Transmit Diversity v. Spatial Multiplexing in Modern MIMO Systems

Angel Lozano, and Nihar Jindal

Abstract

A contemporary perspective on transmit antenna diversity and spatial multiplexing is provided. It is argued that, in the context of most modern wireless systems and for the operating points of interest, transmission techniques that utilize all available spatial degrees of freedom for multiplexing outperform techniques that explicitly sacrifice spatial multiplexing for diversity. Reaching this conclusion, however, requires that the channel and some key system features be adequately modeled and that suitable performance metrics be adopted; failure to do so may bring about starkly different conclusions. As a specific example, this contrast is illustrated using the 3GPP Long-Term Evolution system design.

I. Introduction

Multipath fading is one of the most fundamental features of wireless channels. Because multiple received replicas of the transmitted signal sometimes combine destructively, there is a significant probability of severe fades. Without any means of mitigating such fading, ensuring reasonable reliability requires hefty power margins.

Fortunately, fades are very localized in space and frequency: a change in the transmitter or receiver location (on the order of a carrier wavelength) or in the frequency (on the order of the inverse of the propagation delay spread) leads to a roughly independent realization of the fading process. Motivated by this *selectivity*, the concept of *diversity*

Angel Lozano is with Universitat Pompeu Fabra, Barcelona 08005, Spain. His work is partially supported by the project CONSOLIDER-INGENIO 2010 CSD2008-00010 ”COMONSENS”.

Nihar Jindal is with the University of Minnesota, Minneapolis, MN55455, USA. His work was partially conducted during a visit to UPF under the sponsorship of Project TEC2006-01428.

is borne: rather than making the success of a transmission entirely dependent on a single fading realization, hedge the transmission's success across multiple realizations in order to decrease the probability of failure. Hedging or diversifying are almost universal actions in the presence of uncertainty, instrumental not only in communications but also in other fields as disparate as economics or biology.

In communications specifically, the term 'diversity' has, over time, acquired different meanings, to the point of becoming overloaded. It is used to signify:

- Variations of the underlying channel in time, frequency, space, etc.
- Performance metrics related to the error probability. Adding nuance to the term, more than one such metric can be defined (cf. Section IV).
- Transmission and/or reception techniques designed to improve the above metrics.

In this paper, we carefully discriminate these meanings. We use 'selectivity' to refer to channel features, which are determined by the environment (e.g., propagation and user mobility) and by basic system parameters (e.g., bandwidth and antenna spacing). In turn, the term 'diversity' is reserved for performance metrics and for specific transmit/receive techniques, both of which have to do with the signal. Note that channel selectivity is a necessary condition for diversity strategies to yield an improvement in some diversity metric.

A. Diversity over Time

Archaic electrical communication systems from a century ago already featured primitive forms of diversity, where operators manually selected the receiver with the best quality. Automatic selection of the strongest among various receivers was discussed as early as 1930 [1]. This naturally led to the suggestion of receive antenna combining, initially for microwave links [2]–[5]. MRC (Maximum Ratio Combining), by far the most ubiquitous combining scheme, was first proposed in 1954 [6]. In addition to receive antenna combining, other approaches such as repeating the signal on two or more frequency channels were also considered for microwave links [7]. (Systems were still analog and thus coding and interleaving was not an option.) Given the cost of spectrum, though, approaches that consume additional bandwidth were naturally unattractive and thus the use of antennas quickly emerged as the preferred diversity approach.

Recognizing this point, receive antenna combining was debated extensively in the 1950s [8]–[11] and has since been almost universally adopted for use at base station sites. The industry, however, remained largely ambivalent about multiple antennas at mobile devices. Although featured in early AMPS trials in the 1970s, and despite repeated favorable studies (e.g., [12]), until recently its adoption had been resisted.[1]

Multiple base station antennas immediately allow for uplink receive diversity. It is less clear, on the other hand, how to achieve diversity in the downlink using only multiple transmit antennas. In Rayleigh fading, transmitting each symbol from every antenna simultaneously is equivalent to using a single transmit antenna [14, Section 7.3.2]. Suboptimal schemes were formulated that convert the spatial selectivity across the transmit antennas into effective time or frequency selectivity. In these schemes, multiple copies of each symbol are transmitted from the various antennas, each subject to either a phase shift [15] or a time delay [16]. From the standpoint of the receiver, then, the effective channel that the signal has passed through displays enhanced time or frequency selectivity and thus a diversity advantage can be reaped with appropriate coding and interleaving (cf. Section I-C).

More refined transmit diversity techniques did not develop until the 1990s. Pioneered in [17], these techniques blossomed into OSTBC (orthogonal space-time block codes) [18] and, subsequently, onto space-time codes at large. Albeit first proposed for single-antenna receivers, OSTBCs can also be used in MIMO (multiple-input multiple-output) communication, i.e., when both transmitter and receiver have a multiplicity of antennas. This yields additional diversity, and thus reliability, but no increases in the number of information symbols per MIMO symbol.

Concurrently with space-time coding, the principles of spatial multiplexing were also formulated in the 1990s [19]–[22]. The tenet in spatial multiplexing is to transmit different symbols from each antenna, and have the receiver discriminate these symbols by taking advantage of the fact that each transmit antenna has a different spatial signature at the receiver (because of spatial selectivity). This does allow for an increased number

[1]The sole exception was the Japanese PDC system [13], which supported dual-antenna terminals since the early 1990's.

of information symbols per MIMO symbol; depending on the particular transmission technique used, reliability benefits may or may not be reaped.

Altogether, the powerful thrust promised by MIMO is finally bringing multiantenna devices to the marketplace. Indeed, MIMO is an integral feature of emerging wireless systems such as 3GPP LTE (Long-Term Evolution) [23], 3GPP2 Ultra Mobile Broadband, and IEEE 802.16 WiMAX [24].

B. Overview of Work

With the advent of MIMO, a choice needs to be made between transmit diversity techniques, which increase reliability (decrease probability of error), and spatial multiplexing techniques, which increase rate but not necessarily reliability. Applications requiring extremely high reliability seem well suited for transmit diversity techniques whereas applications that can smoothly handle loss appear better suited for spatial multiplexing. It may further appear that the SNR (signal-to-noise ratio) and the degree of channel selectivity should also affect this decision.

Our findings, however, differ strikingly from the above intuitions. The main conclusion is that techniques utilizing all available spatial degrees of freedom for multiplexing outperform, at the operating points of interest for modern wireless systems, techniques that explicitly sacrifice spatial multiplexing for transmit diversity. Thus, from a performance perspective there essentially is no decision that need be made between transmit diversity and multiplexing in contemporary MIMO systems (cf. [25]). This conclusion is established on the basis of a suboptimal multiplexing technique, and it is only strengthened with optimal multiplexing.

There are a number of different arguments that lead to this conclusion, and which will be elaborated upon:

- Modern systems use link adaptation to maintain a target error probability and there is essentially no benefit in operating below this target. This makes diversity metrics, which quantify the speed at which error probability is driven to zero with the signal-to-noise ratio, beside the point.
- Wireless channels in modern systems generally exhibit a notable amount of time and frequency selectivity, which is naturally converted into diversity benefits through

coding and interleaving. This renders additional transmit diversity superfluous.

- Block error probability is the relevant measure of reliability. Since the channel codes featured in contemporary systems allow for operation close to information-theoretic limits, such block error probability is well approximated by the mutual information outages. Although uncoded error probability is often quantified, this is only an indirect performance measure and incorrect conclusions can be reached by considering only uncoded performance.

It is also imperative to recognize that the notion of diversity is indelibly associated with channel uncertainty. If the transmitter has instantaneous CSI (channel-state information), then it can match the rate to the channel rendering the error probability dependent only on the noise. Diversity techniques, which aim precisely at mitigating the effects of channel uncertainty, are then beside the point. Although perhaps evident, this point is often neglected. In some models traditionally used to evaluate diversity techniques, for instance, the channel fades very slowly yet there is no transmitter adaptation. As we shall see, these models do not reflect the operating conditions of most current systems.

