Transmit Diversity Versus Opportunistic Beamforming in Data Packet Mobile Downlink Transmission

Mari Kobayashi, *Member, IEEE*, Giuseppe Caire, *Fellow, IEEE*, and David Gesbert, *Senior Member, IEEE*

***Abstract*—We compare space–time coding (transmit diversity) and random "opportunistic" beamforming in a space-division multiple access/time-division multiple access 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. The ability of accurately predicting the channel signal-to-noise ratio dominates the performance of opportunistic beamforming, even under the optimistic assumption that the sequence of beamforming matrices is perfectly known *a priori* by the receivers. Our results show that 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.**

***Index Terms*—Non-perfect channel state information at the transmitter (CSIT), opportunistic beamforming, space–time coding (STC), stability.**

I. INTRODUCTION

WE CONSIDER the downlink of a wireless system where the 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 with one user.

The base station operates in space-division multiple access (SDMA)/time-division multiple access (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 simultaneously serves B users at any point in time. 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 an SDMA/TDMA policy.

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 [1]–[3]. When the base station has multiple antennas, random "opportunistic" beamforming has been proposed in [4] and [5]. These systems generate $1 \leq B \leq M$ random time-varying beams. The scheduling algorithm allocates the best user on each beam at any point in time. The ability of a system to serve a user in its peak rate conditions is referred to as "multiuser diversity."

A different use of the M antennas consists of improving the reliability of transmission for each user by space–time coding (STC). In this case, we have $B = 1$ (TDMA), since all antennas are used to achieve *transmit diversity* to a single user. Comparisons between STC (TDMA) and opportunistic beamforming (SDMA/TDMA) have been provided, for example, in [6]–[8]. These works, as well as many others, indicate that transmit diversity always *decreases* the downlink throughput if sufficient multiuser diversity (sufficiently large K) is available.

The conclusions of these works strongly depend on the key assumption that decoding errors never occur, i.e., once a user is scheduled and it is allocated a given (channel-dependent) rate, the message will be successfully received with probability 1. This assumption is valid for sufficiently large T, if the maximum rate supported by the channel in each slot is perfectly known at the transmitter.

We take a different look at the problem and compare STC with opportunistic beamforming in the case of random bit arrival and non-ideal channel state information at the transmitter (CSIT) due to a delay in the channel state feedback loop. Under random arrivals, system *stability* (defined formally later) subsumes any reasonable "fairness" criterion. Furthermore, under non-perfect CSIT, the probability of decoding error cannot be arbitrarily small: there is a non-zero probability that the scheduled rate is above the actual channel *instantaneous* capacity. This makes the comparison between transmit diversity and multiuser diversity non-trivial.

In our setting, the ability of accurately predicting the channel signal-to-noise ratio (SNR) clearly emerges as one of the main limiting factors of multiuser diversity schemes, even under some optimistic assumptions (see later) in favor of the random beamforming schemes. While for slowly varying channels the opportunistic beamforming with $B = M$ beams achieves the best average delay, STC performs better for larger channel Doppler bandwidth.

Paper approved by A. Zanella, the Editor for Wireless Systems of the IEEE Communications Society. Manuscript received June 20, 2005; revised April 6, 2006. This work was supported in part by France Télécom, and in part by the Institut Eurécom under Grant 424 87 987. This paper was presented in part at SPAWC, New York, NY, 2005, and in part at IST, Dresden, Germany, 2005.

M. Kobayashi is with the Centre Tecnologic de Telecomunicacions de Catalunya (CTTC), 08860 Castelldefels, Barcelona, Spain (e-mail: mari.kobayashi@eurecom.fr).

G. Caire is with the Department of Electrical Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089 USA.

D. Gesbert is with the Institut Eurecom, B.P. 193, 06904 Sophia Antipolis Cedex, France.

Digital Object Identifier 10.1109/TCOMM.2006.885090

II. System Model and Definition of Stability

We assume a frequency nonselective block-fading channel [9] where the signal received at user k terminal in slot t is given by

$$\boldsymbol{y}_k(t) = \boldsymbol{X}(t)\boldsymbol{h}_k(t) + \boldsymbol{w}_k(t) \quad (1)$$

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

The base station has K queues, one for each user in the system. The arrival process of queue k, denoted by $A_k(t)$, is an ergodic process with *arrival rate* $\lambda_k \triangleq (1/T)\mathbb{E}[A_k(t)]$ (bits per channel use). The buffer size of queue k is denoted by $S_k(t)$ (bits).

An SDMA/TDMA policy is generally a function of the CSIT and of the state of the transmitter queues. For the systems considered in this paper, specific idealized models for CSIT and for the signaling strategies will be specified in Section III. We assume that at the beginning of each slot t, both a CSIT signal $\boldsymbol{\alpha}(t)$ and the queue buffer states $\{S_k(t) : k = 1,\dots,K\}$ are revealed to the transmitter.

