IMPACT OF MULTIPLE TRANSMIT ANTENNAS IN A QUEUED SDMA/TDMA DOWNLINK

Mari Kobayashi, Giuseppe Caire, and David Gesbert

Institut EURECOM, Sophia-Antipolis, France
E-mail: firstname.name@eurecom.fr

ABSTRACT
We investigate the impact of multiple transmit antennas on the performance of a SDMA/TDMA single-cell downlink system under random packet arrivals, correlated block-fading channels and non-perfect channel state information at the transmitter due to a delay in the feedback link. We derive the arrival rate stability region and the adaptive scheduling policy that stabilizes any arrival rate point inside the region without knowing explicitly the fading and arrival statistics. Then, we apply these results to the case of “opportunistic” beamforming and space-time coding. 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 by the receivers in order to improve the channel state information quality. Under this scheme, it appears that for given feedback delay the relative merit of opportunistic beamforming and space-time coding (transmit diversity) strongly depends on the channel Doppler bandwidth.

1. MOTIVATION
In 3G wireless Internet applications, a high data rate downlink for delay-tolerant packet communications is needed. The downlink of a single cell system is modeled as a fading Gaussian broadcast channel, whose capacity region has been completely characterized under different assumptions in several papers (e.g., [1]). In particular, it is known that, under fading ergodicity, 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 serving the user with the largest fading coefficient at each time instant (e.g., [2]). Motivated by this result, downlink scheduling schemes such as the High-Data Rate (HDR) [3] or the 1xEV-DO [4] have been proposed. Such systems assume that all connected users have infinite backlog (i.e., all data present at the base station, no arrival processes).

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 [5] 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 so called opportunistic beamforming proposed in [2], where the multiple antennas are used to generate a random 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 fairly among the users. A spatial-multiplexing version of the opportunistic beamforming is proposed and analyzed in [6], where \( M \) mutually orthogonal random beams are simultaneously used and the best \( M \) users are selected at each time. It is shown that for \( K \gg M \) and assuming perfect SNR instantaneous feedback, the same multiplexing gain of \( M \) as for the case of perfect CSI is achievable.

In parallel with the development of opportunistic schemes, the current research and standardization trend in wireless cellular systems has 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 [7]. In the most realistic scenario where the base station is equipped with \( M \) antennas and the mobile terminal has a single antenna, STC achieves \( M \)-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.

It is natural to investigate the interaction of STC and opportunistic scheduling algorithms and which use of multiple antennas at the base station one can eventually make. Based on the optimistic assumptions of perfect SNR feed-
back and infinite backlog, a number of recent works showed that the transmit diversity achieved by STC is detrimental for the multiuser diversity effect connected to opportunistic beamforming/scheduling schemes [8, 9]. These results led to the naive conclusion that STC should be avoided in high data rate delay-tolerant downlink applications. In this paper we take a deeper look into this problem and consider two fundamental aspects that were neglected in works such as [2, 6, 8, 9]: we consider random packet arrivals with finite transmission buffers, and we consider time-varying fading channels with a delay in the feedback link. In the case of finite queue buffers and random packet arrivals the traditional notion of fairness is replaced by the notion of stability region and we may think of the second as belonging to the system stability region.

The realistic assumption of feedback link delay prevents the transmitter to know exactly the receivers SNR. This makes the information outage probability non-zero, and therefore there exists a non-trivial tradeoff between the transmit diversity achieved by STC and the multiuser diversity achieved by opportunistic schemes.

By formulating the problem in this way, we compare some common SDMA/TDMA downlink schemes inspired by current technology: STC (transmit diversity) and random beamforming with $1 \leq B \leq M$ beams. The ability of accurately predicting the channel SNR dominates the performance of opportunistic beamforming. Hence, we propose a new scheme based on pseudo-random unitary beamforming matrices known by the receivers (in analogy with random-spreading CDMA, where the downlink scrambling sequence is synchronized and known to all users in the cell). In this way, the user terminals have only to track and predict the underlying physical channel which can be much slower than the variation of the pseudo-random beam pattern. Under this scheme, it appears that for given feedback delay the relative merit of opportunistic beamforming and STC strongly depends on the channel Doppler bandwidth. In particular, it appears that for slowly-varying channels the multiple-beam scheme with $B = M$ [6] achieves the best average delay, while for faster channels STC is better. In light of these results, the utility of random beamforming with $B = 1$, as in [2], is questionable.