C. Time/Frequency Diversity v. Antenna Diversity

Perhaps the simplest manifestation of the efficacy of diversity in the aforementioned traditional settings is receive antenna combining: if two receive antennas are sufficiently spaced, the same signal is received over independently faded paths. Even with simple selection combining, this squares the probability of error; MRC performs even better.

Based upon the specifics of receive antenna combining, it may appear that multiple, independently faded copies of the same signal are required to mitigate fading. Although this is an accurate description of receive combining, it is an overly stringent requirement in general. This point is clearly illustrated if one considers a frequency-selective channel. One simple but naïve method of mitigating fading in such a channel is to repeat the same signal on two sufficiently spaced frequency channels. Unlike receive combining, this technique doubles the number of symbols transmitted and therefore the necessary bandwidth. Is repetition, which seems inefficient, the only way to take advantage of frequency diversity? It is not—if coding is taken into consideration. By applying a channel code to a sequence of information bits, the same benefit is gained by transmitting different portions of the coded block over different frequency channels. No repetition is necessary; rather, information bits are coded and interleaved, and then the first half of the coded block is transmitted on the first frequency and the other half on the second frequency. The information bits can be correctly decoded as long as both frequencies are not badly faded. The same principle applies to time selectivity: instead of repeating the same signal at different time instants, transmit a coded and interleaved block over an appropriate time period [26].[2]

II. Modeling Modern Wireless Systems

Wireless systems have experienced dramatic changes as they evolved from their initial analog forms to today's advanced digital formats. Besides MIMO, features of modern systems—that in many cases were completely absent in earlier designs—include:

- Wideband channelizations and OFDM.
- Packet switching, complemented with time- and frequency-domain scheduling for low-velocity users.
- Powerful channel codes [27]–[29].
- Link adaptation, and specifically rate control via variable modulation and coding [30].
- ARQ (automatic repeat request) and H-ARQ (hybrid-ARQ) [31].

These features have had a major impact on the operational conditions:

- There is a target block error probability, on the order of 1%, at the output of the decoder. (When H-ARQ is in place, this target applies at termination.) A link

[2]When explaining the exploitation of selectivity through coding and interleaving, it is important to dispel the misconception that channel coding incurs a bandwidth penalty. If the constellation is kept fixed, then coding does reduce the rate relative to an uncoded system. However, there is no rate penalty if the constellation size is flexible as in modern systems. For instance, a system using QPSK with a rate-1/2 binary code and an uncoded BPSK system both have an information rate of 1 bit/symbol. For a reasonably strong code, though, the coded system will achieve a considerably smaller bit error probability than the uncoded one. More importantly, the advantage of the coded system in terms of *block* error probability is even larger and this advantage increases with blocklength: the block error probability of a coded system decreases with the blocklength whereas, without coding, it actually increases with the blocklength. As will be emphasized throughout the paper, modern wireless systems cannot be conceived without powerful channel coding.

adaptation loop is then tasked with selecting the rate in order to maintain the performance tightly around this operating point. The rationale for this is two-fold:

i) There is little point in spending resources pushing the error probability on the traffic channels much below the error probability on the control plane, which, by its very nature (short messages and tight latency requirements), cannot be made arbitrarily small [32].

ii) Lower error probabilities often do not improve end-to-end performance: in some applications (e.g., voice) there is simply no perceivable improvement in the user experience while, in others (e.g., data communication requiring very high reliability), it is more cost effective to let the upper protocol layers handle the losses [33].

- The fading of low-velocity users can be tracked and fed back to the transmitter thereby allowing for link adaptation to the supportable rate, scheduling on favorable time/frequency locations, and possibly beamforming and precoding.
- The channels of high-velocity users vary too quickly in time to allow for feedback of CSI or even of the supportable rates. Thus, the signals of such users are dispersed over the entire available bandwidth thereby taking advantage of extensive frequency selectivity. In addition, time selectivity is naturally available because of the high velocity.

The above points evidence the disparity between the low- and high-velocity regimes and hence it is necessary to distinguish between them.

III. The Low-Velocity Regime

At low velocities, timely feedback regarding the current state of the channel becomes feasible. This fundamentally changes the nature of the communication problem: all uncertainty is removed except for the noise. With powerful coding handling that remaining uncertainly, outages are essentially eliminated.[3] Transmit diversity techniques, whose goal is precisely to reduce outages, are then beside the point. Rate maximization

[3]The rate supported by the channel may be essentially zero at some time/frequency points, but with proper link adaptation this does not constitute an outage in the sense that data is not lost (cf. Eq. 4).

becomes the overriding transmission design principle, and the optimum strategy in this known-channel setting is spatial waterfilling [21].

Although the above consideration posited perfect CSI at the transmitter, it also extends to imperfect-CSI settings (caused by limited rate and/or delay in the link adaptation loop). At a minimum, the supportable rate can be fed back; this still removes outages. Additional CSI feedback enables adaptive techniques such as scheduling, power control, beamforming and precoding [34].[4]

In multiuser settings, furthermore, CSI feedback is collected from many users and time- and frequency-domain scheduling offers additional degrees of freedom. In this case, transmit diversity techniques can actually be detrimental because they harden the possible transmission rates to different users thereby reducing potential multiuser scheduling gains [35], [36].

These conclusions apply almost universally to indoor systems, which conform to this low-velocity regime, as long as their medium-access control features the necessary functionalities. In outdoor systems, they apply to stationary and pedestrian users.

IV. The High-Velocity Regime

Having established that diversity is not an appropriate perspective in the low-velocity regime, we henceforth focus exclusively on the high-velocity regime. This is the regime of interest for vehicular users in outdoor systems. At high velocities, the fading (and therefore the time-varying mutual information) is too rapid to be tracked. The link adaptation loop can therefore only match the rate to the average channel conditions. The scheduler, likewise, can only respond to average conditions and thus it is not possible to transmit only to users with favorable instantaneous channels; we thus need not distinguish between single-user and multiuser settings.

[4]Feedback mechanisms are sometimes studied under the assumption that they convey information regarding the transmit strategy, e.g., which beamformer or precoder to use, but not regarding rate selection, in which case outages still occur. This, however, is not well aligned with modern system designs in which rate control is paramount.

A. Channel Model and Performance Metrics

Let n_{T} and n_{R} denote, respectively, the number of transmit and receive antennas. Assuming that OFDM (orthogonal frequency division multiplexing), the prevalent signalling technique in contemporary systems, is used to decompose a possibly frequency-selective channel into N parallel, non-interfering tones, the received signal on the ith tone is

$$\mathbf{y}_i = \mathbf{H}_i \mathbf{x}_i + \mathbf{n}_i \tag{1}$$

where $\mathbf{H}_i$ is the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ channel matrix on that tone, $\mathbf{y}_i$ is the $n_{\mathrm{R}} \times 1$ received signal, $\mathbf{n}_i$ is the $n_{\mathrm{R}} \times 1$ thermal noise, IID circularly symmetric complex Gaussian with unit variance, and $\mathbf{x}_i$ is the $n_{\mathrm{T}} \times 1$ transmitted signal subject to a power constraint SNR, i.e., $E[||\mathbf{x}_i||^2] \leq \mathsf{SNR}$. The receiver has perfect knowledge of the N channel matrices, the joint distribution of which is specified later.

For a particular realization of $\mathbf{H}_1, \ldots, \mathbf{H}_N$, the average mutual information thereon is

$$\mathcal{I}(\mathsf{SNR}) = \frac{1}{N} \sum_{i=1}^{N} I(\mathbf{x}_i; \mathbf{y}_i). \tag{2}$$

This quantity is in bits per (complex) modulation symbol, and thus it represents spectral efficiency in bits/s/Hz under the standard assumption of one symbol/s/Hz. The mutual information on each tone is determined by the chosen signal distribution. If the signals are IID complex Gaussian[5] with $E[\mathbf{x}_i \mathbf{x}_i^\dagger] = \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \mathbf{I}$, then

$$I(\mathbf{x}_i; \mathbf{y}_i) = \log_2 \det \left(\mathbf{I} + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \mathbf{H}_i \mathbf{H}_i^\dagger \right). \tag{3}$$

Since approaching this mutual information may entail high complexity, simpler MIMO strategies with different (lower) mutual informations are often used. Expressions for these are given later in this section.