We let $\boldsymbol{\alpha}_{1,\dots,t} = \{\boldsymbol{\alpha}(\tau) : \tau = 1,\dots,t\}$ denote the sequence of CSIT signals up to time t, and $\boldsymbol{S}_{1,\dots,t} = \{S_1(\tau),\dots,S_K(\tau) : \tau = 1,\dots,t\}$ denote the queue buffer state sequence up to time t. An SDMA/TDMA resource allocation policy is defined by a sequence of resource and rate allocation functions

$$\boldsymbol{P}(t) = \{p_{k,j}(t) : k = 1,\dots,K; j = 1,\dots,B\}$$

and

$$\boldsymbol{R}(t) = \{R_{k,j}(t) : k = 1,\dots,K; j = 1,\dots,B\}$$

for $t = 1,2,\dots,$ such that user k on each slot t transmits at rate $R_{k,j}(t)$ bits per channel over a fraction $p_{k,j}(t)$ of dimensions of stream $j = 1,\dots,B$, where both $p_{k,j}(t)$ and $R_{k,j}(t)$ are, in general, functions of $(\boldsymbol{\alpha}_{1,\dots,t}, \boldsymbol{S}_{1,\dots,t})$. Notice that the allocation functions depend *causally* on the CSIT and queue state sequences. Furthermore, $\boldsymbol{P}(t)$ must satisfy the SDMA/TDMA feasibility constraint

$$\sum_{k=1}^{K} p_{k,j}(t) \leq 1 \quad (2)$$

for each $j = 1,\dots,B$ and for all t.

Coding and decoding is performed on a slot-by-slot basis.[1] Decoding errors are handled by a standard automatic-repeat request (ARQ) protocol such that the unsuccessfully decoded information bits are left in the transmission buffer and shall be rescheduled for transmission at a later time. For a given signaling strategy and SDMA/TDMA policy, let $C_{k,j}(t)$ denote the supremum of the coding rates supported by the channel for user k on stream j in slot t. This rate depends on the instantaneous channel realization, and is therefore a random variable. The probability of decoding error for user k scheduled on stream j at time t, with rate $R_{k,j}(t)$, is given by the information outage probability $\mathbb{P}(C_{k,j}(t) \leq R_{k,j}(t))$. Under the above assumptions, the queue buffer states evolve in time according to the stochastic difference equation

$$S_k(t+1) = \left[S_k(t) - T\sum_{j=1}^{B} p_{k,j}(t)R_{k,j}(t) \cdot \mathbf{1}\left\{R_{k,j}(t) < C_{k,j}(t)\right\}\right]_+ + A_k(t) \quad (3)$$

for all $k = 1,\dots,K$, where $[\cdot]_+ \triangleq \max\{\cdot, 0\}$. The indicator function $1\{R_{k,j}(t) < C_{k,j}(t)\}$ reflects the fact that, due to the ARQ protocol, the corresponding information bits are removed from the transmission buffer only if decoding is successful.

Following the definition of system stability of [10], we define the buffer overflow function

$$g_k(S) = \limsup_{t\to\infty} \frac{1}{t}\sum_{\tau=1}^{t} \mathbf{1}\{S_k(\tau) > S\}$$

and say that the system is stable if $\lim_{S\to\infty} g_k(S) = 0$ for all $k = 1,\dots,K$. The SDMA/TDMA stability region Ω is the set of all arrival rates K-tuples $\boldsymbol{\lambda} \in \mathbb{R}_+^K$ such that there exists a feasible SDMA/TDMA policy for which the system is stable.

We introduce two further technical assumptions: A1) the channel vectors $\{\boldsymbol{h}_k(t) : k = 1,\dots,K\}$, the CSIT $\boldsymbol{\alpha}(t)$, and the arrival processes $\{A_k(t) : k = 1,\dots,K\}$ are jointly stationary ergodic and Markov; A2) for a given signaling strategy, for every sufficiently large t, the following Markov chain holds:

$$\boldsymbol{\alpha}_{1,\dots,t-1} \to \boldsymbol{\alpha}(t) \to \{C_{k,j}(t) : k = 1,\dots,K, j = 1,\dots,B\}. \quad (4)$$

In particular, A2) 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

$$\mathbb{P}\left(C_{k,j}(t) \leq R | \boldsymbol{\alpha}_{1,\dots,t}\right) = \mathbb{P}\left(C_{k,j}(t) \leq R | \boldsymbol{\alpha}(t)\right). \quad (5)$$

We have the following result.