2. SDMA/TDMA DOWNLINK SYSTEM MODEL

We consider a base station with $M$ antennas transmitting to $K$ user terminals. Transmission is slotted and each slot comprises $N$ channel uses (complex dimensions). The signal received at user $k$ terminal in slot $t$ is given by

$$
y_k(t) = X(t)h_k(t) + w_k(t)
$$

where $X(t) \in \mathbb{C}^{N \times M}$ is the transmitted codeword, $h_k(t) \in \mathbb{C}^{M \times 1}$ denotes the $M$-input 1-output channel response for the user $k$ channel in slot $t$, assumed time-invariant over each slot and $w_k(t) \in \mathbb{C}^{N \times 1}$ is complex circularly symmetric AWGN with components $\mathcal{C} \mathcal{N}(0,1)$. The base station has fixed transmit power $\gamma$ in each slot, that is, $\text{tr}(X(t)X(t)^H) \leq \gamma N$ for all $t$. 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 codes exist whose block error probability is essentially given by the information outage probability [7]

$$
P_{\text{out}}(R) = \Pr \left( \log_2 (1 + \beta(t) \gamma) \leq R \right)
$$

where $\beta(t) \gamma$ denotes the instantaneous received SNR and $R$ the coding rate (bit/channel use).

We assume a SDMA/TDMA downlink system. Namely, at each point in time, a subset of $1 \leq B \leq M$ out of $K$ users is selected and independent information messages are sent to these users via $B$ independently selected codewords. The codewords can be linearly combined via a beamforming matrix. The ensemble of user selection, beamforming and rate allocation rules for an SDMA/TDMA policy. The policy is generally a function of the feedback signals sent back by the users over a feedback link. Information packets destined to the users arrive randomly at the base station from some underlying wired network, and are stored into $K$ queues, where queue $k$ is associated to user $k$. The arrival process of queue $k$ is denoted by $A_k(t)$, with arrival rate $\lambda_k = \mathbb{E}[A_k(t)]$ (bit/channel use), and the buffer size of queue $k$ is denoted by $S_k(t)$ (bit). At the beginning of each slot, a DRC signal $\alpha(t) = \{ \alpha_1(t), \ldots, \alpha_K(t) \}$ is revealed to the transmitter. The transmitter is characterized by certain feasible rate functions, denoted by $p_{k,j} R_{k,j}(\alpha)$, where $R_{k,j}(\alpha)$ is a function that will be specified later, $p_k$ is a SDMA/TDMA resource-sharing matrix, and $\alpha$ is the current value of the DRC signal. The resource sharing matrix $p_k$ has the following meaning: $p_{k,j} \geq 0$ is the fraction of the current slot allocated to user $k$ on beam $j$ or, equivalently, it is the probability with which the whole slot is allocated to user $k$ on beam $j$. As it will be clear from the following treatment, these two interpretations yield the same results in terms of stability region and we may think of the second as a more practical option (only one user per beam transmit at any slot instead of partitioning the slot time into sub-slots).

The set of all feasible resource-sharing matrices is

$$
\mathcal{F} = \left\{ p \in \mathbb{R}_+^{K \times B} : \sum_{k=1}^{K} p_{k,j} \leq 1, \; \forall \; j \right\}
$$

(2)

With some abuse of notation, we denote by $\mathcal{F}$ also the set of resource-sharing feasible functions, i.e., the set of all functions that map the DRC signal into $\mathcal{F}$. Since the DRC signal is not ideal, there exists a non-zero probability that any
specified transmission rate $R$ cannot be supported by the channel. We assume an ARQ protocol such that an unsuccess-fully decoded packet remains in the transmission buffer and is re-scheduled for transmission at a later time. This is a good model for a delay-tolerant data packet system, that is the focus of this paper. We let the rate function $R_{k,j}(\alpha)$ to be the average rate achieved by the ARQ protocol for user $k$ over beam $j$ conditioned with respect to the current DRC signal $\alpha$ and maximized over the choice of the instantaneous coding rate, i.e.,

