^{(t)}, \mathbf{R}^{(t)}\}$, let $\overline{R}_{k,j}(t)$ denote the supremum of the coding rates supported by the channel for user k on stream j in slot t. Notice that $\overline{R}_{k,j}(t)$ depends on the instantaneous channel realization, and is therefore a random variable. Under the assumptions given above, the queue buffer states evolve in time according to the stochastic difference equation

$$S_k(t+1) = \left[S_k(t) - T \sum_{j=1}^{B} p_{k,j}^{(t)} R_{k,j}^{(t)} \, \mathbf{1}\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\} \right]_+ + A_k(t) \tag{4}$$

for all $k = 1, \ldots, K$, where $[\cdot]_+ \triangleq \max\{\cdot, 0\}$. The presence of the indicator function $\mathbf{1}\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$ is due to the ARQ mechanism illustrated before: only if decoding is successful, i.e., if the scheduled rate $R_{k,j}^{(t)}$ is indeed achievable, the corresponding information bits are removed from the transmission buffer.

In order to define stability, we follow [21] and define the buffer overflow function $g_k(S) = \limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} \mathbf{1}\{S_k(\tau) > S\}$. We say that the system is stable if $\lim_{S\to\infty} g_k(S) = 0$ for all $k = 1, \ldots, K$. For a given signaling strategy, we define the system SDMA/TDMA stability region Ω as the set of all arrival rates K-tuples $\boldsymbol{\lambda} \in \mathbb{R}_+^K$ such that there exists a feasible policy in the form (2) for which the system is stable.

3 Stability region and max-stability adaptive policy

In this section we make some key simplifying assumptions that allow us to obtain a very simple characterization of Ω and a simple adaptive SDMA/TDMA policy that achieves any $\boldsymbol{\lambda} \in \Omega$. We assume that: 1) the channel vectors $\{\mathbf{h}_k(t) : k = 1, \ldots, K\}$, the CSIT signal $\boldsymbol{\alpha}(t)$ and the arrival processes $\{A_k(t) : k = 1, \ldots, K\}$ evolve according to a jointly stationary ergodic Markov process; 2) the arrival processes have $\mathbb{E}[A_k(t)^2] \leq \infty$; 3) for a given signaling strategy, for every sufficiently large t, the following Markov chain holds:

$$\boldsymbol{\alpha}_1^{t-1} \to \boldsymbol{\alpha}(t) \to \{\overline{R}_{k,j}(t) : k = 1, \ldots, K, j = 1, \ldots, B\} \tag{5}$$

In particular, the third condition implies that for all k, j and sufficiently large t, the conditional outage probability depends only on the current CSIT value, i.e., it satisfies

$$\mathbb{P}(\overline{R}_{k,j}(t) \leq R | \boldsymbol{\alpha}_1^t) = \mathbb{P}(\overline{R}_{k,j}(t) \leq R | \boldsymbol{\alpha}(t)) \tag{6}$$

We can cast the stability problem for the general SDMA/TDMA system into a stability-wise equivalent, but simpler, problem for which the results of [21] are almost immediately applicable. In a new (virtual) system, $\boldsymbol{\alpha}(t)$ takes on the role of the channel state, and the instantaneous *effective* rate $R_{k,j}^{(t)}\mathbf{1}\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$ is replaced by the conditional outage rate, defined by

$$R_{k,j}^{\rm out}(\mathbf{a}) = \sup_{R\geq 0} \quad R(1 - \mathbb{P}(\overline{R}_{k,j} \leq R | \boldsymbol{\alpha} = \mathbf{a})) \tag{7}$$

which is a function of the current CSIT value $\boldsymbol{\alpha} = \mathbf{a} \in \mathcal{A}$.[3] In Appendix A we prove:

Theorem 1 [Stability region]. Under the above system assumptions, for any fixed signaling strategy, the system stability region is given by

$$\Omega = \text{coh} \bigcup_{\mathbf{P}\in\mathcal{P}} \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^K \quad : \quad \lambda_k \leq \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\rm out}(\boldsymbol{\alpha})\right] \right\} \tag{8}$$

where "coh" means *closure of the convex hull* and where $\mathcal{P}$ is the set of *stationary* SDMA/TDMA policies $\mathbf{P} : \mathcal{A} \to [0,1]^{K,B}$ that map the current CSIT $\boldsymbol{\alpha} = \mathbf{a} \in \mathcal{A}$ into an array $[p_{k,j}(\mathbf{a})]$ satisfying (3). This stability region is achieved by the stationary rate allocation policy $\mathbf{R}^\star$, defined by

$$R_{k,j}^\star(\mathbf{a}) = \arg \sup_{R\geq 0} \quad R(1 - \mathbb{P}(\overline{R}_{k,j} \leq R | \boldsymbol{\alpha} = \mathbf{a})) \tag{9}$$

□

The following comments are in order: 1) Under the conditions assumed in this paper, we have that the rate allocation function $\mathbf{R}^\star$ in (9) is optimal for any stationary $\mathbf{P} \in \mathcal{P}$ and any SDMA/TDMA signaling scheme. This reduces the problem of the stability-optimal resource allocation policy to the determination of $\mathbf{P} \in \mathcal{P}$ alone; 2) under any stationary policy $(\mathbf{P}, \mathbf{R}^\star)$, the user average service rates $\widetilde{\mu}_k(t) = T \sum_{j=1}^{B} p_{k,j}(\boldsymbol{\alpha}(t)) R_{k,j}^{\rm out}(\boldsymbol{\alpha}(t))$ (bit/slot) are linear functions of $\mathbf{P} \in \mathcal{P}$ and $\mathcal{P}$ is a convex set. It follows that no convex-hull operation is needed in (8); 3) we can introduce the class $\mathcal{P}_{\rm on-off}$ of *randomized* on-off SDMA/TDMA policies such that for every $\mathbf{P} \in \mathcal{P}$, there exists $\mathbf{P}' \in \mathcal{P}_{\rm on-off}$ defined as follows: for all $\mathbf{a} \in \mathcal{A}$ and $j = 1, \ldots, B$, let user κ_j be served on the whole slot on stream j where $\kappa_j \in \{1, \ldots, K\}$ is a random variable generated according to the probability mass function $(p_{1,j}(\mathbf{a}), \ldots, p_{K,j}(\mathbf{a}))$.

[3]Here and in the following we let $\boldsymbol{\alpha}$ denote a random vector over $\mathcal{A}$ with the same distribution of $\boldsymbol{\alpha}(t)$, which is independent of t because of stationarity.