Theorem 1 [Stability Region]: Under Assumptions A1 and A2, for any fixed signaling strategy, the system stability region is given by

$$\Omega = \mathrm{coh} \bigcup_{\boldsymbol{P}\in\mathcal{P}} \left\{ \boldsymbol{\lambda} \in \mathbb{R}_+^K \;:\; \lambda_k \leq \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha})R_{k,j}^{\mathrm{out}}(\boldsymbol{\alpha})\right] \right\} \quad (6)$$

where "coh" means *closure of the convex hull*, $\mathcal{P}$ is the set of *stationary* SDMA/TDMA policies $\boldsymbol{P} = \{p_{k,j}\}$ that map the current CSIT $\boldsymbol{\alpha}(t) = \boldsymbol{a}$ into an array $[p_{k,j}(\boldsymbol{a})]$ satisfying (2), and where we define the "conditional outage rate"

$$R_{k,j}^{\mathrm{out}}(\boldsymbol{a}) = \sup_{R\geq 0} R\left(1 - \mathbb{P}(C_{k,j} \leq R | \boldsymbol{\alpha} = \boldsymbol{a})\right). \quad (7)$$

The stability region is achieved by the stationary rate-allocation policy $\boldsymbol{R}^\star(\boldsymbol{a})$ achieving the supremization in (7).

Proof: See Appendix A □

[1]We assume that T is large enough such that the probability of decoding error is essentially characterized by the information outage probability [9].

The following comments are in order: 1) under the conditions assumed in this paper, we have that the rate-allocation function $\boldsymbol{R}^\star$ is optimal for any stationary $\boldsymbol{P} \in \mathcal{P}$ and any SDMA/TDMA signaling scheme. This reduces the problem of the stability-optimal resource-allocation policy to the determination of $\boldsymbol{P} \in \mathcal{P}$ alone; 2) under any stationary policy $(\boldsymbol{P}, \boldsymbol{R}^\star)$, the user average service rates $\widetilde{\mu}_k(t) = T\sum_{j=1}^{B} p_{k,j}(\boldsymbol{\alpha}(t))R_{k,j}^{\rm out}(\boldsymbol{\alpha}(t))$ (bits per slot) are linear functions of $\boldsymbol{P} \in \mathcal{P}$, and $\mathcal{P}$ is a convex set. It follows that the convex hull in (6) can be removed; 3) we can introduce the class $\mathcal{P}_{\text{on-off}}$ of *randomized* on–off SDMA/TDMA policies such that for every $\boldsymbol{P} \in \mathcal{P}$, there exists $\boldsymbol{P}' \in \mathcal{P}_{\text{on-off}}$ defined as follows: for all CSIT vectors $\boldsymbol{a}$ and $j = 1, \dots, B$, let user κ_j be served on the whole slot on stream j where $\kappa_j \in \{1, \dots, K\}$ is a random variable generated according to the probability mass function $(p_{1,j}(\boldsymbol{a}), \dots, p_{K,j}(\boldsymbol{a}))$. Clearly, Ω is achieved by restricting the union to the policies in $\mathcal{P}_{\text{on}-\text{off}}$. In practice, randomized on–off policies are preferable, since handling a single user per slot per stream is much easier. Therefore, we shall subsequently restrict to these policies; 4) from the convexity of Ω, points on the region boundary $\partial\Omega$ can be obtained by letting $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K) \in \mathbb{R}_+^K$ and finding $\lambda_k(\boldsymbol{\theta}) = \sum_{j=1}^{B} \mathbb{E}[p_{k,j}(\boldsymbol{\alpha})R_{k,j}^{\rm out}(\boldsymbol{\alpha})]$, for the policy $\boldsymbol{P} \in \mathcal{P}$ that solves the maximization problem

$$\max_{\boldsymbol{P}\in\mathcal{P}} \sum_k \theta_k \sum_{j=1}^{B} \mathbb{E}\left[p_{k,j}(\boldsymbol{\alpha})R_{k,j}^{\rm out}(\boldsymbol{\alpha})\right]. \tag{8}$$

All such points $\boldsymbol{\lambda}(\boldsymbol{\theta})$ lie on $\partial\Omega$. For given $\boldsymbol{\theta} \in \mathbb{R}_+^K$, the solution of (8) is readily obtained as

$$\widehat{p}_{k,j}(\boldsymbol{a}) = \begin{cases} 1, & k = \arg\max_{k'} \theta_{k'} R_{k',j}^{\rm out}(\boldsymbol{a}) \\ 0, & k \neq \arg\max_{k'} \theta_{k'} R_{k',j}^{\rm out}(\boldsymbol{a}). \end{cases} \tag{9}$$

This means that, as expected, the points on $\partial\Omega$ are achieved either by *deterministic* on–off policies that schedule the best user, given by $k = \arg\max_{k'} \theta_{k'} R_{k',j}^{\rm out}(\boldsymbol{a})$ on each stream j, or by convex combinations thereof (time sharing). Notice also 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 $\boldsymbol{\lambda} \in \Omega$, by definition, there exists some memoryless stationary policy $\boldsymbol{P}_{\boldsymbol{\lambda}} \in \mathcal{P}_{\text{on-off}}$ that stabilizes the system. However, in order to determine $\boldsymbol{P}_{\boldsymbol{\lambda}}$, 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 [10], [11]. This adaptive policy, referred to as "max-stability adaptive policy," is established by the following.