$$R_{k,j}(\alpha) = \max_{R \geq 0} R \left( 1 - P_{\text{out}}(R|\alpha) \right)$$

where

$$P_{\text{out}}(R|\alpha) \triangleq \Pr \left( \log_2(1 + \beta_{k,j}\gamma) \leq R|\alpha \right)$$

and where $\beta_{k,j}\gamma$ is the received SNR for user $k$ associated with the signal sent on beam $j$. The rate $R_{k,j}(\alpha)$ is achieved on average, if whenever the DRC signal is equal to $\alpha$ user $k$ is scheduled on beam $j$ and allocated an instantaneous rate equal to $R^*$, achieving the maximum in (3).

For a given SDMA/TDMA resource allocation policy $p(t)$, the queue buffers evolve in time according to the stochastic difference equation

$$S_k(t+1) = \left[ S_k(t) - N \sum_{j=1}^{B} p_{k,j}(t) R_{k,j}(\alpha(t)) \right]_+ + A_k(t)$$

for all $k = 1, \ldots, K$, where $[.]_+ \triangleq \max\{.,0\}$. In order to define stability, we follow [10] and define the buffer overflow function $g_k(S) = \limsup_{t \to \infty} \frac{1}{t} \sum_{\tau=1}^{t} 1\{S_k(\tau) > S\}$. We say that the system is stable if $\lim_{S \to \infty} g_k(S) = 0$ for all $k$. 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 resource-sharing policy for which the system is stable. Clearly, for the system defined above the main goal of a SDMA/TDMA policy is to stabilize the system whenever $\lambda \in \Omega$. In this setting, achieving any point in the stability region subsumes any reasonable fairness criterion and may be considered as the single most important goal of a downlink resource-allocation policy.

3. MAIN RESULTS

The stability theory of [10] can be easily extended to our setting, where the role of the power allocation in [10] is played by the resource-sharing allocation $p_{k,j}$ and the role of the channel state in [10] is played by the DRC signal $\alpha(t)$. A slight modification of the proofs in [10] is required to take into account the fact that here we have $B$ beams, each of which can be shared by several users. However, this modification is rather trivial and the details are omitted for the sake of space limitation. Under the following assumptions:

i) $\{A_k(t) : k = 1, \ldots, K\}$ is a set of jointly stationary ergodic Markov arrival processes with rates $\lambda = (\lambda_1, \ldots, \lambda_K)$ and $\mathbb{E}[A_k(t)] < \infty$;

ii) $\alpha(t)$ is a jointly stationary ergodic Markov $K$-dimensional DRC process independent of the arrival processes;

iii) $\{\alpha(1), \ldots, \alpha(t-1)\} \to \alpha(t) \to \{\beta_{k,j}(t) : k = 1, \ldots, K, j = 1, \ldots, B\}$ is a Markov chain; we have the following result.

**Theorem 1 [stability region].** Under assumptions i), ii) and iii), the stability region of the SDMA/TDMA downlink system defined above is given by

$$\Omega = \text{co} \left\{ \lambda \in \mathbb{R}^K_+ : \lambda_k \leq \sum_{j=1}^{B} \mathbb{E} [p_{k,j}(\alpha) R_{k,j}(\alpha)], \forall k \right\}$$

where $\text{co}$ means “closure of the convex hull”. □

For any $\lambda \in \Omega$ there exists a memoryless stationary policy $p$ (i.e., a function of the instantaneous DRC signal $\alpha$ at time $t$ only) that stabilizes all queues. However, for any given $\lambda$ the stabilizing policy is, in general, a function of $\lambda$ and of the statistics of $\alpha$. An adaptive policy is a function $p$ of the instantaneous buffer sizes $\{S_k(t)\}$ and of the DRC signal $\alpha(t)$ such that, even not knowing the arrival rates and the statistics of $\alpha(t)$, it stabilizes the queues whenever $\lambda \in \Omega$ [10, 11]. This is given by the next result.