Clearly, Ω is achieved by restricting the union to the policies in $\mathcal{P}_{\text{on-off}}$. In practice, randomized on-off policies are preferable since handling a single user per slot per stream is much easier. Therefore, we shall restrict to these policies in the following; 4) from the convexity of Ω, the region boundary $\partial\Omega$ can be obtained by letting $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K) \in \mathbb{R}_+^K$ and finding $\lambda_k(\boldsymbol{\theta}) = \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\text{out}}(\boldsymbol{\alpha})\right]$, for the policy $\mathbf{P} \in \mathcal{P}$ solution of the maximization problem

$$\max_{\mathbf{P} \in \mathcal{P}} \sum_k \theta_k \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\text{out}}(\boldsymbol{\alpha})\right] \tag{10}$$

Then, $\partial\Omega$ is the convex upper envelope of the points $\{\boldsymbol{\lambda}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \mathbb{R}_+^K, \sum_k \theta_k = 1\}$.

For given values $\boldsymbol{\theta} \in \mathbb{R}_+^K$, the solution of (10) is readily obtained as

$$\widehat{p}_{k,j}(\mathbf{a}) = \begin{cases} 1 & k = \arg\ \max_{k'} \theta_{k'} R_{k',j}^{\text{out}}(\mathbf{a}) \\ 0 & k \neq \arg\ \max_{k'} \theta_{k'} R_{k',j}^{\text{out}}(\mathbf{a}) \end{cases} \tag{11}$$

This means that, as expected, the points on $\partial\Omega$ are achieved by time-sharing (convex-hull operation) of *deterministic* on-off policies that schedule the best user, given by $k = \arg\ \max_{k'} \theta_{k'} R_{k',j}^{\text{out}}(\mathbf{a})$, on each stream j. Notice that time-sharing of deterministic on-off policies yields the class of randomized on-off policies $\mathcal{P}_{\text{on-off}}$, as expected.

For given $\boldsymbol{\lambda} \in \Omega$ there exists some memoryless stationary policy $\mathbf{P}_{\boldsymbol{\lambda}} \in \mathcal{P}_{\text{on-off}}$ that stabilizes the system. However, in order to determine $\mathbf{P}_{\boldsymbol{\lambda}}$ the a priori knowledge of $\boldsymbol{\lambda}$ is generally required. This might not be available in practice. Hence, a policy that achieves stability for all $\boldsymbol{\lambda} \in \Omega$ *adaptively* (i.e., without prior knowledge of the arrival rates) is of great practical interest [21, 22]. We shall refer to this policy as the "max-stability" adaptive policy. We have,

Theorem 2 [Max-stability adaptive policy]. Under the above system assumptions, for any fixed signaling strategy, the max-stability adaptive policy is given by

$$\widehat{p}_{k,j}(\mathbf{a}, \mathbf{S}) = \begin{cases} 1 & k = \arg\ \max_{k'} \theta_{k'} S_{k'} R_{k',j}^{\text{out}}(\mathbf{a}) \\ 0 & k \neq \arg\ \max_{k'} \theta_{k'} S_{k'} R_{k',j}^{\text{out}}(\mathbf{a}) \end{cases} \tag{12}$$

for any strictly positive weights $\theta_k > 0$. □

Theorem 2 follows by applying the theory of Lyapunov drift [21,22] and follows closely the proof of Theorem 3 in [21]. It is omitted for the sake of space limitation. The max-stability adaptive policy allocates on each stream j in slot t the user maximizing the weighted outage rate $\theta_{k'} S_{k'}(t) R^{\rm out}_{k',j}(\boldsymbol{\alpha}(t))$. The *adaptive* nature of the max-stability policy is evident from the fact that $\widehat{\mathbf{P}}$ is a memoryless stationary function of the current CSIT $\boldsymbol{\alpha}(t)$ and of the current queues state $\mathbf{S}(t)$, rather than of $\boldsymbol{\alpha}(t)$ alone. Intuitively, the buffer state represents an empirical observation of the arrival rate.

4 Space-time coding vs. opportunistic beamforming

In this section we apply the max-stability adaptive policy obtained before to STC (transmit diversity) and random opportunistic beamforming. We assume that the channel vectors $\mathbf{h}_k(t)$ are mutually statistically independent for different index k and i.i.d. for different antennas. Focusing without loss of generality on a scalar channel coefficient $h(t)$ where we drop the user and antenna index because of the i.i.d. assumption, we assume that $h(t)$ evolves from slot to slot according to a stationary ergodic L-order Gauss-Markov process, given by

$$h(t) = -\sum_{\ell=1}^{L} g_\ell h(t-\ell) + \nu(t) \tag{13}$$

where $\nu(t) \sim \mathcal{CN}(0, \sigma^2)$ is an i.i.d. process. We make the optimistic assumption that the receivers can estimate *exactly* their channel vector and feed back the corresponding value without distortion (unquantized and noiseless). However, due to a fixed delay of d slots in the feedback link, the CSIT is given by $\boldsymbol{\alpha}(t) = \{\mathbf{h}_k(t-\ell-d) : k = 1, \ldots, K, \ell = 0, \ldots, L-1\}$. We hasten to say that this CSIT model is just a *convenient idealization* for which the assumptions yielding the stability region and max-stability policy obtained in Section 3 hold exactly. In practice, many other sources of uncertainty are present, such as noisy channel observations [25], a rate-constrained feedback link [3,26], or an unquantized but noisy feedback link [25]. As a matter of fact, for a fixed signaling strategy and a fixed delay d

in the feedback, any *degraded version* of $\boldsymbol{\alpha}(t)$ defined above can only worsen the performance. This is an immediate consequence of the Markov nature of the channel processes. Hence, our assumption yields an optimistic scenario in favor of the opportunistic schemes.

Notice also that we consider only fixed signaling strategies that do not make full use of the CSIT $\boldsymbol{\alpha}(t)$. In fact, the goal here is not to find the best possible signaling strategy for a given type of CSIT, but compare given fixed strategies under the max-stability policy. It is clear that neither opportunistic beamforming nor STC would be the best strategy under the assumed CSIT if one has full freedom of optimizing the system with respect to the signaling strategy. We compare the following signaling strategies.

Space Time Coding. In this case, $\mathbf{X}(t) \in \mathbb{C}^{T \times M}$ denotes the transmitted space-time codeword. One user is served in each slot, i.e., $B = 1$ (in this case we drop the stream index j for notation simplicity). The system cannot exploit spatial multiplexing since the user terminals have only one antenna. Hence, STC yields only M-fold transmit diversity. The supremum of the rates supported by the channel under any coding scheme is given by the channel instantaneous capacity $\overline{R}_k(t) = \log_2\left(1 + \frac{\gamma}{M}|\mathbf{h}_k(t)|^2\right)$. The conditional outage probability is given by

$$\mathbb{P}(\overline{R}_k(t) \leq R|\boldsymbol{\alpha}(t) = \mathbf{a}) = 1 - \mathcal{Q}_M\left(\sqrt{\frac{2|\mathbf{g}_k(t)|^2}{\sigma_e^2}}, \sqrt{\frac{2M(2^R - 1)}{\gamma\sigma_e^2}}\right) \tag{14}$$

where $\mathcal{Q}_M$ denotes generalized Marcum's Q function [27], σ_e^2 denotes the prediction MMSE [4] of $\mathbf{h}_k(t)$ from $\boldsymbol{\alpha}(t)$ and where $\mathbf{g}_k(t)$ is the MMSE predictor of $\mathbf{h}_k(t)$ from $\boldsymbol{\alpha}(t)$. Expression (14) is easily obtained by using the fact that $\mathbf{h}_k(t)$ and $\boldsymbol{\alpha}(t)$ are jointly Gaussian. In a practical implementation, each user sends back (a suitably quantized version of) the estimated channel gain $|\mathbf{g}_k(t)|^2/M$. Hence, the amount of feedback required by this scheme is very small.