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

$$\widehat{p}_{k,j}(\boldsymbol{a}, \boldsymbol{S}) = \begin{cases} 1, & k = \arg\max_{k'} \theta_{k'} S_{k'} R_{k',j}^{\rm out}(\boldsymbol{a}) \\ 0, & k \neq \arg\max_{k'} \theta_{k'} S_{k'} R_{k',j}^{\rm out}(\boldsymbol{a}) \end{cases} \tag{10}$$

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

Theorem 2 follows by applying the theory of Lyapunov drift [10], [11] and follows closely the proof of [10, Th. 3]. It is omitted for the sake of space limitation.

III. STC Versus Opportunistic Beamforming

In this section, we make use of the general theory illustrated before to compare random opportunistic beamforming and STC. We assume that the channel vectors $\boldsymbol{h}_k(t)$ are mutually statistically independent for different index k and independent and identically distributed (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

$$h(t) = -\sum_{\ell=1}^{L} g_\ell h(t-\ell) + \nu(t) \tag{11}$$

where $\nu(t) \sim \mathcal{CN}(0, \sigma^2)$. We assume 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

$$\boldsymbol{\alpha}(t) = \{\boldsymbol{h}_k(t-\ell-d) : k = 1, \dots, K, \ell = 0, \dots, L-1\}.$$

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 II hold exactly. In practice, many other sources of uncertainty are present, such as noisy channel observations [12], a rate-constrained feedback link [13], [14], or an unquantized but noisy feedback link [12]. 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 easy immediate consequence of the Markov nature of the channel processes. Hence, our assumption yields an optimistic scenario in favor of the opportunistic schemes.

STC: In this case, $\boldsymbol{X}(t) \in \mathbb{C}^{T\times M}$ denotes the transmitted space–time codeword. STC yields only M-fold transmit diversity and no spatial multiplexing $(B = 1)$. In this case, we drop the stream index j for notation simplicity. The instantaneous capacity for user k at time t is given by $C_k(t) = \log(1 + \gamma|\boldsymbol{h}_k(t)|^2/M)$ and the corresponding conditional outage probability is

$$\mathbb{P}\left(C_k(t) \le R | \boldsymbol{\alpha}(t) = \boldsymbol{a}\right) = 1 - \mathcal{Q}_M\left(\sqrt{\frac{2^R - 1}{\gamma\sigma_e^2/2M}}, \sqrt{\frac{|\boldsymbol{g}_k(t)|^2}{\sigma_e^2/2}}\right) \tag{12}$$

where $\mathcal{Q}_M$ denotes generalized Marcum's Q-function, σ_e^2 denotes the prediction minimum mean-square error (MMSE)[2] of $\boldsymbol{h}_k(t)$ from $\boldsymbol{\alpha}(t)$, and where $\boldsymbol{g}_k(t)$ is the MMSE predictor of $\boldsymbol{h}_k(t)$ from $\boldsymbol{\alpha}(t)$. Expression (12) is easily obtained by using the fact that $\boldsymbol{h}_k(t)$ and $\boldsymbol{\alpha}(t)$ are jointly Gaussian (see [15]). Each user sends back (a suitably quantized version of) the estimated

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

channel gain $|\boldsymbol{g}_k(t)|^2/M$. Hence, the amount of feedback required by this scheme is very small.

Opportunistic Beamforming: We consider opportunistic beamforming using $1 \le B \le M$ mutually orthogonal pseudorandom beams, as proposed in [4] and [5]. The transmitted signal is given by

$$\boldsymbol{X}(t) = \sum_{j=1}^{B} \boldsymbol{s}_j(t)\boldsymbol{\phi}_j^T(t) \tag{13}$$

where $\boldsymbol{s}_j(t) \in \mathbb{C}^{T\times 1}$ is the signal associated with 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

$$\mathrm{SINR}_{k,j}(t) = \frac{\left|\boldsymbol{\phi}_j^T(t)\boldsymbol{h}_k(t)\right|^2}{B/\gamma + \sum_{m\neq j}\left|\boldsymbol{\phi}_m^T(t)\boldsymbol{h}_k(t)\right|^2}. \tag{14}$$

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 $C_{k,j}(t) = \log(1 + \mathrm{SINR}_{k,j}(t))$. In the schemes analyzed in [4] and [5], each user measures its received SINR for each of the B transmit beams and feeds back its best SINR and the index of the beam achieving it. In these works, it is not clear how the SINRs are estimated and at what rate the random beamforming vectors change. For the sake of simplicity, we make the *optimistic* assumption that the beamforming vectors sequence $\{\boldsymbol{\phi}_m\}$ is perfectly known *a priori* by the receivers that make use of this knowledge in order to estimate the channels.[3] Furthermore, they feed back all their B SINR values and not only the best one. Under this assumption, the ability of estimating the channel is independent of the speed of variation of the beamforming vectors, and we let the matrix $[\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_B]$ drawn i.i.d. according to the Haar measure (the beamforming vectors form a random orthonormal B-dimensional basis at any point in time).