Once a transmission strategy has been specified, the corresponding outage probability for rate R (bits/s/Hz) is then

$$P_{\mathsf{out}}(\mathsf{SNR}, R) = \Pr\{\mathcal{I}(\mathsf{SNR}) < R\}. \tag{4}$$

[5]Actual systems use discrete constellations, for which counterparts to (3) exist in integral form [37]. As long as the cardinality of the constellation is large enough relative to the SNR, the gap between the actual mutual information and (3) is small and is inconsequential to our conclusions.

With suitably powerful codes, the error probability when not in outage is very small and therefore the outage probability is an accurate approximation for the actual block error probability [38]–[40]. We shall therefore use both notions interchangeably henceforth.

As justified in Section II, modern systems operate at a target error probability. Hence, the primary performance metric is the maximum rate, at each SNR, such that this target is not exceeded, i.e.,

$$R_\epsilon(\mathsf{SNR}) = \max_{\zeta}\{\zeta : P_{\mathsf{out}}(\mathsf{SNR}, \zeta) \leq \epsilon\} \tag{5}$$

where ϵ is the target.

B. The Outage-Rate Tradeoff and the DMT

Eq. (4) fully specifies the tradeoff between outage and rate at any particular SNR, but closed forms do not exist in general for (4). This led to the introduction of metrics whose tradeoff can be more succinctly characterized. In particular, the diversity order was introduced as a proxy for the outage probability. The traditional notion of diversity order equals the asymptotic slope of the outage-SNR curve (in log-log scale) for a fixed R. Although meaningful in early wireless systems, where R was indeed fixed, this is not particularly indicative of contemporary systems in which R is increased with SNR. A more general formulation was introduced in [41], where R depends on SNR according to some function $R = f(\mathsf{SNR})$. The diversity order

$$d = -\lim_{\mathsf{SNR}\to\infty} \frac{\log P_{\mathsf{out}}(\mathsf{SNR}, f(\mathsf{SNR}))}{\log \mathsf{SNR}} \tag{6}$$

still captures the asymptotic slope of the outage-SNR curve (in log-log scale), albeit now for increasing R. A proxy for rate, termed the multiplexing gain, was further introduced as

$$r = \lim_{\mathsf{SNR}\to\infty} \frac{f(\mathsf{SNR})}{\log \mathsf{SNR}}, \tag{7}$$

which is the asymptotic slope, in bits/s/Hz/(3 dB), of the rate-SNR curve.

The DMT (diversity-multiplexing tradeoff) specifies the (r, d) pairs that are achievable for $\mathsf{SNR} \to \infty$, and thus characterizes the tradeoff between r and d [41]. For a quasi-static channel model where each coded block is subject to a single realization of the fading process, the DMT specifies that $\min(n_{\mathrm{T}}, n_{\mathrm{R}}) + 1$ distinct DMT points are feasible, each corresponding to a multiplexing gain $0 \leq r \leq \min(n_{\mathrm{T}}, n_{\mathrm{R}})$ and a diversity order

$$d(r) = (n_{\mathrm{T}} - r)(n_{\mathrm{R}} - r). \tag{8}$$

The full DMT frontier of achievable (r, d) pairs is obtained by connecting these points with straight lines.[6] More precisely, this is the optimum DMT frontier, corresponding to the mutual information in (3). Each transmit-receive architecture is associated with a DMT frontier, associated with the outage-rate relationship for that architecture, which may be smaller than this optimum.

The DMT governs the speed at which outage decreases with SNR: if the rate is increased as $r \log \mathsf{SNR}$, then the outage decreases (ignoring sub-polynomial terms) as $\mathsf{SNR}^{-d(r)}$. The DMT is thus a coarse description, through the proxies d and r, of the fundamental tradeoff between outage and rate. This coarseness arises from the definitions of r and d, which:

1) are asymptotic, thereby restricting the validity of the insights to the high-power regime,[7] and
2) involve only the slopes of the outage-SNR and rate-SNR curves, thereby ignoring constant offsets.[8]

Indeed, d does not suffice to determine the outage probability at a given SNR but simply quantifies the speed at which the outage falls with SNR. Similarly, r does not suffice to determine the rate, but it only quantifies how it grows with SNR.

Notice that $d(0)$ corresponds to the traditional notion of diversity order, i.e., with a fixed rate. A multiplexing gain $r = 0$ signifies a rate that does not increase (polynomially) with the SNR while $d = 0$ indicates an outage probability that does not decrease (polynomially) with the SNR.

[6]If the coded block spans several fading realizations, then this additional time/frequency selectivity leads to larger diversity orders but does not increase the maximum value of r [41], [42].

[7]Non-asymptotic DMT formulations have been put forth but they lack the simplicity and generality of (8) [43], [44].

[8]The choice of specific signal combining schemes at the receiver or of specific signal covariances at the transmitter is inconsequential to the DMT [45]–[47]

It is important to bear in mind that r need not coincide with the number of information symbols per MIMO symbol, which is the intuitive notion of spatial multiplexing. Likewise, d is not solely determined by the channel selectivity. Rather, both r and d reflect how aggressively link adaptation is performed; r measures how quickly the rate increases with SNR while d measures how rapidly the outage decreases with SNR.

C. Transmit Diversity v. Spatial Multiplexing in Modern Systems

In this high-velocity scenario, frequency-flat analyses are likely to indicate that dramatic reductions in outage probability can be had by increasing d. On these grounds, transmission strategies that operate efficiently at the full-diversity DMT point have been developed. The value of these strategies for modern wireless systems, however, is questionable because:

1) The outage need not be reduced below the target error probability.
2) Diversity is plentiful already:
   i) By the same token that the fading is too rapid to be tracked, it offers time selectivity.
   ii) Since, in this regime, modern systems disperse the signals over large swaths of bandwidth, there tends to be abundant frequency selectivity.

The DMT describes (coarsely) the entire outage-rate frontier, but modern wireless systems operate at a target outage probability ϵ and the quantity of interest is $R_\epsilon(\mathsf{SNR})$. Within the DMT framework, a fixed outage corresponds to the $d = 0$ point. As a result, all that one can infer about $R_\epsilon(\mathsf{SNR})$ on the basis of the DMT is that its asymptotic slope

$$\lim_{\mathsf{SNR}\to\infty} \frac{R_\epsilon(\mathsf{SNR})}{\log \mathsf{SNR}} \tag{9}$$

equals the maximum multiplexing gain (i.e., the value of r when $d = 0$) for the transmit-receive strategy being used. This quantity can, at most, equal $\min(n_{\mathrm{T}}, n_{\mathrm{R}})$ and, as recognized in [25], any strategy that does not attain $r = \min(n_{\mathrm{T}}, n_{\mathrm{R}})$ is strictly suboptimal in terms of $R_\epsilon(\mathsf{SNR})$ for $\mathsf{SNR} \to \infty$. Hence, we conclude that the optimum strategies are those that utilize all spatial degrees of freedom for multiplexing. Although this holds asymptotically in the SNR, the extent to which it holds for SNR values of interest in a selective channel can only be determined through a more detailed (non-asymptotic) study. To shed light on this point, a case study is presented next.