Improved Opportunistic Beamforming. We consider opportunistic beamforming using $1 \leq B \leq M$ mutually orthogonal pseudorandom beams, as proposed in [6, 7]. The

[4]We define $\sigma_e^2 = \mathbb{E}[|h(t) - \mathbb{E}[h(t)|h(t-d), \ldots, h(t-d-L+1)]|^2]$, which is the same for all components of $\mathbf{h}_k(t)$ and all users, under the symmetry assumptions considered.

transmitted signal is given by

$$\mathbf{X}(t) = \sum_{j=1}^{B} \mathbf{s}_j(t)\boldsymbol{\phi}_j^T(t) \tag{15}$$

where $\mathbf{s}_j(t) \in \mathbb{C}^{T\times 1}$ is the signal associated to beam (stream) j, $\boldsymbol{\phi}_j(t) \in \mathbb{C}^{M\times 1}$ is the beamforming vector for beam j in slot t, and it is assumed that $\boldsymbol{\phi}_j^H(t)\boldsymbol{\phi}_m(t) = \delta_{j,m}$. The signal-to-interference plus noise ratio (SINR) of user k in beam j is equal to

$$\mathsf{SINR}_{k,j}(t) = \frac{|\boldsymbol{\phi}_j^T(t)\mathbf{h}_k(t)|^2}{B/\gamma + \sum_{m\neq j} |\boldsymbol{\phi}_m^T(t)\mathbf{h}_k(t)|^2} \tag{16}$$

Assuming user codes drawn from an i.i.d. Gaussian distribution (or making a Gaussian approximation of interference), the supremum of the rates supported by beam j for user k is given by $\overline{R}_{k,j}(t) = \log_2(1 + \mathsf{SINR}_{k,j}(t))$. In the schemes analyzed in [6,7], each user measures its SINR for each of the B transmit beams and feeds back its best SINR and the index of the beam achieving it. The SINR values are considered instantaneously and perfectly known to the transmitter. This is clearly not a realistic assumption. Moreover, in these works it is not clear how the SINRs are estimated and at which rate the random beamforming matrix changes. The next result, proved in Appendix B, establishes a simple relationship between the average transmission delays and the speed of variation of the random beamformer, in the ideal case of perfect CSIT (zero outage probability). By Little's theorem [28] we let $\mathbb{E}[D_k] = \frac{\mathbb{E}[S_k]}{T\lambda_k}$ denote average delay of user k expressed in slots.

Theorem 3 [Bound on the average delay]. Assume that CSIT is perfect, i.e., $R_{k,j}^{\text{out}}(t) = \overline{R}_{k,j}(t)$, and that the channels change independently every N slots. Then, the average user delays $\mathbb{E}[D_k]$ obtained by the max-stability adaptive policy satisfies

$$\sum_k \theta_k \lambda_k \mathbb{E}[D_k] \leq \frac{N\mathcal{K}}{2\delta} \tag{17}$$

where $\mathcal{K} = \sum_{k=1}^{K} \theta_k \mathbb{E}\left[\left(\frac{A_k(t)}{T}\right)^2\right] + B\sum_{k=1}^{K}\theta_k\sum_{j=1}^{B}\mathbb{E}\left[(\overline{R}_{k,j}(t))^2\right]$, and where $\delta > 0$ indicates the distance of the arrival rate vector from the stability region boundary. □

Under perfect CSIT and static physical channels, the average delays are bounded by a function that increases linearly with N, the number of slots over which the combination of physical channels and random beamformers remain constant. Hence, it is clear that under perfect CSIT it is convenient to make the random beamformers vary as fast as possible (i.e., change independently at each slot). In practice, however, if the beamformers change too rapidly the users are not able to reliably estimate their SINRs. We conclude that, under any realistic scheme to obtain CSIT, there must be a tradeoff between the goal of letting the channels rapidly varying and the need of estimating the SINRs accurately.

Motivated by this consideration, we propose the following improvement: each user in the system is synchronized with a common random number generator that generates the random beamforming matrices. Hence, the matrices can be considered *a priori* known by all users logged into the system. Moreover, since they are unitary, they have no impact on the estimation of the underlying physical channel that can be achieved by usual pilot-aided schemes. In this way, the rate of variation of the random beams is independent of the ability of estimating the channels, that depends uniquely on the physical channel Doppler bandwidth. In the numerical results in Section 5 , we let $N = 1$, which can be regarded as a best case for the average transmission delay.

By using standard methods of characteristic functions of Hermitian quadratic forms of Gaussian random variables [29] it is possible to compute numerically the conditional outage probability, which is needed to compute the instantaneous rate request $R^{\star}_{k,j}(t)$ and the corresponding outage rate. It can be shown that this probability depends on $\boldsymbol{\alpha}(t)$ only through the two real numbers $|\boldsymbol{\phi}_j^T \mathbf{g}_k(t)|^2$ and $\sum_{m \neq j} |\boldsymbol{\phi}_m^T \mathbf{g}_k(t)|^2$, where again $\mathbf{g}_k(t)$ denotes the MMSE prediction of $\mathbf{h}_k(t)$ from $\boldsymbol{\alpha}(t)$. The details of this calculation are given in [30]. In a practical implementation, each user feeds back (a suitably quantized version of) the B real numbers $\{|\boldsymbol{\phi}_j^T \mathbf{g}_k(t)|^2 : j = 1, \ldots, B\}$, from which the max-stability policy can be computed. Hence, also in this case the amount of feedback required by this scheme is moderately small.

5 Numerical Results

Simulation setting. We considered 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 at the beginning of slot t and $\{b_{k,j}(t)\}$ are i.i.d. exponentially distributed random variables expressing the number of bits per packet. We take $\mathbb{E}[b_{k,j}(t)] = T$ ($T = 2000$ in our simulations), so that λ_k coincides with the average number of packets arrived in a slot (T channel uses). We considered a Gauss-Markov process of order $L = 5$ where the coefficients in (13) are chosen in order to approximate (see [31,32]) Jakes' autocorrelation model [33], typical of wireless mobile channels. Inspired by the HDR system [3], we let the duration of a slot be 1.67 msec, and the feedback delay $d = 2$ slots. Under this setting, the mobile speeds $v = 0$, 25, 40, 60, 80 km/h yield a channel prediction MMSE $\sigma_e^2 = 0.00$, 0.05, 0.10, 0.40, 0.60 respectively. The average SNR is set equal to $\gamma = 10$ dB. For opportunistic beamforming, we generate a new independent set of random beamforming vectors in every slot that are assumed to be known a priori by the users as explained before.