Computation of the conditional outage probability in this case is also possible by using standard methods of Hermitian quadratic forms of Gaussian random variables. The details of this calculation are given in [15]. This probability depends only on B real numbers that must be fed back by each user. Hence, also in this case, the amount of feedback required is moderately small.

IV. Numerical Results

We considered mutually independent arrival processes given by $A_k(t) = \sum_{j=1}^{M_k(t)} b_{k,j}(t)$, where $M_k(t)$ is an i.i.d. Poisson distributed sequence that counts the number of packets arriving at the kth 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 arriving in a slot (T channel uses). We

[3]Notice that in practical code-division multiple-access (CDMA) systems, each user is synchronized with the random spreading/scrambling code of the base station. Therefore, this assumption is, indeed, realistic also in a practical implementation.

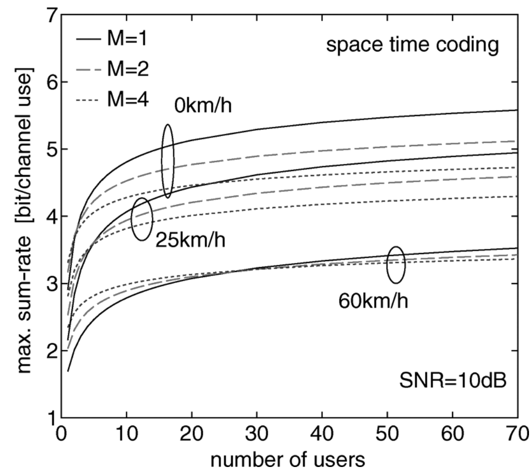


Fig. 1. Maximum sum-rate versus the number of users (STC).

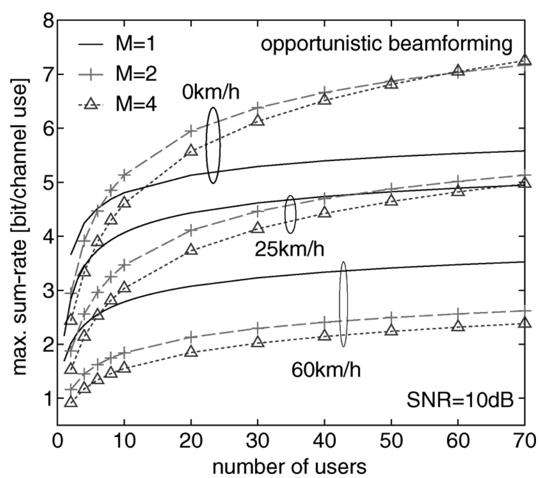


Fig. 2. Maximum sum-rate versus the number of users (beamforming with $B = M$).

considered a Gauss–Markov process of order $L = 5$, where the coefficients in (11) are chosen in order to approximate (see [16] and [17]) Jakes' autocorrelation model, typical of wireless mobile channels. Inspired by the high-data-rate (HDR) system [1], we let the duration of a slot be 1.67 ms, 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.

A. 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 adaptive policy. The maximum sum-rate is obtained by scheduling at each time and stream the user with the largest outage rate, irrespectively of the buffers state.

Figs. 1 and 2 show the maximum sum-rate versus 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

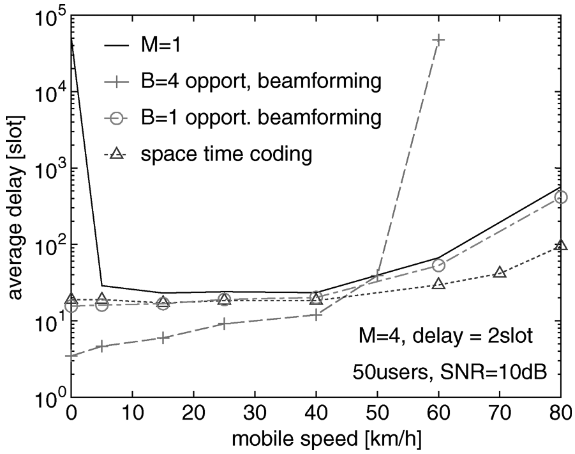


Fig. 3. Average delay versus mobile speed for the case of $K = 50$ users and $M = 4$ transmit antennas. The cases of random beamforming with $B = 1$ and $B = 4$ beams, and STC are compared. The case $M = 1$ (single antenna) is also included for comparison.

Fig. 2, 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 [4] with $M > 1$ and $B = 1$.

In Fig. 1, we observe that the 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. 2, 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.

B. Average Delay Versus 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). The average delay is obtained from the average buffer size by Little's theorem. We consider the symmetric arrival case. Fig. 3 shows the average delay versus the mobile speed for a system with 50 users with $M = 4$ antennas, in the cases of STC, random beamforming with $B = 1$, and random beamforming with $B = 4$, respectively. The case $M = 1$ is included for comparison and is referred to as 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 channel-hardening effect: ergodicity which is lost in time is eventually recovered in the spatial domain by increasing the number of transmit antennas.