D. Case Study: A Contemporary MIMO-OFDM System

Let us consider the exemplary system described in Table I, which is loosely based on the 3GPP LTE design [23]. (With only slight modifications, this system could be made to conform with 3GPP2 UMB or with IEEE 802.16 WiMAX.) Every feature relevant to the discussion at hand is modeled:

- A basic resource block spans 12 OFDM tones over 1 ms. Since 1 ms corresponds to 14 OFDM symbols, a resource block consists of 168 symbols. In the high-velocity regime being considered, the 12 tones are interspersed uniformly over 10 MHz of bandwidth. There are 600 usable tones on that bandwidth, guards excluded, and hence every 50th tone is allocated to the user at hand while the rest are available for other users.[9]
- Every coded block spans up to 6 H-ARQ transmission rounds, each corresponding to a basic resource block, with successive rounds spaced by 6 ms for a maximum temporal span of 31 ms. (This is an acceptable delay for most applications, including Voice-over-IP.) The H-ARQ process terminates as soon as decoding is possible. An error is declared if decoding is not possible after 6 rounds.
- The channel exhibits continuous Rayleigh fading with a Clarke-Jakes spectrum and a 180-Hz maximum Doppler frequency. (This could correspond, for example, to a speed of 100 Km/h at 2 GHz.) The power delay profile is given by the 12-ray TU (typical urban) channel detailed in Table II. The r.m.s. delay spread equals 1 μs.
- The antennas are uncorrelated to underscore the roles of both diversity and multiplexing. Some comments on antenna correlation are put forth in Section IV.E.

The impulse response describing each of the $n_{\rm T} n_{\rm R}$ entries of the channel matrix is

$$h(t,\tau) = \sum_{j=1}^{12} \sqrt{\alpha_j} c_j(t) \delta(t-\tau_j) \tag{10}$$

[9]For low velocity users, in contrast, the 12 tones in a resource block are contiguous so that their fading can be efficiently described and fed back for link adaptation and scheduling purposes as discussed in Section II.

where the delays $\{\tau_j\}_{j=1}^{12}$ and the powers $\{\alpha_j\}_{j=1}^{12}$ are specified in Table II and $\{c_j(t)\}_{j=1}^{12}$ are independent complex Gaussian processes with a Clarke-Jakes spectrum. Although time-varying, the channel is suitably constant for the duration of an OFDM symbol such that it is meaningful to consider its frequency response as in (1).

The variability of the channel response over the multiple tones and H-ARQ rounds of a coded block is illustrated in Fig. 1. Note the very high degree of frequency selectivity and how the channel decorrelates during the 6 ms separating H-ARQ rounds.

Without H-ARQ, rate and outage are related as per (4). With H-ARQ, on the other hand, the length of each coded block becomes variable. With IR (incremental redundancy) specifically, mutual information is accumulated over successive H-ARQ rounds [48]. If we let $\mathcal{M}_k(\mathsf{SNR})$ denote the mutual information after k rounds, then the number of rounds needed to decode a particular block is the smallest integer K such that

$$\mathcal{M}_K(\mathsf{SNR}) > 6\, R_\epsilon(\mathsf{SNR}) \tag{11}$$

where $K \leq 6$. A one-bit notification of success/failure is fed back after the receiver attempts to decode following each H-ARQ round. An outage is declared if $\mathcal{M}_6(\mathsf{SNR}) \leq 6\, R_\epsilon(\mathsf{SNR})$ and the effective rate (long-term average transmitted rate) is

$$\mathcal{R}_\epsilon(\mathsf{SNR}) = \frac{6\, R_\epsilon(\mathsf{SNR})}{E[K]}. \tag{12}$$

The initial rate is selected such that the outage at H-ARQ termination is precisely ϵ. This corresponds to choosing an initial rate of $6\, R_\epsilon$ where R_ϵ corresponds to the quantity of interest defined in (5) with the mutual information averaged over the 168 symbols within each H-ARQ round and then summed across the 6 rounds.

In order to contrast the benefits of transmit diversity and spatial multiplexing, we shall evaluate two representative transmission techniques that achieve the respective extremes of the optimal DMT frontier:

- A transmit diversity strategy that converts the MIMO channel into an effective scalar channel with signal-to-noise ratio

$$\frac{\mathsf{SNR}}{n_{\mathrm{T}}} \,\mathrm{Tr}\left\{\mathbf{H}_i(k)\mathbf{H}_i^\dagger(k)\right\} \tag{13}$$

  where $\mathbf{H}_i(k)$ denotes the channel for the ith symbol on the kth H-ARQ round. By applying a strong outer code to this effective scalar channel, the mutual information after k rounds is, at most [41]

$$\mathcal{M}_k(\mathsf{SNR}) = \sum_{\ell=1}^{k} \frac{1}{168} \sum_{i=1}^{168} \log\left(1 + \frac{\mathsf{SNR}}{n_{\mathrm{T}}} \operatorname{Tr}\left\{\mathbf{H}_i(\ell)\mathbf{H}_i^{\dagger}(\ell)\right\}\right). \tag{14}$$

This transmit diversity strategy provides full diversity order with reduced complexity, but its multiplexing gain cannot exceed $r = 1$, i.e., one information symbol for every vector $\mathbf{x}_i$ in (1). Note that, when $n_{\mathrm{T}} = 2$, (14) is achieved by Alamouti transmission [17].

- A basic MMSE-SIC spatial multiplexing strategy where a separate coded signal is transmitted from each antenna, all of them at the same rate [49]. The receiver attempts to decode the signal transmitted from the first antenna. An MMSE filter is applied to whiten the interference from the other signals, which means that the first signal experiences a signal-to-noise ratio

$$\mathbf{h}_{i,1}^{\dagger}(k)\left(\mathbf{H}_{i,1}(k)\mathbf{H}_{i,1}^{\dagger}(k) + \frac{n_{\mathrm{T}}}{\mathsf{SNR}}\mathbf{I}\right)^{-1}\mathbf{h}_{i,1}(k) \tag{15}$$

during the kth H-ARQ round. If successful, the effect of the first signal is subtracted from the received samples and decoding of the second signal is attempted, and so forth. No optimistic assumption regarding error propagation is made: an outage is declared if any of the n_{T} coded signals cannot support the transmitted rate. The aggregate mutual information over the n_{T} antennas after k H-ARQ rounds is

$$\mathcal{M}_k(\mathsf{SNR}) = n_{\mathrm{T}} \min_{m=1,\cdots,n_{\mathrm{T}}} \left\{\sum_{\ell=1}^{k} \frac{1}{168} \sum_{i=1}^{168} \log\left(1 + \mathbf{h}_{i,m}^{\dagger}(\ell)\left(\mathbf{H}_{i,m}(\ell)\mathbf{H}_{i,m}^{\dagger}(\ell) + \frac{n_{\mathrm{T}}}{\mathsf{SNR}}\mathbf{I}\right)^{-1}\mathbf{h}_{i,m}(\ell)\right)\right\} \tag{16}$$

where $\mathbf{h}_{i,m}(\ell)$ is the mth column of $\mathbf{H}_i(\ell)$ and $\mathbf{H}_{i,m}(\ell) = [\mathbf{h}_{i,m+1}(\ell)\mathbf{h}_{i,m+2}(\ell)\cdots\mathbf{h}_{i,n_{\mathrm{T}}}(\ell)]$. While deficient in terms of diversity order, this strategy yields full multiplexing gain, $r = \min(n_{\mathrm{T}}, n_{\mathrm{R}})$, when $d = 0$. This MMSE-SIC structure is representative of the single-user MIMO mode in LTE [23].

Let $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$, the high-end configuration for LTE, and consider first a simplistic model where the fading is frequency-flat and there is no H-ARQ. Every coded block is therefore subject to essentially a single realization of the Rayleigh fading process. Under such model, the spectral efficiencies achievable with 1% outage, $R_{0.01}(\mathsf{SNR})$, are compared in Fig. 2 alongside the corresponding efficiency for the non-MIMO reference ($n_{\mathrm{T}} = 1$,

$n_{\mathrm{R}} = 4$). Transmit diversity is uniformly superior to spatial multiplexing in the SNR range of interest. In fact, spatial multiplexing results in a loss even with respect to non-MIMO transmission with the same number of receive antennas. The curves eventually cross, as the DMT predicts[10] and the inset in Fig. 2 confirms, but this crossover does not occur until beyond 30 dB.