Maximum sum rate. We evaluate the maximum sum rate of STC and opportunistic beamforming. Since the maximum sum-rate is given by the intersection between the boundary of the stability region and the symmetric arrival vector $\lambda_1 = \cdots = \lambda_K$, this allows us to know exactly the total arrival rate where the buffer size diverges under the max-stability policy. The maximum sum-rate is obtained by scheduling at each time and on each stream the user with the largest outage rate, irrespectively of the buffers state.

Figs. 2 and 3 show the maximum sum-rate vs. the number of users for mobile speed $v = 0$ km/h and $v = 25, 60$ km/h by using STC and opportunistic beamforming respectively. In Fig. 3 we let $B = M$. The case $M = 1$ in both figures is the same and coincides also with the performance of the opportunistic single-beamforming [6] with $M > 1$ and $B = 1$.

In Fig. 2 we observe that number of users after which transmit diversity becomes harmful depends heavily on the CSIT quality. We have $K = 2$ for ideal CSIT ($\sigma_e^2 = 0$), and $K = 5, 28$ for $\sigma_e^2 = 0.05, 0.40$, respectively. In Fig. 3 we observe large gain with $M = 2, 4$ beams especially for $K \geq 10, 15$, for the case of perfect CSIT. Unfortunately, the advantage of multiple beams decreases dramatically as the quality of the CSIT worsens. For $\sigma_e^2 = 0.40$, multiple beams seem to be harmful even for a large number of users in the system, i.e., multiuser diversity has completely disappeared.

Impact of the beamforming variation speed on delay. We consider opportunistic beamforming with $B = M = 4$ over a static channel for a system with 50 users with symmetric arrival. Fig. 4 shows the average delay, given by $\sum_k \mathbb{E}[D_k]/K$, as well as its upper bound derived in Theorem 3 for different N. The upper bound becomes relatively tighter as the arrival rate gets closer to the stability region boundary. Clearly, at arrival rates between zero and the stability boundary, $N = 1$ yields the smallest average delay.

Average delay vs. mobile speed. We evaluated the average delay of STC and opportunistic beamforming as a function of the mobile speed by letting the total arrival rate fixed (to 2.5 bit/channel use in this case). We consider the symmetric arrival case. Figs. 5,6 and 7 show the average delay vs. the mobile speed for a system with 50 users with STC, random beamforming with $B = 1$ and random beamforming with $B = M$, respectively. Clearly, the case $M = 1$ is the same in all three figures and refers to a standard single-antenna system.

For a very slowly-varying channels (close to $v = 0$ km/h) the STC system becomes non-ergodic and there is a positive probability of buffer overflow. This probability is reduced by increasing transmit diversity, thanks to channel-hardening effect: ergodicity which is lost in time is eventually recovered in the spatial domain by increasing the number of transmit antennas.

As seen from Figs. 6 and 7, opportunistic random beamforming decreases the average delay by making the channel vary almost i.i.d.. When the channel is slow (below 40 km/h), opportunistic beamforming with M beams achieves the smallest delay. As v increases (i.e., the quality of CSIT becomes worse), STC outperforms random beamforming due to its better outage rate. Interestingly, the opportunistic beamforming systems become unstable (the average delay diverges) with $M = 2, 4$ and v larger than 60 km/h. This means that at this speed users are essentially allocated on the wrong beam with high probability. It is also noticed that, with the parameters of this simulation, the M-antenna $B = 1$ random beamforming scheme proposed in [6] is outperformed by the $B = M$ random beamforming scheme for low mobile speed and by STC for larger mobile speed. Therefore, it is never useful.

6 Conclusions

We have compared two simple SDMA/TDMA schemes for downlink transmission in a mobile wireless system where the base station has multiple antennas and the user terminals have one antenna: space-time coding and random "opportunistic" beamforming. Beyond their simplicity, these schemes are also relevant since they are currently considered for standardization in evolutionary 3G systems. Our comparison is made under a general max-stability SDMA/TDMA scheduling and rate allocation policy that is relevant and more meaningful than "fairness" policies in the case of random arrivals and finite-length transmission queues. Moreover, unlike previous works, we took into account the key aspect of non-perfect CSIT, which allows for decoding errors, and a simple ARQ protocol that retransmits packets that are not successfully decoded.

Our results evidenced that the ability of accurately estimating the channel (or beams) SINRs has a fundamental impact on the performance of opportunistic beamforming schemes. In the case of mobile communications, with a delay in the feedback, the transmitter cannot have perfect CSIT and there exists a non-trivial tradeoff between multiuser diversity and transmit diversity. It clearly appears that the multiuser diversity gain disappears as soon as the channels change too rapidly. Hence, while random beamforming with $B = M$ beams should be chosen for very slow channels, space-time coding should be chosen for faster mobility terminals. This result cannot be observed under the somehow naive assumption of no feedback delay made by other works.

We have also proposed and demonstrated an improvement of the random beamforming system such that the users are synchronized and have a priori knowledge of the (pseudo-) random beamforming matrices. This decouples the problem of channel estimation and prediction from the speed of variation of the beamformers.

Curiously, it appears that the opportunistic single beamforming scheme is not attractive since its performance is dominated by either STC for large Doppler bandwidth or by the opportunistic M-beam scheme for small Doppler bandwidth.

Appendix

A Proof of Theorem 1.

For a fixed policy $\{\mathbf{P}^{(t)}, \mathbf{R}^{(t)}\}$, the instantaneous service rate (information bits per slot) for user k is given by (see (4)) $\mu_k(t) = T\sum_{j=1}^{B} p_{k,j}^{(t)} R_{k,j}^{(t)} \ \mathbf{1}\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$. The results in [21] applied to our setting yield that under the assumption of Section 3 the system is stable if and only if

$$\lambda_k \leq \underline{\mu}_k = \liminf_{t\to\infty} \frac{1}{tT} \sum_{\tau=1}^{t} \mu_k(\tau), \quad k = 1, \ldots, K \tag{18}$$

Let $\widetilde{\Omega}$ denote the stability region of a new system with channel state $\boldsymbol{\alpha}(t)$ and feasible rate $\mathbf{R}^{\text{out}}(t)$. The instantaneous service rate (information bits per slot) of user k in the new system is given by $\widetilde{\mu}_k(t) = T\sum_{j=1}^{B} p_{k,j}^{(t)} R_{k,j}^{\text{out}}(\boldsymbol{\alpha}(t))$, which is linear (hence concave) and non-decreasing in $\mathbf{P}^{(t)}$. Theorem 1 of [21] yields that the stability region $\widetilde{\Omega}$ of the new system is given by (8). Hence, the theorem is proved if we show that the two stability regions coincide.