**Theorem 2 [max-stability adaptive policy].** Under the same assumptions of Theorem 1, the SDMA/TDMA adaptive resource-sharing policy given by

$$\hat{p} = \arg \max_{p \in F} \left\{ \sum_{k=1}^{K} \theta_k S_k \sum_{j=1}^{B} p_{k,j} R_{k,j}(\alpha) \right\}$$

for any strictly positive weights $\theta_k > 0$, stabilizes the system for all $\lambda \in \Omega$. □

The solution of (7) is readily given explicitly by

$$\hat{p}_{k,j}(S_k, \alpha) = \begin{cases} 1 & k = \arg \max_{\theta_k, \theta_j} \theta_k S_k R_{k,j}(\alpha) \\ 0 & k \neq \arg \max_{\theta_k, \theta_j} \theta_k S_k R_{k,j}(\alpha) \end{cases}$$

The max-stability adaptive policy allocates on each beam $j$ the user maximizing the product $\theta_k S_k(t) R_{k,j}(\alpha(t))$ at a slot $t$. The parameters $\theta_k$ can be used in order to provide different quality-of-service to the users, as they have an influence on the average individual delays [10].

4. APPLICATION TO PRACTICAL SCHEMES

In this section we consider a specific model for the fading channel $h_k(t)$ and the DRC signal $\alpha(t)$, and particularize the general treatment of before to some schemes inspired by the current trend in high-speed downlink technology. We assume that the channel vectors $h_k(t)$ are mutually statistically independent for different index $k$ and i.i.d. for different antennas. $h_k(t)$ is constant over each slot of $N$ channel
uses, and changes from slot to slot according to a stationary ergodic Gauss Markov process, given by \( h_k(t + 1) = \sqrt{n} h_k(t) + \sqrt{1 - n} \rho z_k(t) \), where \( h_k(t) \sim CN(0, I) \). Then, we let \( \alpha(t) \) be a function of the \( d \)-step MMSE predictor \( g_k(t) = \rho^{k/2} h_k(t - d) \) of \( h_k(t) \) given \( h_k(t - d) \), where \( d \) denotes the feedback delay measured in slots. This assumption is made in order to make the assumptions ii) and iii) of Section 3. Nevertheless, we shall apply the stability policy obtained before to more realistic fading statistics such as the Jake’s correlation model [12], and define the DRC signal to be a function of the \( d \)-step MMSE predictor of \( h_k(t) \) from \( \{h_k(t - d), h_k(t - d - 1), \ldots \} \). We compare the following system choices.

**Space Time Coding (STC).** In this case, \( X(t) \in C^{N \times M} \) denotes the transmitted space-time codeword, assumed to be drawn from a Gaussian i.i.d. ensemble. The system cannot exploit spatial multiplexing since the user terminals have one antenna each. Hence, STC yields only \( M \)-fold transmit diversity, the instantaneous channel gain of user \( k \) is given by \( \beta_j \beta_j(t) = \frac{1}{|h_k(t)|^2} \). Each user feeds back its DRC \( \alpha_0(t) = \frac{1}{|h_k(t)|^2} \) such that the total number of feedbacks is \( K \) (suitably quantized) real values. All the results of Section 3 apply with \( B = 1 \), since a single user is served on each slot.

**Opportunistic beamforming.** We consider opportunistic beamforming using \( B \leq M \) mutuallly orthogonal beams. In [6] \( B = M \) while in [2] \( B = 1 \) with \( M > 1 \) antennas. It is clear that the quality of the DRC signal depends critically on the ability of predicting the physical channels \( h_k(t) \). Then, we propose a modification of [2, 6]: as in usual random-spreading CDMA, 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. Moreover, since they are unitary, they have no impact on the estimation of the underlying physical channel that can be achieved with usual pilot-aided schemes and linear prediction. In this way, the speed of variation of the random beams is independent of the ability of estimating the channels, that depends uniquely on the Doppler bandwidth. Therefore, we let the random beams change independently at each slot. We have \( X(t) = \sum_{j=1}^{B} s_j(t) \phi_j^H(t), \) where \( s_j(t) \in C^{N \times 1} \) is the signal associated to beam \( j \), \( \phi_j(t) \in C^{M \times 1} \) is the beamforming vector for beam \( j \) in slot \( t \), and it is assumed that \( \phi_j^H(t) \phi_j(t) = \delta_{j,m} \). User \( k \) “sees” SINR for the signal in beam \( j \) equal to