Still with $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$, consider now the richer model described in Tables I–II. The effective mutual information for each block is averaged over tones and symbols and accumulated over H-ARQ rounds. The corresponding comparison is presented in Fig. 3. In this case, transmit diversity offers a negligible advantage whereas spatial multiplexing provides ample gains with respect to non-MIMO.

The stark contrast between the behaviors observed under the different models can only be explained by the abundant time and frequency selectivity neglected by the simple model and actually present in the system. This renders transmit antenna diversity superfluous, not only asymptotically but at every SNR. Under the simple model, the signal from the first antenna does not benefit from any diversity and thus limits the overall rate. Under the richer channel model, however, that first signal reaps diversity from time/frequency selectivity and thus the lack of spatial diversity is mostly inconsequential. To further highlight this effect, consider the outage ϵ-$\mathcal{R}_\epsilon$ curves in Fig. 4 for the simple and rich channel models. (Recall that, in the simple model, $\mathcal{R}_\epsilon = R_\epsilon$ because there is no H-ARQ). The enrichment of the channel greatly increases the MMSE-SIC rate, but has a much smaller effect on the transmit diversity rate due to the diminishing returns of diversity.

Although the above results were for a highly selective channel, this behavioral contrast is robust. Even if the speed is reduced down to where the low-velocity regime might start, as in Fig. 5, the behaviors are hardly affected because there is still significant selectivity. Likewise, the performances are largely preserved if the bandwidth is diminished

[10]For transmit diversity, the DMT frontier is a straight line connecting $(r = 0, d = 16)$ and $(r = 1, d = 0)$, while for MMSE-SIC it connects $(r = 0, d = 1)$ and $(r = 4, d = 0)$. These frontiers intersect at $(r = 20/21, d = 16/21)$ and thus transmit diversity outperforms MMSE-SIC (as SNR $\to \infty$) for $d > 16/21$. In our $d = 0$ setting, transmit diversity only achieves a slope of $r = 1$ bit/s/Hz/(3 dB) compared to $r = 4$ for MMSE-SIC. This difference in slope explains the eventual crossover.

significantly below 10 MHz or the delay spread is reduced below 1 μs.

E. Ergodic Modeling

As it turns out, the time/frequency selectivity in modern systems is so substantial as to justify the adoption of an ergodic model altogether. Shown in Fig. 6 is the correspondence between the exact rates achievable with 1% outage in the channel described in Tables I–II and the respective ergodic rates.

From a computational standpoint, this match is welcome news because of the fact that convenient closed forms exist for the rates achievable in an ergodic Rayleigh-faded channel [50]. Moreover, the optimum transmission strategies and the impact upon capacity of more detailed channel features such as antenna correlation, Rice factors, colored out-of-cell interference, etc, can then be asserted by virtue of the extensive body of results available for the ergodic setting [45], [51].

Antenna correlation, for example, leads to a disparity in the distribution of the spatial eigenmodes that effectively reduces the spatial multiplexing capability. Such effects should, of course, be taken under consideration when determining the appropriate transmission strategy.

F. Optimal Spatial Multiplexing

While in the case study we considered the performance of a low complexity but suboptimal detection scheme for spatial multiplexing, the continual increase in computational power is now rendering optimal or near-optimal spatial multiplexing feasible. Rather than transmitting separate coded signals from the n_{T} antennas, a single one can be interleaved over time, frequency and the transmit antennas. At the receiver side, each vector symbol is then fed to a detector that derives soft estimates of each coded bit—possibly by use of a sphere decoder—to a standard outer decoder (e.g., message-passing decoder), with subsequent iterations between the MIMO detector and the decoder [52]. Such techniques, and others such as mutual information lossless codes [53]–[55], can approach the mutual information in (3).

In the context of our comparison between transmit diversity and spatial multiplexing, it is worthwhile to note that the mutual information in (3) is greater than that in (13) for *any* channel matrix $\mathbf{H}$. Denoting by λ_ℓ the ℓth eigenvalue of $\mathbf{H}\mathbf{H}^\dagger$,

$$\log\det\left(\mathbf{I}+\frac{\mathsf{SNR}}{n_{\mathrm{T}}}\mathbf{H}\mathbf{H}^\dagger\right) = \log\left(\prod_{\ell=1}^{n_{\mathrm{R}}}\left(1+\frac{\mathsf{SNR}}{n_{\mathrm{T}}}\lambda_\ell\right)\right) \tag{17}$$

$$\geq \log\left(1+\frac{\mathsf{SNR}}{n_{\mathrm{T}}}\sum_{\ell=1}^{n_{\mathrm{R}}}\lambda_\ell\right) \tag{18}$$

$$= \log\left(1+\frac{\mathsf{SNR}}{n_{\mathrm{T}}}\mathrm{Tr}\left\{\mathbf{H}\mathbf{H}^\dagger\right\}\right) \tag{19}$$

where (17) holds because the determinant equals the product of the eigenvalues, (18) comes from dropping terms in the product, and (19) follows from $\mathrm{Tr}\left\{\mathbf{H}\mathbf{H}^\dagger\right\} = \sum_{\ell=1}^{n_{\mathrm{R}}}\lambda_\ell$.

Hence, spatial multiplexing with optimal detection is uniformly superior to rate-sacrificing transmit diversity in the sense that it achieves a smaller outage probability for *any* rate and SNR.[11] Drawing parallels with the discussion in Section I about the sub-optimality of repeating the same signal on two frequency channels versus transmitting different portions of a coded block thereon, one could equate transmit diversity with the former and the optimum MIMO strategy with the latter.

Optimum spatial multiplexing allows achieving both the rate and the outage benefits of multiantenna communication, whereas transmit diversity and MMSE-SIC spatial multiplexing obtain only the outage or only the rate benefit, respectively. Nonetheless, a system employing optimum spatial multiplexing is still subject to the fundamental outage-rate tradeoff governed by the optimal DMT region specified in (8).

In Fig. 7, the spectral efficiencies of Alamouti transmission and spatial multiplexing (for optimal and MMSE-SIC) are shown for $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$, for both the frequency-flat model and the richer model in Tables I–II. Optimal spatial multiplexing is superior to Alamouti with both models, as per the above derivation, but the difference is considerably larger when the rich model is used. Consistent with the earlier case study, MMSE-SIC performs well below Alamouti in the frequency-flat setting but outperforms it under the rich model. Spatial multiplexing with MMSE-SIC and optimal spatial multiplexing

[11] The mutual information in (3) is also greater than the MMSE-SIC mutual information in (16), because (3) equals the sum of the n_{T} mutual informations whereas (16) is n_{T} times the minimum. One can make the two equal by separately rate controlling each coded signal based on instantaneous channel conditions [56], but this is infeasible in the high-velocity regime.

have the same maximum multiplexing gain and thus the same slope, $r = 2$ bits/s/Hz/(3 dB), but the DMT cannot capture the constant offset between the two.

V. Uncoded Error Probability: A Potentially Misleading Metric

In the previous section, the superiority of spatial multiplexing relative to rate-sacrificing transmit diversity was illustrated in the context of modern wireless systems, which exhibit abundant time and frequency selectivity and utilize powerful outer coding. However, a different conclusion is sometimes reached if one compares the error probabilities of the two schemes in the absence of outer coding.