By restricting the resource allocation policies for the original system to the stationary policies $\{\mathbf{P}, \mathbf{R}^\star\}$, for $\mathbf{P} \in \mathcal{P}$ and $\mathbf{R}^\star$ defined by (9), the arrival processes $\{A_k(t)\}$ and the service rates $\{\mu_k(t)\}$ become jointly Markov modulated [21] and it is immediate to show, by ergodicity, that $\liminf_{t\to\infty} \frac{1}{tT}\sum_{\tau=1}^{t} \mu_k(\tau) = \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\rm out}(\boldsymbol{\alpha})\right]$. Hence, any point $\boldsymbol{\lambda} \in \widetilde{\Omega}$ is also in Ω and it is stabilized by a policy $\{\mathbf{P}, \mathbf{R}^\star\}$ for some $\mathbf{P} \in \mathcal{P}$. This shows that $\widetilde{\Omega} \subseteq \Omega$.

In order to show that $\Omega \subseteq \widetilde{\Omega}$, let's assume that the channels and CSIT signals take on values in finite discrete sets (the proof can be extended for well-behaved continuous processes by using standard discretization and continuity arguments). By ergodicity and from the definition of liminf, for any $\epsilon > 0$ there exists a sufficiently large t_ϵ such that, simultaneously,

$$\begin{aligned} \frac{|N_{\mathbf{a}}(t_\epsilon)|}{t_\epsilon} &\leq P_\alpha(\mathbf{a}) + \epsilon \\ \frac{1}{t_\epsilon T}\sum_{\tau=1}^{t_\epsilon}\sum_{j=1}^{B} \mu_{k,j}(\tau) &\geq \underline{\mu}_k - \epsilon \\ \frac{1}{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|} \sum_{\tau\in N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)} \mathbf{1}\{r_{k,j} < \overline{R}_{k,j}(\tau)\} &\leq 1 - \mathbb{P}(\overline{R}_{k,j} \leq r_{k,j} | \boldsymbol{\alpha} = \mathbf{a}) + \epsilon \end{aligned} \tag{19}$$

where we define the sets $N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t) = \left\{\tau \in \{1,\ldots,t\} \;:\; \boldsymbol{\alpha}(\tau) = \mathbf{a}, \mathbf{P}^{(\tau)} = \mathbf{p}, \mathbf{R}^{(\tau)} = \mathbf{r}\right\}$, $N_{\mathbf{a},\mathbf{p}}(t) = \bigcup_{\mathbf{r}} N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t)$ and $N_{\mathbf{a}}(t) = \bigcup_{\mathbf{p}} N_{\mathbf{a},\mathbf{p}}(t)$, and where $P_\alpha(\mathbf{a}) = \mathbb{P}(\boldsymbol{\alpha} = \mathbf{a})$ denotes the stationary probability of $\boldsymbol{\alpha}(t)$, which exists by assumption, and where we use the short-hand notation $\mu_{k,j}(t) = T p_{k,j}^{(t)} R_{k,j}^{(t)} \mathbf{1}\{R_{k,j}^{(t)} < \overline{R}_{k,j}(t)\}$ to denote the instantaneous service rate for user k on stream j in slot t.

The last inequality in (19) follows since by assumption any feasible resource allocation policy is a causal function of the CSIT process, the CSIT process is ergodic, and $\overline{R}_{k,j}(\tau)$ is independent of $\boldsymbol{\alpha}_1^{\tau-1}$ given $\boldsymbol{\alpha}(\tau)$.

Consider any stable arrival rate vector $\boldsymbol{\lambda}$. By the necessary condition of [21, Lemma 1](see above), there exists some not necessarily stationary policy $\{\mathbf{P}^{(t)}, \mathbf{R}^{(t)}\}$ such that, for all $k = 1, \ldots, K$,

$$\lambda_k \leq \underline{\mu}_k \leq \frac{1}{t_\epsilon T}\sum_{\tau=1}^{t_\epsilon}\sum_{j=1}^{B} \mu_{k,j}(\tau) + \epsilon \tag{20}$$

By using (19), we have

$$
\begin{aligned}
\lambda_k &\leq \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a}) \frac{1}{|N_{\mathbf{a}}(t_\epsilon)|} \sum_{\tau\in N_{\mathbf{a}}(t_\epsilon)} \sum_{j=1}^{B} p_{k,j}^{(\tau)} R_{k,j}^{(\tau)} \mathbf{1}\{R_{k,j}^{(\tau)} < \overline{R}_{k,j}(\tau)\} + \epsilon' \\
&\leq \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a}) \sum_{\mathbf{p},\mathbf{r}} \frac{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|} \frac{1}{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|} \sum_{\tau\in N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)} \sum_{j=1}^{B} p_{k,j} r_{k,j} \mathbf{1}\{r_{k,j} < \overline{R}_{k,j}(\tau)\} + \epsilon' \\
&\leq \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a}) \sum_{\mathbf{p},\mathbf{r}} \frac{|N_{\mathbf{a},\mathbf{p},\mathbf{r}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|} \sum_{j=1}^{B} p_{k,j} r_{k,j} (1 - \mathbb{P}(\overline{R}_{k,j} \leq r_{k,j} | \boldsymbol{\alpha} = \mathbf{a})) + \epsilon'' \\
&\overset{(a)}{\leq} \sum_{\mathbf{a}\in\mathcal{A}} P_{\boldsymbol{\alpha}}(\mathbf{a}) \sum_{j=1}^{B} \widetilde{p}_{k,j}(\mathbf{a}) R_{k,j}^{\mathrm{out}}(\mathbf{a}) + \epsilon'' \\
&= \widetilde{\lambda}_k + \epsilon''
\end{aligned} \tag{21}
$$

where (a) follows from the definition of $R_{k,j}^{\mathrm{out}}(\mathbf{a})$ (see (7)) and by letting

$$
\widetilde{p}_{k,j}(\mathbf{a}) = \sum_{\mathbf{p}} \frac{|N_{\mathbf{a},\mathbf{p}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|} p_{k,j}
$$

Since by assumption $\sum_{k=1}^{K} p_{k,j} \leq 1$ for all $j = 1, \ldots, B$ on every slot, and since $\sum_{\mathbf{p}} \frac{|N_{\mathbf{a},\mathbf{p}}(t_\epsilon)|}{|N_{\mathbf{a}}(t_\epsilon)|} = 1$, it follows that $\sum_{k=1}^{K} \widetilde{p}_{k,j}(\mathbf{a}) \leq 1$ for all j and $\mathbf{a} \in \mathcal{A}$, i.e., the stationary policy $\{\widetilde{p}_{k,j}\}$ defined above is feasible. Hence, $\widetilde{\boldsymbol{\lambda}}$ with k-th component given by the last line of (21) is a point in $\widetilde{\Omega}$. Since $\epsilon > 0$ is arbitrary, and $\epsilon'' \to 0$ as $\epsilon \to 0$, we have that $\boldsymbol{\lambda} \leq \widetilde{\boldsymbol{\lambda}} \in \widetilde{\Omega}$, which eventually implies that $\widetilde{\Omega}$ and Ω coincide.