\[
\text{SINR}_{b,j}(t) = \frac{|\phi_j^H(t) h_k(t)|^2}{B / \gamma + \sum_{m \neq j} |\phi_m^H(t) h_k(t)|^2} \tag{9}
\]

for \( j = 1, \ldots, B \). The instantaneous channel gain is given by \( \beta_{b,j} = \text{SINR}_{b,j}(t) / \gamma \). The outage rate (3) conditioned on the prediction \( g_k(t) \) of the channel can be computed by numerical integration (details are omitted for the sake of space limitation). As a matter of fact, each user feeds back \( B \) outage rates for each of the beams such that the total number of feedbacks is \( KB \) (suitably quantized) real values.

In order to reduce further the number of real values to feed back, following [6] we may assume that each user feeds back only its best outage rate and the index of the beam that achieves it. We have checked by simulation, in accordance with the asymptotic results of [6], that the difference between these two schemes is minimal and therefore, since it is conceptually simpler, we shall consider the case where each user feeds back all \( B \) outage rate values in our numerical results.

5. NUMERICAL RESULTS AND CONCLUSIONS

**Simulation setting.** We considered mutually independent arrival processes such that \( A_k(t) = \sum_{j=1}^{M} \beta_{b,j}(t) b_{b,j}(t) \), where \( M_0 \) 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 \( b_{b,j}(t) \) are i.i.d. exponentially distributed random variables expressing the number of bits per packet. We take \( \mathbb{E}[b_{b,j}(t)] = N \) \((N = 2000 \) in our simulations), so that \( A_k \) coincides with the average number of packets arrived in a slot \( N \) channel uses). The fading \( h_k(t) \sim CN(0, I) \) is a stationary ergodic Gaussian complex circularly symmetric vector process. We consider Jake’s Doppler model, with correlation function \( J_B(2\pi f_D T_{d,}\ell, t) \) and \( f_D \) and \( T_{d,\ell} \) denoting the one-sided Doppler bandwidth (in Hz) and the slot duration (in seconds), respectively. Since we assume the fading constant over a slot, this model is meaningful for \( f_D T_{d,\ell} \ll 1 \). We let \( T_{d,\ell} = 1.67 \) msec [3], the channel prediction order \( n = 8 \), the feedback delay \( d = 2 \) slot. Under this setting, the mobile speeds \( v = 0, 25, 40, 60, 80 \) km/h yield a channel prediction error \( \sigma_d^2 = 0.00, 0.05, 0.10, 0.40, 0.60 \) respectively. For opportunistic beamforming, we generate a new set of random beams every slot.

**Average delay performance.** We evaluated the average delay of STC and opportunistic beamforming as a function of the mobile speed in km/h by letting the total arrival rate to 2.5 bit/channel use. By Little’s theorem, the average delay is given by \( \overline{D} = \frac{\sum_{k=1}^{K} \sum_{\ell=1}^{L} S_k / \lambda_k}{\overline{S}} \) measured in slot where \( S_k \) denotes the \( k \) user’s time-averaged buffer size in bit. We consider the symmetric arrival case. Figs. 1, 2, 3 shows the average delay for a system with 50 users, average SNR \( \gamma = 10 \) dB 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 it is introduced for the sake of comparison with a standard single-antenna system.

For a very slowly-varying channel (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 the so called “channel-hardening effect” [9]: ergodicity is recovered in the spatial domain by increasing the number of transmit antennas.

As seen from Fig. 2 and 3, opportunistic random beamforming decreases the average delay by making the channel vary almost i.i.d.. When the channel is slow (up to 40km/h), opportunistic beamforming with \( M \) beams achieves the smallest delay. As the mobile speed increases (i.e., the quality of DRC becomes worse), STC outperforms the random beamforming schemes 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.

These results show that the ranking of STC and opportunistic beamforming is not clear and depends critically on the ability of feeding back accurate SNR measurements or predictions. Generally speaking, it appears that the opportunistic single beamforming is not very attractive because its performance is dominated by either STC for large Doppler bandwidth or the opportunistic \( M \)-beamforming for small Doppler bandwidth.

6. REFERENCES