Consider $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$ for the sake of specificity. Comparisons must be conducted at equal SNR and rate, e.g., Alamouti with 16-QAM v. spatial multiplexing with 4-QAM or Alamouti with 256-QAM v. spatial multiplexing with 16-QAM. These constitute different space-time modulation formats with 4 and 8 bits per MIMO symbol, respectively. Fig. 8 presents the symbol error probabilities, averaged over the fading distribution, for a maximum-likelihood detector with no outer coding. (Note that, in the absence of outer coding, changing the modulation format is the only way to coarsely perform link adaptation). Consistent with previous sections, one can compare the different schemes at a particular operating point, which in this case corresponds to a fixed symbol error probability. From the two sets of curves in the figure, spatial multiplexing outperforms Alamouti if the error probability is above 10^{-2}. For lower error probabilities, which is where an uncoded system would likely have to operate, Alamouti outperforms spatial multiplexing at 4 bits/symbol but not necessarily at 8 bits/symbol. In fact, the behavior of the uncoded error probability as a function of SNR is qualitatively similar to that of the mutual information outages in a non-selective channel: from Fig. 4, at outage probabilities above roughly 0.2 spatial multiplexing outperforms transmit diversity (in flat fading with no H-ARQ) whereas for lower outage probabilities the opposite is true.[12]

[12] The uncoded symbol error probability is the average of the uncoded symbol error probability conditioned on $\mathbf{H}$. This conditional error probability, which is determined only by the noise, is essentially 0 or 1 for most channel realizations depending on whether the mutual information corresponding to $\mathbf{H}$ is larger or smaller, respectively, than the transmit rate. Intuitively then, the average symbol error probability roughly mirrors the mutual information outage probability in non-selective channels.

Hence, uncoded analysis does not always correctly predict the behavior with strong outer coding in much the same way that a mutual information analysis without selectivity does not (cf. Section IV). In a system such as the one described in Tables I and II, the outer code makes use of frequency selectivity across tones and time selectivity across H-ARQ rounds. Without an outer code, on the other hand, this selectivity would not be exploited and thus averaging uncoded error probabilities does not have the same operational meaning of averaging mutual informations. Since modern communication systems rely on powerful channel codes, it is the coded performance that is relevant. One should be careful when predicting such performance on the basis of uncoded error probabilities.

VI. Conclusion

Since the 1970's, antenna diversity had been a preferred weapon used by mobile wireless systems against the deleterious effect of fading. While narrowband channelizations and non-adaptive links were the norm, antenna diversity was highly effective. In modern systems, however, this is no longer the case. Link adaptivity and scheduling have rendered transmit diversity undesirable for low-velocity users whereas abundant time/frequency selectivity has rendered transmit diversity superfluous for high-velocity users. Moreover, the prevalence of MIMO has opened the door for a much more effective use of antennas: spatial multiplexing. Indeed, the spatial degrees of freedom created by MIMO should be regarded as additional 'bandwidth' and, for the same reason that schemes based on time/frequency repetition are not used for they waste bandwidth, rate-sacrificing transmit diversity techniques (e.g., OSTBC) waste 'bandwidth'.

Of all possible DMT points, therefore, the zero-diversity one stands out in importance. Techniques, even suboptimum ones, that can provide full multiplexing are most appealing to modern wireless systems whereas techniques that achieve full diversity order but fall short on multiplexing gain are least appealing. Our findings further the conclusion in [25], where a similar point is made solely on the basis of the multiplexing gain for frequency-flat channels. Although our conclusion has been reached on the premise that the coded error probabilities of discrete constellations are well approximated by the mutual information outages of Gaussian codebooks, we expect it to hold in any situation where the code operates at a (roughly) constant gap to the mutual information.

The trend for the foreseeable future is a sustained increase in system bandwidth, which is bound to only shore up the above conclusion. LTE, which for our case study was taken to use 10 MHz, is already moving towards 20 MHz channelizations.

At the same time, exceptions to the foregoing conclusion do exist. These include, for example, control channels that convey short messages. Transmit diversity is fitting for these channels, which do benefit from a lower error probability but lack significant time/frequency selectivity. Other exceptions may be found in applications such as sensor networks or others where the medium access control is non-existent or does not have link adaptation and retransmission mechanisms.

Our study has only required evaluating well-known techniques under realistic models and at the appropriate operating points. Indeed, a more general conclusion that can be drawn from the discussion in this paper is that, over time, the evolution of wireless systems has rendered some of the traditional models and wisdoms obsolete. In particular:

- Frequency and time selectivity should always be properly modeled.
- Performance assessments are to be made at the correct operating point, particularly in terms of error probability.
- The assumptions regarding transmit CSI must be consistent with the regime being considered. At low velocities, adaptive rate control based on instantaneous CSI should be incorporated; at high velocities, only adaptation to average channel conditions should be allowed.
- Coded block error probabilities or mutual information outages, rather than uncoded error probabilities, should be used to gauge performance.

Proper modeling is essential in order to evaluate the behavior of transmission and reception techniques in contemporary and future wireless systems. As our discussion on transmit diversity and spatial multiplexing demonstrates, improper modeling can lead to misguided perceptions and fictitious gains.

VII. Acknowledgments

The authors gratefully acknowledge comments, suggestions and discussions with Profs. Jeff Andrews and Robert Heath Jr., from the University of Texas at Austin, Prof. Uri Erez, from Tel-Aviv University, Prof. Andrea Goldsmith, from Stanford University, Prof. Albert Guillen, from the University of Cambridge, Prof. Daniel P. Palomar, from the Hong Kong University of Science and Technology, and Mr. Chris Ramming from Intel Corp., as well as the comments of the anonymous reviewers.

References

[1] W. C Bohn, "Automatic selection of receiving channels," *U. S. Patent No. 1747218*, Feb. 1930.

[2] H. H. Beverage and H. O. Peterson, "Diversity receiving system of RCA for radiotelegraphy," *Proc. of IRE*, vol. 19, pp. 531–584, Apr. 1931.

[3] F. J. Altman and W. Sichak, "A simplified diversity communication system for beyond-the-horizon links," *IRE Trans. Comm. Systems*, vol. 4, pp. 50–55, Mar. 1956.

[4] P. Bello and B. Nelin, "Predetection diversity combining with selectively fading channels," *IRE Trans. Comm. Systems*, vol. 9, pp. 32–42, Mar. 1963.

[5] H. Makino and K. Morita, "Design of space diversity receiving and transmitting systems for line-of-sight microwave links," *IRE Trans. Comm. Tech.*, vol. 15, pp. 603–614, Aug. 1967.

[6] L. R. Kahn, "Ratio squarer," *Proc. IRE*, vol. 42, pp. 1704, Nov. 1954.

[7] W. T. Barnett, "Microwave line-of-sight propagation with and without frequency diversity," *Bell Syst. Tech. Journal*, vol. 49, no. 8, pp. 1827–1871, Oct. 1970.

[8] D. G. Brennan, "Linear diversity combining techniques," *Proc. IRE*, vol. 47, pp. 1075–1102, 1959.

[9] M. Schwartz, W. R. Bennett, and S. Stein, *Communication Systems and Techniques*, McGraw-Hill, New York, 1965.

[10] A. Vigants, "Space diversity performance as a function of antenna separation," *IEEE Trans. Comm. Tech.*, vol. 16, no. 6, pp. 831–836, Dec. 1968.

[11] William C. Jakes, *Microwave Mobile Communications*, New York, IEEE Press, 1974.

[12] D. C. Cox, "Antenna diversity performance in mitigating the effects of portable radiotelephone orientation and multipath propagation," *IEEE Trans. Commun.*, vol. 31, pp. 620–628, May 1983.

[13] N. Nakajima K. Kinoshita, M. Kuramoto, "Development of a TDMA digital cellular system based on a Japanese standard," *Proc. of IEEE Vehic. Tech. Conf. (VTC'91)*, pp. 642–645, 1991.

[14] A. J. Goldsmith, *Wireless Communications*, Cambridge, 2005.

[15] A. Hiroike and F. Adachi, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Veh. Technol.*, vol. 33, pp. 37–43, Feb. 1984.

[16] A. Wittneben, "A new bandwidth efficient transmit antenna modulation diversity scheme for linear digital modulation," *Proc. IEEE Int'l Conf. on Commun. (ICC'93)*, vol. 3, pp. 1630–1634, 1993.