B Proof of Theorem 3.

The feasible rates $\overline{R}_{k,j}(t) = \log_2(1 + \mathsf{SINR}_{k,j}(t))$ change randomly and independently every N slots. We shall refer to a block of N slots over which the feasible rates are constant as a *frame*. Theorem 3 can be proved by considering the stability of a frame-based modification of max-stability policy, which maximizes $\sum_k \theta_k S_k(t) \widetilde{\mu}_k(t)$ subject to the SDMA/TDMA feasibility constraint (3) at every frame and keeps the same policy during all N slots in a frame irrespectively of the arrival processes.

For the general theory of the Lyapunov drift (see [21,34]), defining the Lyapunov function

$$L(\mathbf{S}) = \sum_{k=1}^{K} \theta_k S_k^2 \tag{22}$$

if for some constants $\mathcal{K}_1$ and ν_k the condition

$$\mathbb{E}[L(\mathbf{S}(t+N)) - L(\mathbf{S}(t))|\mathbf{S}(t)] \leq \mathcal{K}_1 - \sum_{k=1}^{K} \nu_k S_k(t) \tag{23}$$

on the Lyapunov drift holds, then $\sum_k \nu_k \mathbb{E}[S_k(t)] \leq \mathcal{K}_1$ and by Little's theorem, we obtain $\sum_k \nu_k \lambda_k \mathbb{E}[D_k] \leq \frac{\mathcal{K}_1}{T}$. Let consider an arbitrary time t_0 and the N-slot buffer evolution

$$S_k(t_0+N) = \left[S_k(t_0) - NT\sum_{j=1}^{B} p_{k,j}(t_0)\overline{R}_{k,j}(t_0)\right]_+ + \sum_{\tau=t_0}^{t_0+N-1} A_k(\tau) \tag{24}$$

It is immediate to show that the Lyapunov drift can be bounded as

$$\mathbb{E}\left[L(\mathbf{S}(t_0+N)) - L(\mathbf{S}(t_0))|\mathbf{S}(t_0)\right] \leq N^2T^2\widetilde{\mathcal{K}} - 2NT\sum_k \theta_k S_k(t_0)\left(\mathbb{E}[\overline{\mu}_k(t_0)|\mathbf{S}(t_0)] - \lambda_k\right) \tag{25}$$

where we define $\overline{\mu}_k(t_0) = \sum_{j=1}^{B} p_{k,j}(t_0) R_{k,j}(t_0)$ and $A_k = \frac{1}{N}\sum_{\tau=t_0}^{t_0+N-1} A_k(\tau)$ and where $\widetilde{\mathcal{K}} = \sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k}{T}\right)^2\right] + \sum_k \theta_k \mathbb{E}\left[\overline{\mu}_k^2(t_0)|\mathbf{S}(t_0)\right]$. In order to bound this term by a constant, we have

$$\begin{aligned}
\sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k}{T}\right)^2\right] &\overset{(a)}{\leq} \sum_k \theta_k \frac{1}{N}\sum_{\tau=t_0}^{t_0+N-1} \mathbb{E}\left[\left(\frac{A_k(\tau)}{T}\right)^2\right] \\
&\overset{(b)}{=} \sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k(\tau)}{T}\right)^2\right]
\end{aligned}$$

where inequality (a) follows from Jensen's inequality and equality (b) follows from the i.i.d. arrival assumption.

Then, we also have that

$$\begin{aligned}
\sum_{k=1}^{K} \theta_k \mathbb{E}\left[\overline{\mu}_k^2(t_0)|\mathbf{S}(t_0)\right] &\overset{(c)}{\leq} \sum_{k=1}^{K} \theta_k \mathbb{E}\left[\left(\sum_{j=1}^{B} p_{k,j}^2(t_0)\right)\left(\sum_{j=1}^{B} \overline{R}_{k,j}^2(t_0)\right)|\mathbf{S}(t_0)\right] \\
&\overset{(d)}{\leq} B\sum_{k=1}^{K} \theta_k \sum_{j=1}^{B} \mathbb{E}\left[\overline{R}_{k,j}^2(t_0)\right]
\end{aligned}$$

where inequality (c) follows from Cauchy-Schwartz inequality and inequality (d) follows from $\sum_{j=1}^{B} p_{k,j}^2 \leq B$ for any k. By letting $\mathcal{K} = \sum_k \theta_k \mathbb{E}\left[\left(\frac{A_k(t)}{T}\right)^2\right] + B\sum_{k=1}^{K} \theta_k \sum_{j=1}^{B} \mathbb{E}\left[R_{k,j}^2(t)\right]$ we have $\widetilde{\mathcal{K}} \leq \mathcal{K}$. It follows that Theorem 3 holds for the frame-based policy. Finally, we argue that the frame-based policy is a suboptimal policy that deliberately ignores the buffer state at times $t \neq t_0 + mN$, for $m = 1, 2, \ldots$. With little effort it is possible to show that the frame-based policy also achieves the system stability region for any finite N and that if the max-stability (slot-based) policy and the frame-based policy are applied to the same initial state $\mathbf{S}(t_0)$ then $\sum_k \theta_k \mathbb{E}[S_k^{slot}(t)] \leq \sum_k \theta_k \mathbb{E}[S_k^{frame}(t)]$ for all $t \geq t_0$. Then, Theorem 3 applies also to the max-stability policy.

References

[1] 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, pp. 1083–1102, March 2001.

[2] D.N.C.Tse, "Optimal Power Allocation over parallel Gaussian Broadcast Channels," *unpublished.*

[3] 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.

[4] "Cdma2000 high rate packet dataair interface specification," *TIA/EIA/3GPP2 Standard IS-856/3GPP2 C.S.0024, v3.0*, December 2001.

[5] T.E.Kolding et al., "High Speed Downlink Packet Access : WCDMA Evolution," *IEEE Vehic. Tech. Society News*, pp. 4–10, February 2003.

[6] P.Viswanath, D.N.C.Tse, and R.Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Trans. on Inform. Theory*, vol. 48, no. 6, June 2002.

[7] M.Sharif and B.Hassibi, "On the Capacity of MIMO Broadcast Channel with Partial Side Information," *IEEE Trans. on Inform. Theory*, vol. 51, no. 2, pp. 506 – 522, February 2005.

[8] R.Knopp and P.A.Humblet, "Information capacity and power control in single-cell multiuser communications," *Proc. IEEE ICC*, vol. 1, pp. 331–335, 1995.

[9] R.Knopp, "Coding and Multiple-Access over Fading Channels," *Ph.D dissertation, Ecole Polytechnique Fédérale de Lausanne*, 1997.

[10] E.G.Larsson and P.Stoica, "Space-time block coding for wireless communications," *Cambridge University Press*, 2003.

[11] A.Paulraj, R.Nabar, and D.Gore, "Introduction to space-time wireless communications," *Cambridge University Press*, 2003.

[12] E.Biglieri and G.Taricco, "Transmission and reception with multiple antennas: Theoretical foundations," *Foundations and Trends in Communications and Information Theory*, 2004.