When the channel varies slowly, so that the CSIT quality is good, opportunistic random beamforming decreases the average delay by making the channel vary almost i.i.d. and choosing the best set of users at each time. In our example, for speeds below 40 km/h, opportunistic beamforming with $B = M$ beams achieves the smallest delay. As the speed increases (i.e., the quality of CSIT degrades), STC outperforms random beamforming due to its better outage rate. Interestingly, the opportunistic beamforming systems become unstable (the average delay diverges) with $M = 4$ and speed 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 [4] is outperformed by the $B = M$ random beamforming scheme for low mobile speed and by STC for higher mobile speed. Therefore, it is not useful in any mobile speed range.

V. Conclusion

We have compared transmit diversity (STC) and opportunistic random beamforming for downlink transmission in a mobile wireless system where the base station has multiple antennas and the user terminals have one antenna. Beyond their simplicity, these schemes are also relevant since they are currently considered for standardization in evolutionary 3G systems. Our comparison assumes random arrivals and considers the transmission average delay and system stability, under a general max-stability SDMA/TDMA scheduling and rate-allocation policy. 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 showed that the ability of accurately predicting the channel (or beams) SINRs has a fundamental impact on the performance of opportunistic beamforming schemes. In the case of mobile communications with a feedback delay, the transmitter cannot have perfect CSIT due to the channel non-zero Doppler bandwidth. Hence, 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, STC 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.

Appendix

A. Proof of Theorem 1

For a fixed policy $\{\boldsymbol{P}(t), \boldsymbol{R}(t)\}$, the instantaneous service rate (information bits per slot) for user k is given by [see (3)]

$$\mu_k(t) = T\sum_{j=1}^{B} p_{k,j}(t)R_{k,j}(t)\mathbf{1}\{R_{k,j}(t) < C_{k,j}(t)\}.$$

The results in [10] applied to our setting yield that under the assumption of Section II, the system is stable if and only if

$$\lambda_k \leq \underline{\mu}_k = \liminf_{t\to\infty} \frac{1}{tT}\sum_{\tau=1}^{t} \mu_k(\tau), \quad k = 1, \ldots, K. \tag{15}$$

Let $\widetilde{\Omega}$ denote the stability region of a new system with channel state $\boldsymbol{\alpha}(t)$ and feasible rate $\boldsymbol{R}^{\text{out}}(t)$. The instantaneous service rate (information bits per slot) of user k in the new system is given by

$$\widetilde{\mu}_k(t) = T \sum_{j=1}^{B} p_{k,j}(t) R_{k,j}^{\mathrm{out}}(\boldsymbol{\alpha}(t))$$

which is linear (hence, concave) and non-decreasing in the elements $p_{k,j}(t)$ of $\boldsymbol{P}(t)$. [10, Th. 1] yields that the stability region $\widetilde{\Omega}$ of the new system is given by (6). Hence, the theorem is proved if we show that the two stability regions coincide.

By restricting to the stationary resource-allocation policies $\boldsymbol{P} \in \mathcal{P}$ and to the rate-allocation policy $R^\star$ given in *Theorem 1*, the arrival processes $\{A_k(t)\}$ and the service rates $\{\mu_k(t)\}$ become jointly Markov modulated [10], and it is immediate to show, by ergodicity, that

$$\liminf_{t\to\infty} \frac{1}{tT} \sum_{\tau=1}^{t} \mu_k(\tau) = \sum_{j=1}^{B} \mathbb{E}[p_{k,j}(\boldsymbol{\alpha}) R_{k,j}^{\mathrm{out}}(\boldsymbol{\alpha})].$$

Hence, any point $\boldsymbol{\lambda} \in \widetilde{\Omega}$ is also in Ω, and it is stabilized by a policy $\{\boldsymbol{P}, \boldsymbol{R}^\star\}$ for some $\boldsymbol{P} \in \mathcal{P}$. This shows that $\widetilde{\Omega} \subseteq \Omega$.

In order to show that $\Omega \subseteq \widetilde{\Omega}$, we assume that the channels and CSIT signals take on values in finite discrete sets $\mathcal{H}$ and $\mathcal{A}$, respectively (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

$$\begin{aligned}
&\frac{|N_{\boldsymbol{a}}(t_\epsilon)|}{t_\epsilon} \le P_\alpha(\boldsymbol{a}) + \epsilon \\
&\frac{1}{t_\epsilon T} \sum_{\tau=1}^{t_\epsilon} \sum_{j=1}^{B} \mu_{k,j}(\tau) \ge \underline{\mu}_k - \epsilon \\
&\frac{1}{|N_{\boldsymbol{a},\boldsymbol{p},\boldsymbol{r}}(t_\epsilon)|} \sum_{\tau \in N_{\boldsymbol{a},\boldsymbol{p},\boldsymbol{r}}(t_\epsilon)} \mathbf{1}\left\{ r_{k,j} < \overline{R}_{k,j}(\tau) \right\} \\
&\qquad\qquad \le 1 - \mathbb{P}(\overline{R}_{k,j} \le r_{k,j} | \boldsymbol{\alpha} = \boldsymbol{a}) + \epsilon \qquad (16)
\end{aligned}$$

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

The last inequality in (16) follows, since, by assumption, any feasible resource-allocation policy is a causal function of the CSIT process, the CSIT process is ergodic, and $C_{k,j}(\tau)$ is independent of $\boldsymbol{\alpha}_{1,\dots,\tau-1}$ given $\boldsymbol{\alpha}(\tau)$.