[17] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.

[18] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.

[19] G. J. Foschini and M. J. Gans, "On the limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, pp. 315–335, 1998.

[20] G. J. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs Tech. J.*, pp. 41–59, 1996.

[21] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecom*, vol. 10, pp. 585–595, Nov. 1999.

[22] G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, Mar. 1998.

[23] S. Sesia, I. Toufik, and M. Baker (Editors), *The UMTS Long Term Evolution: From Theory to Practice*, Wiley, 2009.

[24] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX*, Prentice Hall, 2007.

[25] B. Varadarajan and J. R. Barry, "The outage capacity of linear space-time codes," *IEEE Trans. Wireless Communications*, vol. 6, no. 6, pp. 2642–2648, Nov. 2005.

[26] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge University Press, 2005.

[27] R. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 8, no. 1, pp. 21–28, Jan. 1962.

[28] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo-codes," *Proc. of IEEE Int'l Conf. in Communic. (ICC'93)*, pp. 1064–1070, May 1993.

[29] G. D. Forney and D. J. Costello, "Channel coding: The road to channel capacity," *Proceedings of IEEE*, vol. 95, no. 6, pp. 1150–1177, June 2007.

[30] A. J. Goldsmith, "The capacity of downlink fading channels with variable rate and power," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 569–580, Aug. 1997.

[31] D. Costello, J. Hagenauer, H. Imai, and S. B. Wicker, "Applications of error-control coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2531–2560, Oct. 1998.

[32] M. Meyer, H. Wiemann, M. Sagfors, J. Torsner, and J-F. Cheng, "ARQ concept for the UMTS Long-Term Evolution," *Proc. of IEEE Vehic. Technol. Conf. (VTC'2006 Fall)*, Sept. 2006.

[33] P. Wu and N. Jindal, "Coding versus ARQ in fading channels: How reliable should the PHY be?," Apr. 2009, arXiv:0904.0226.

[34] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas in Comm.*, vol. 26, pp. 1341–1365, Oct. 2008.

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

[36] H. Viswanathan and S. Venkatesan, "The impact of antenna diversity in packet data systems with scheduling," *IEEE Trans. on Communications*, vol. 52, pp. 546–549, Apr. 2004.

[37] A. Lozano, A. M. Tulino, and S. Verdú, "Power allocation over parallel Gaussian channels with arbitrary input distributions," *IEEE Trans. Inform. Theory*, vol. 52, no. 7, pp. 3033–3051, July 2006.

[38] L. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, pp. 359–378, May 1994.

[39] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1468–1489, July 1999.

[40] N. Prasad and M. K. Varanasi, "Outage theorems for MIMO block-fading channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5284–5296, Dec. 2006.

[41] L. Zheng and D. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[42] P. Coronel and H . Bolcskei, "Diversity-multiplexing tradeoff in selective-fading MIMO channels," *Proc. of IEEE Int'l Symp. on Inform. Theory (ISIT'07)*, June 2007.

[43] R. Narasihman, "Finite-SNR diversity-multiplexing tradeoff for correlated Rayleigh and Rician MIMO channels," *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 3956–3979, Sept. 2006.

[44] K. Azarian and H. El Gamal, "The throughput-reliability tradeoff in block-fading MIMO channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 488–501, Feb. 2007.

[45] A. Lozano, A. M. Tulino, and S. Verdu, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

[46] N. Jindal, "High SNR analysis of MIMO broadcast channels," *Proc. of IEEE Int'l Symp. on Inform. Theory (ISIT'05)*, Sept. 2005.

[47] N. Prasad and M. K. Varanasi, "MIMO outage capacity in the high SNR regime," *Proc. Int'l Symp. on Inform. Theory (ISIT 2005)*, pp. 656–660, Sept. 2005.

[48] G. Caire and D. Tuninetti, "The throughput of hybrid-ARQ protocols for the Gaussian collision channel," *IEEE Trans. Inform. Theory*, vol. 47, no. 5, pp. 1971–1988, Jul. 2001.

[49] G. J. Foschini, G. D. Golden, R. A. Valenzuela, and P. W. Wolnianski, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE J. Select. Areas Commun.*, vol. 17, no. 11, pp. 1841–1852, Nov. 1999.

[50] H. Shin and J. H. Lee, "Capacity of multiple-antenna fading channels: Spatial fading correlation, double scattering and keyhole," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2636–2647, Oct. 2003.

[51] A. M. Tulino, A. Lozano, and S. Verdu, "Impact of correlation on the capacity of multiantenna channels," *IEEE Trans. Inform. Theory*, vol. 51, pp. 2491–2509, July 2005.

[52] B. M. Hochwald and S. ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, Mar. 2003.

[53] B. A. Sethuraman, B. S. Rajan, and V. Shashidhar, "Full-diversity, high-rate space-time block codes from division algebras," *IEEE Trans. Inform. Theory*, vol. 49, pp. 2596–2616, Oct. 2003.

[54] F. Oggier, J.-C. Belfiore, and E. Viterbo, "Cyclic division algebras: A tool for space-time coding," *Foundations and Trends in Communications and Information Theory*, vol. 4, no. 1.

[55] V. Shashidhar, B. S. Rajan, and B. A. Sethuraman, "Information-lossless space-time block codes from crossed-product algebras," *IEEE Trans. Inform. Theory*, vol. 52, pp. 3913–3935, Sept. 2006.

[56] S. T. Chung, A. Lozano, H. C. Huang, A. Sutivong, and J. M. Cioffi, "Approaching the MIMO capacity with V-BLAST: theory and practice," *EURASIP Journal on Applied Signal Processing (special issue on MIMO)*, vol. 2002, pp. 762–771, May 2004.

TABLE I
MIMO-OFDM SYSTEM PARAMETERS

Tone spacing	15 kHz
OFDM Symbol duration	71.5 μs
Bandwidth	10 MHz (600 tones, excluding guards)
Resource block	12 tones over 1 ms (168 symbols)
H-ARQ	Incremental redundancy
H-ARQ round spacing	6 ms
Max. number H-ARQ rounds	6
Power delay profile	12-ray TU
Doppler spectrum	Clarke-Jakes
Max. Doppler frequency	185 Hz
Antenna correlation	None

TABLE II
TU POWER DELAY PROFILE

Delay (μs)	Power (dB)
0	-4
0.1	-3
0.3	0
0.5	-2.6
0.8	-3
1.1	-5
1.3	-7
1.7	-5
2.3	-6.5
3.1	-8.6
3.2	-11
5	-10