[13] 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, May 2003.

[14] B.M.Hochwald, T.L.Marzetta, and V.Tarokh, "Multiple-Antenna Channel Hardening and its Implications for Rate Feedback and Scheduling," *IEEE Trans. on Inform. Theory*, vol. 50, no. 9, September 2004.

[15] H.Viswanathan and S.Venkatesan, "The Impact of Antenna Diversity in Packet Data Systems with Scheduling," *IEEE Trans. on Commun.*, pp. 546–549, April 2004.

[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, pp. 1073 – 1096, May 2003.

[17] S.Borst and P.Whiting, "The Use of Diversity Antennas in High-Speed Wireless Systems: Capacity Gains, Fairness Issues," *Bell Laboratories Technical Memorandum, 2001*, 2001.

[18] T.Bonald, "Flow-level performance analysis of some opportunistic scheduling algorithms," *Euro. Trans. on Telecomms.*, vol. 16, pp. 65–75, 2005.

[19] S.Shakkottai and A.Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," *Proc. of the 17th International Teletraffic Congress (ITC-17)*, September 2001.

[20] A.Sang, X.Wang, M.Madihian, and R.D.Gitlin, "Downlink scheduling schemes in cellular packet data systems of multiple-input multiple-output antennas," *Proceeding of IEEE Globecom 2004*, pp. 4021–4–27, December 2004.

[21] 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, pp. 138–152, February 2003.

[22] 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.

[23] L.H.Ozarow, S.Shamai, and A.D.Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. on Vehic. Tech.*, vol. 43, pp. 359 – 378, May 1994.

[24] E. Biglieri, J. Proakis, , and S. Shamai, "Fading channels : information-theoretic and communications aspects," *IEEE Trans. on Inform. Theory*, vol. 44, pp. 2619–2692, October 1998.

[25] D.Samardzija and N.Mandayam, "Unquantized and Uncoded Channel State Information Feedback on Wireless Channels," *Proceeding of IEEE WCNC'2005*, pp. New Orleans, LA, USA, March 2005.

[26] S.Sanayei and A.Nosratinia, "Opportunistic Beamforming with Limited Feedback," *submitted to IEEE International Symposium on Information Theory*, 2005.

[27] John G.Proakis, "Digital communications," *McGraw-Hill Book Company*, 1997.

[28] N.Randolph, "Probability, stochastic process and queueing theory :the mathematics of computer performance modeling," *Springer-Verlag*, 1995.

[29] E.Biglieri, G.Caire, G.Taricco, and J.Ventura-Traveset, "Simple method for evaluating error probabilities," *Electron.Lett.*, vol. 32, pp. 191–192, February 1996.

[30] M.Kobayashi, "On the use of Multiple Antennas for Downlink of Wireless Systems," *Ph.D dissertation, ENST*, in preparation.

[31] W.Turin and R.Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1809–1817, December 1998.

[32] K.E.Baddour and N.C.Beaulieu, "Autoregressive models for fading channel simulation," *Proceeding of IEEE Globecom 2001*, pp. 1187–1192, November 2001.

[33] William C Jakes, "Microwave Mobile Communications," *IEEE Press*, 1974.

[34] M.J.Neely, "Dynamic Power Allocation and Routing for Satellites and Wireless Networks with Time-Varying Channels," *Ph.D dissertation, MIT*, November 2003.

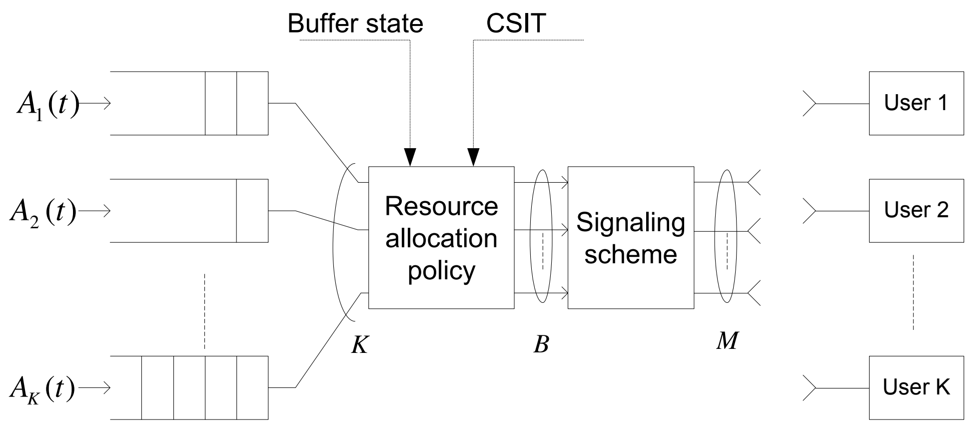

Figure 1: Block diagram of the SDMA/TDMA downlink

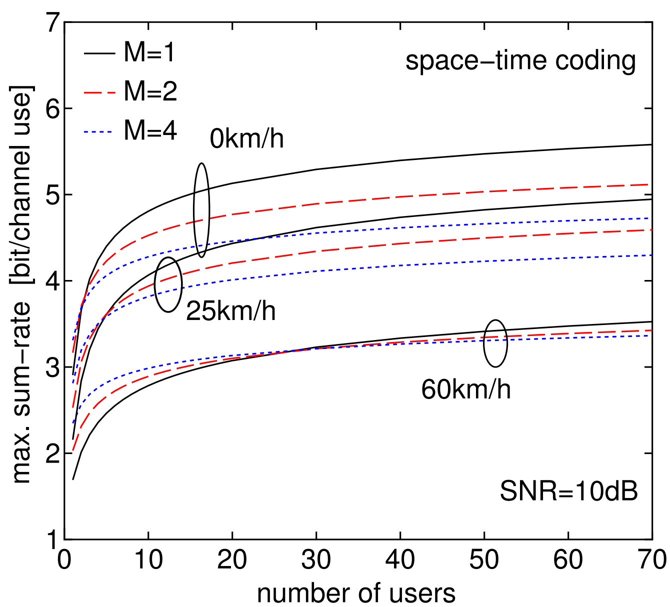

Figure 2: max. sum-rate vs. number of users (STC)

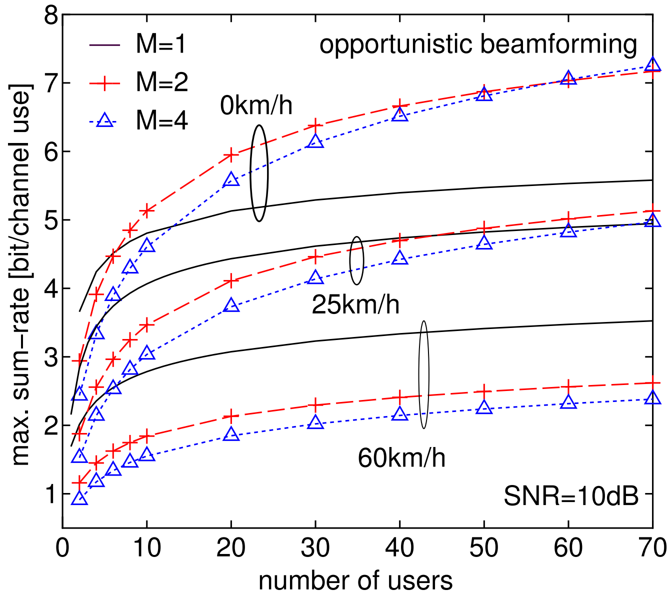

Figure 3: max sum-rate vs. number of users (beamforming with $B = M$)