Consider any stable arrival-rate vector $\boldsymbol{\lambda}$. By the necessary condition of [10, Lemma 1] (see above), there exists some not necessarily stationary policy $\{\boldsymbol{P}(t), \boldsymbol{R}(t)\}$ such that, for all $k = 1, \dots, K$

$$\lambda_k \le \underline{\mu}_k \le \frac{1}{t_\epsilon T} \sum_{\tau=1}^{t_\epsilon} \sum_{j=1}^{B} \mu_{k,j}(\tau) + \epsilon. \qquad (17)$$

By using (16), we have

$$\begin{aligned}
\lambda_k &\le \sum_{\boldsymbol{a}\in\mathcal{A}} P_\alpha(\boldsymbol{a}) \frac{1}{|N_{\boldsymbol{a}}(t_\epsilon)|} \sum_{\tau \in N_{\boldsymbol{a}}(t_\epsilon)} \sum_{j=1}^{B} p_{k,j}(\tau) R_{k,j}(\tau) \\
&\quad \mathbf{1}\{R_{k,j}(\tau) < C_{k,j}(\tau)\} + \epsilon' \\
&\le \sum_{\boldsymbol{a}\in\mathcal{A}} P_\alpha(\boldsymbol{a}) \sum_{\boldsymbol{p},\boldsymbol{r}} \frac{|N_{\boldsymbol{a},\boldsymbol{p},\boldsymbol{r}}(t_\epsilon)|}{|N_{\boldsymbol{a}}(t_\epsilon)|} \frac{1}{|N_{\boldsymbol{a},\boldsymbol{p},\boldsymbol{r}}(t_\epsilon)|} \\
&\quad \sum_{\tau \in N_{\boldsymbol{a},\boldsymbol{p},\boldsymbol{r}}(t_\epsilon)} \sum_{j=1}^{B} p_{k,j} r_{k,j} \mathbf{1}\{r_{k,j} < C_{k,j}(\tau)\} + \epsilon' \\
&\le \sum_{\boldsymbol{a}\in\mathcal{A}} P_\alpha(\boldsymbol{a}) \sum_{\boldsymbol{p},\boldsymbol{r}} \frac{|N_{\boldsymbol{a},\boldsymbol{p},\boldsymbol{r}}(t_\epsilon)|}{|N_{\boldsymbol{a}}(t_\epsilon)|} \\
&\quad \sum_{j=1}^{B} p_{k,j} r_{k,j} \left(1 - \mathbb{P}(C_{k,j} \le r_{k,j} | \boldsymbol{\alpha} = \boldsymbol{a})\right) + \epsilon'' \\
&\overset{(a)}{\le} \sum_{\boldsymbol{a}\in\mathcal{A}} P_\alpha(\boldsymbol{a}) \sum_{j=1}^{B} \widetilde{p}_{k,j}(\boldsymbol{a}) R_{k,j}^{\mathrm{out}}(\boldsymbol{a}) + \epsilon'' \\
&= \widetilde{\lambda}_k + \epsilon'' \qquad (18)
\end{aligned}$$

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

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

Since, by assumption, $\sum_{k=1}^{K} p_{k,j} \le 1$ for all $j = 1, \dots, B$ on every slot, and since $\sum_{\boldsymbol{p}} (|N_{\boldsymbol{a},\boldsymbol{p}}(t_\epsilon)| / |N_{\boldsymbol{a}}(t_\epsilon)|) = 1$, it follows that $\sum_{k=1}^{K} \widetilde{p}_{k,j}(\boldsymbol{a}) \le 1$ for all j and $\boldsymbol{a} \in \mathcal{A}$, i.e., the stationary policy $\{\widetilde{p}_{k,j}\}$ defined above is feasible. Hence, $\widetilde{\boldsymbol{\lambda}}$ with kth component given by the last line of (18) is a point in $\widetilde{\Omega}$. Since $\epsilon > 0$ is arbitrary, and $\epsilon'' \to 0$ as $\epsilon \to 0$, we have that $\boldsymbol{\lambda} \le \widetilde{\boldsymbol{\lambda}} \in \widetilde{\Omega}$, which eventually implies that $\widetilde{\Omega}$ and Ω coincide.

References

[1] 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, Jul. 2000.

[2] *CDMA2000 High Rate Packet Data Air Interface Specification*, TIA/EIA/3GPP2 Std. IS-856/3GPP2 C.S.0024, v3.0, Dec. 2001.

[3] T. E. Kolding *et al.*, "High speed downlink packet access: WCDMA evolution," *IEEE Veh. Technol. Soc. News*, pp. 4–10, Feb. 2003.

[4] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[5] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[6] R. Gozali, R. M. Buehrer, and B. D. Woerner, "The impact of multiuser diversity on space-time block coding," *IEEE Commun. Lett.*, vol. 7, no. 5, pp. 213–215, May 2003.

[7] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1893–1909, Sep. 2004.

[8] H. Viswanathan and S. Venkatesan, "The impact of antenna diversity in packet data systems with scheduling," *IEEE Trans. Commun.*, vol. 52, no. 4, pp. 546–549, Apr. 2004.

[9] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[10] M. J. Neely, E. Modiano, and C. E. Rohrs, "Power allocation and routing in multibeam satellites with time-varying channels," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 138–152, Feb. 2003.

[11] E. M. Yeh and A. S. Cohen, "Information theory, queueing, and resource allocation in multi-user fading communications," in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, Mar. 2004, CD-ROM.

[12] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback on wireless channels," in *Proc. IEEE WCNC*, New Orleans, LA, Mar. 2005, CD-ROM.

[13] S. Sanayei and A. Nosratinia, "Opportunistic beamforming with limited feedback," in *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 2005, CD-ROM.

[14] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[15] M. Kobayashi, "On the use of multiple antennas in a downlink of wireless systems," Ph.D. dissertation, ENST, Paris, France, Jun. 2005.

[16] W. Turin and R. Nobelen, "Hidden Markov modeling of flat fading channels," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1809–1817, Dec. 1998.

[17] K. E. Baddour and N. C. Beaulieu, "Autoregressive models for fading channel simulation," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1650–1662, Jul. 2005.

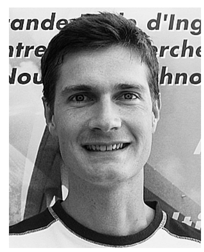
Mari Kobayashi (M'06) received the B.E. degree in electrical engineering from Keio University, Yokohama, Japan, in 1999, and the M.S. degree in mobile radio, and the Ph.D. degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 2000 and 2005, respectively.

Since November 2005, she has been a Postdoctoral Researcher with the Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona, Spain. Her current research interests include MIMO communcation systems and multiuser communication theory.

Giuseppe Caire (S'92–M'94–SM'03–F'05) was born in Torino, Italy, in 1965. He received the B.Sc. degree in electrical engineering from the Politecnico di Torino, Torino, Italy, in 1990, the M.Sc. degree in electrical engineering from Princeton University, Princeton, NJ, in 1992, and the Ph.D. degree from the Politecnico di Torino in 1994.

He was with the European Space Agency (ESTEC), Noordwijk, The Netherlands, from May 1994 to February 1995. He was visiting Princeton University in the summer of 1997 and Sydney University, Sydney, Australia, in the summer of 2000. He was an Assistant Professor in Telecommunications with the Politecnico di Torino, an Associate Professor with the University of Parma, Parma, Italy, a Professor with the Department of Mobile Communications, Eurecom Institute, Sophia Antipolis, France, and he is currently a Professor with the Electrical Engineering Department, University of Southern California, Los Angeles, CA. His current interests are in the field of communications theory, information theory, and coding theory, with particular focus on wireless applications.

Dr. Caire was a recipient of the AEI G. Someda Scholarship in 1991, was a recipient of the COTRAO Scholarship in 1996, and of a CNR Scholarship in 1997. He served as an Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS in 1998–2001, and as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY in 2001–2003. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, and the Joint IT/Comsoc Best Paper Award in 2004. Since November 2004, he has been a member of the Board of Governors of the IEEE Information Theory Society.



David Gesbert (S'97–M'98–SM'06) received the Ph.D degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1997.

He is a Professor with the Eurecom Institute, Sophia Antipolis, France. From 1993 to 1997, he was with France Telecom Research, Paris. From April 1997 to October 1998, he was a Research Fellow with the Information Systems Laboratory, Stanford University, Stanford, CA. He took part in the founding team of Iospan Wireless Inc, San Jose, CA, a startup company pioneering MIMO-OFDM. Starting in 2001, he has been with the University of Oslo, Oslo, Norway, as an adjunct professor. He has published about 100 papers and holds several patents, all in the area of signal processing and communications.

Dr. Gesbert co-edited several special issues for the IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS (JSAC, 2003), EURASIP JASP (2004), and IEEE COMMUNICATIONS MAGAZINE (2006). He is an elected member of the IEEE Signal Processing for Communications Technical Committee. He authored or co-authored papers winning the 2004 IEEE Best Tutorial Paper Award (Communications Society) for a 2003 JSAC paper on MIMO systems, 2005 Best Paper (Young Author) Award for Signal Processing Society journals, and the Best Paper Award for the 2004 ACM MSWiM workshop. He was co-organizer of the IEEE Workshop on Signal Processing Advances in Wireless Communications, 2006 (Cannes, France).