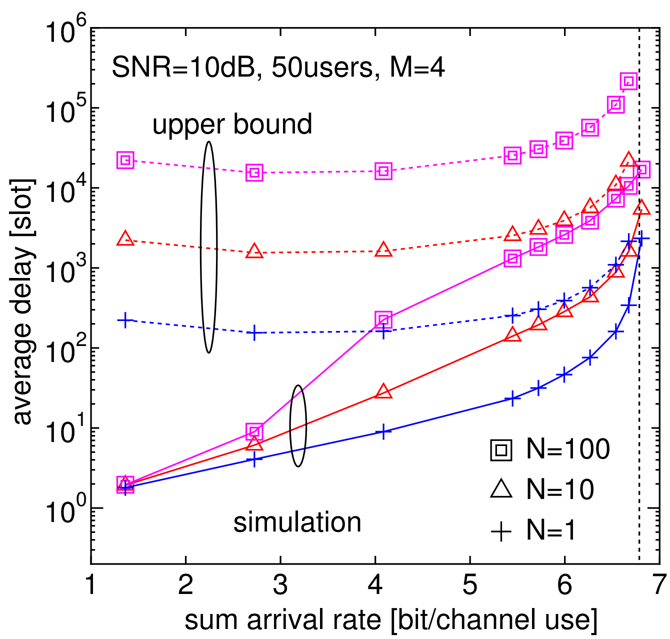

Figure 4: Impact of beamforming variation speed

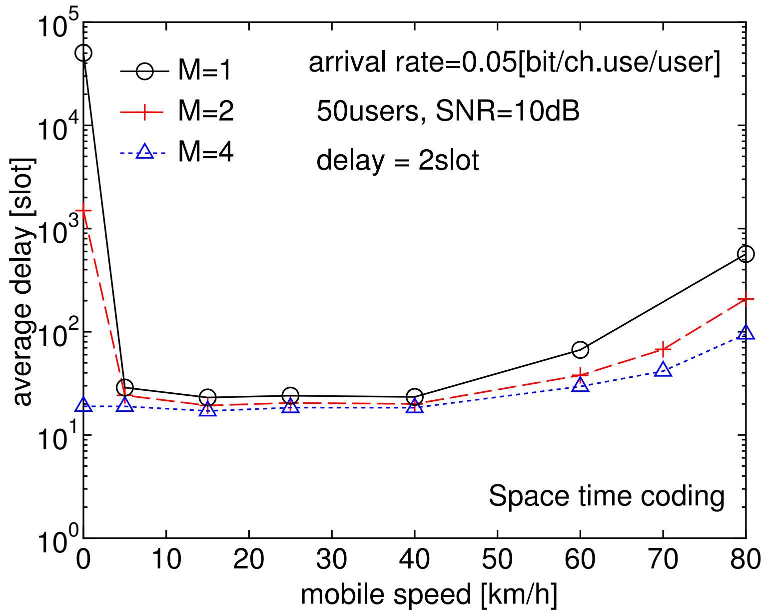

Figure 5: average delay vs. speed (STC)

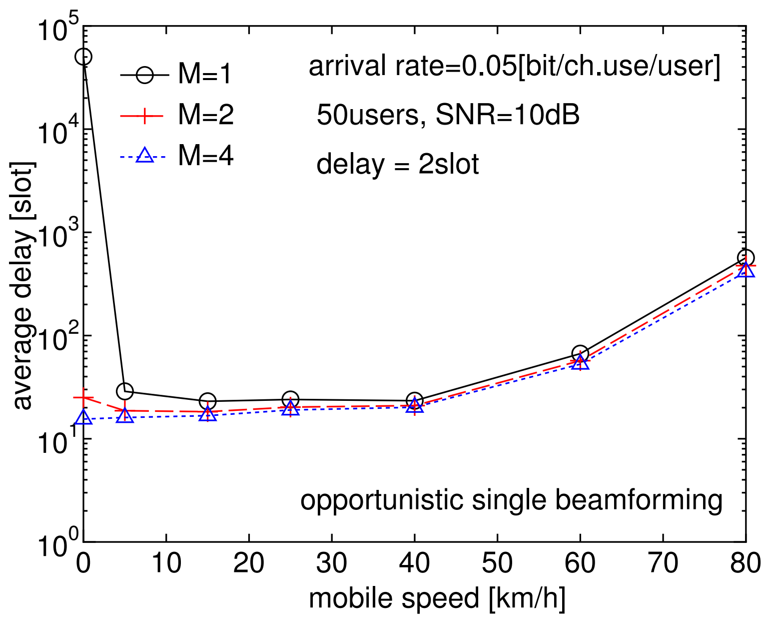

Figure 6: average delay vs. speed ($B = 1$ beamforming)

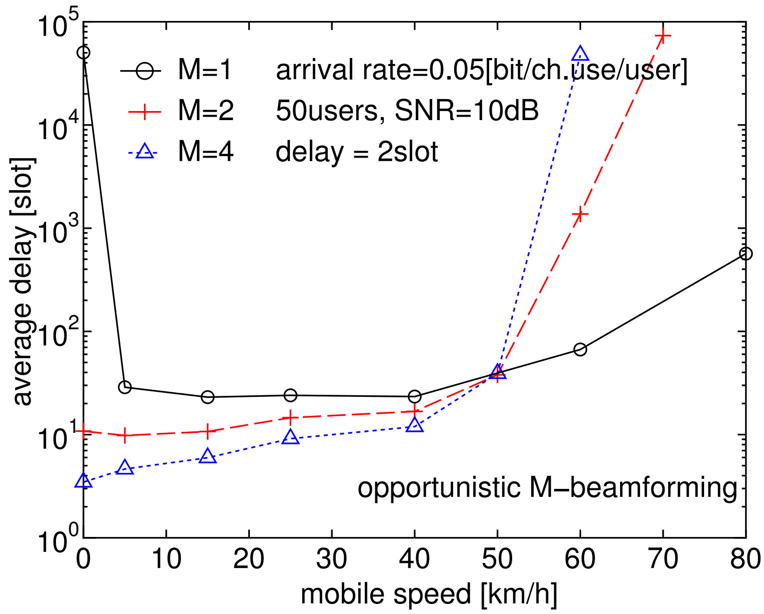

Figure 7: average delay vs. speed ($B = M$ beamforming)