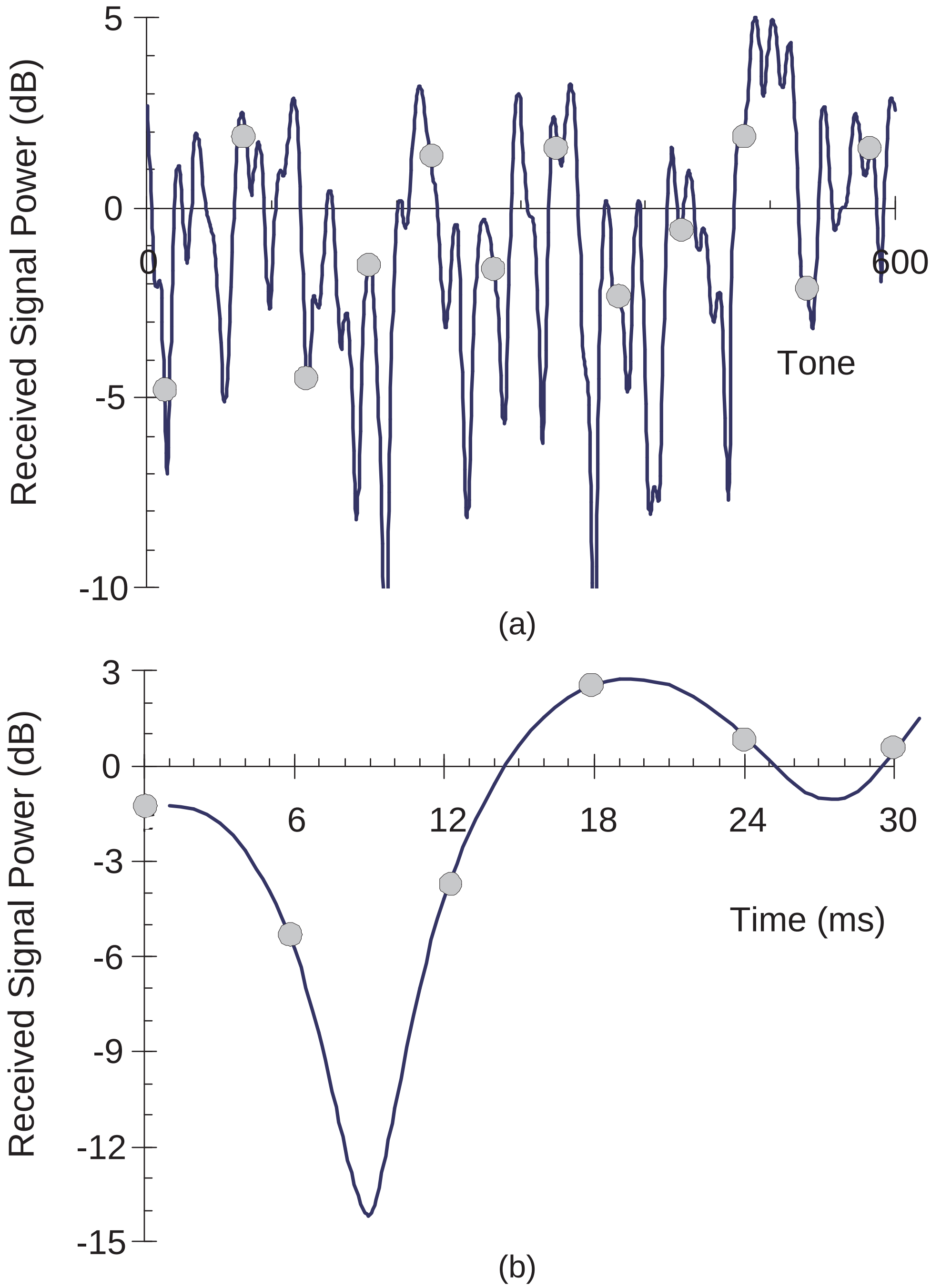

Fig. 1. (a) TU channel fading realization over 600 tones. The circles indicate the locations of the 12 tones that map to a given resource block. (b) TU channel fading realization, for a given tone, over 30 ms. The circles indicate the locations of the 6 H-ARQ rounds.

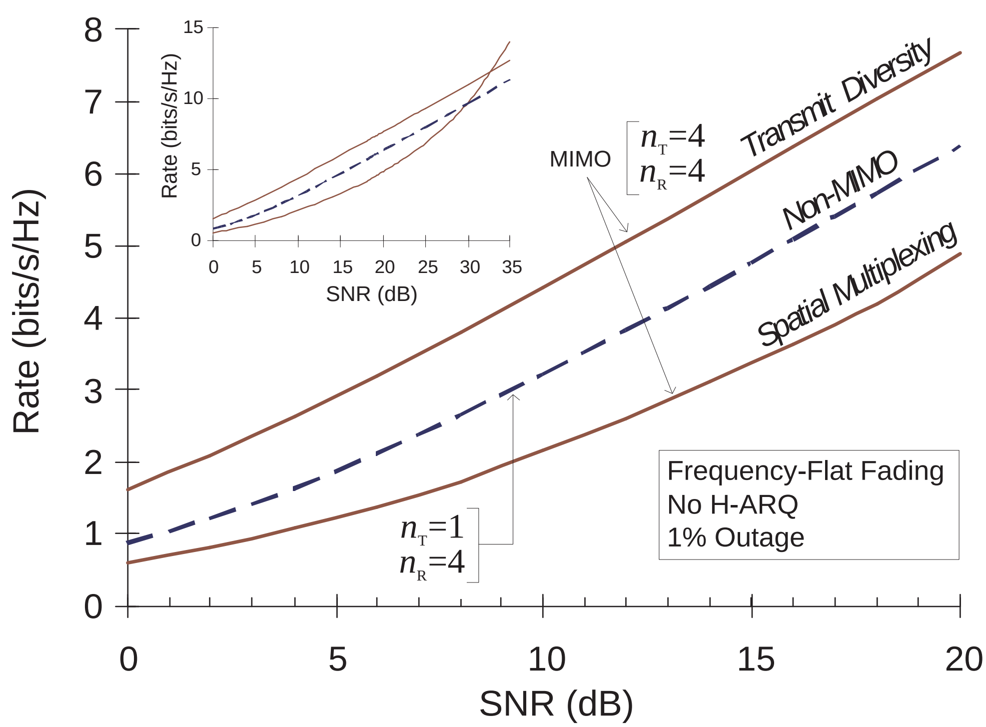

Fig. 2. Main plot: MMSE-SIC spatial multiplexing v. transmit diversity with $n_{\mathrm{T}} = n_{\mathrm{R}} = 4$ in a frequency-flat channel with no H-ARQ. Also shown is the non-MIMO reference ($n_{\mathrm{T}} = 1$, $n_{\mathrm{R}} = 4$). Inset: Same curves over a wider SNR range.

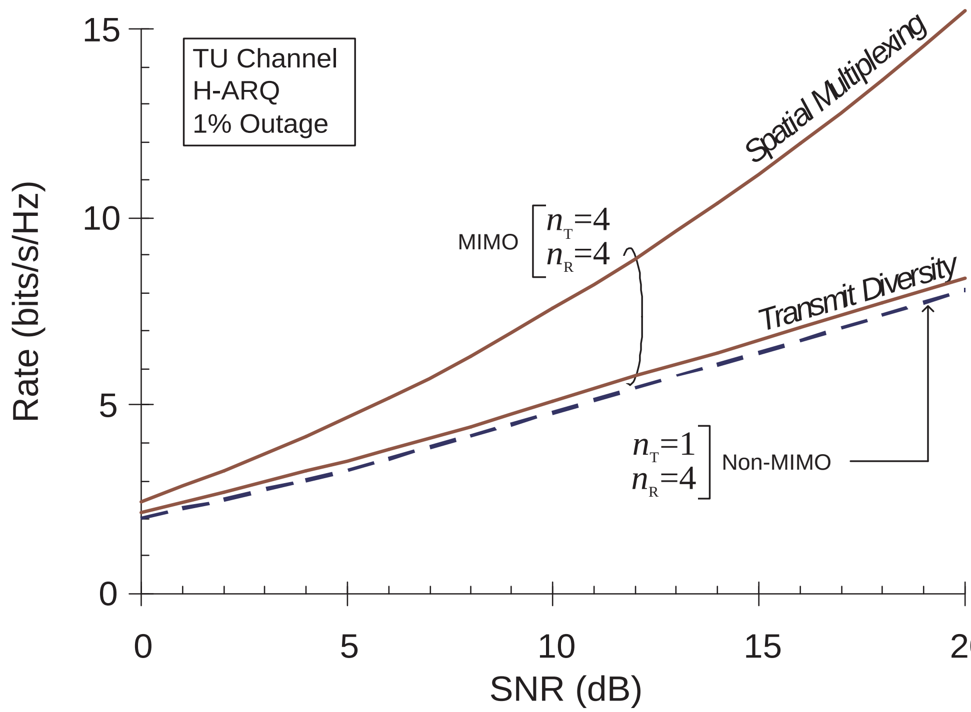

Fig. 3. MMSE-SIC spatial multiplexing v. transmit diversity with $n_T = n_R = 4$ in the channel described in Tables I–II. Also shown is the non-MIMO reference ($n_T = 1$, $n_R = 4$).

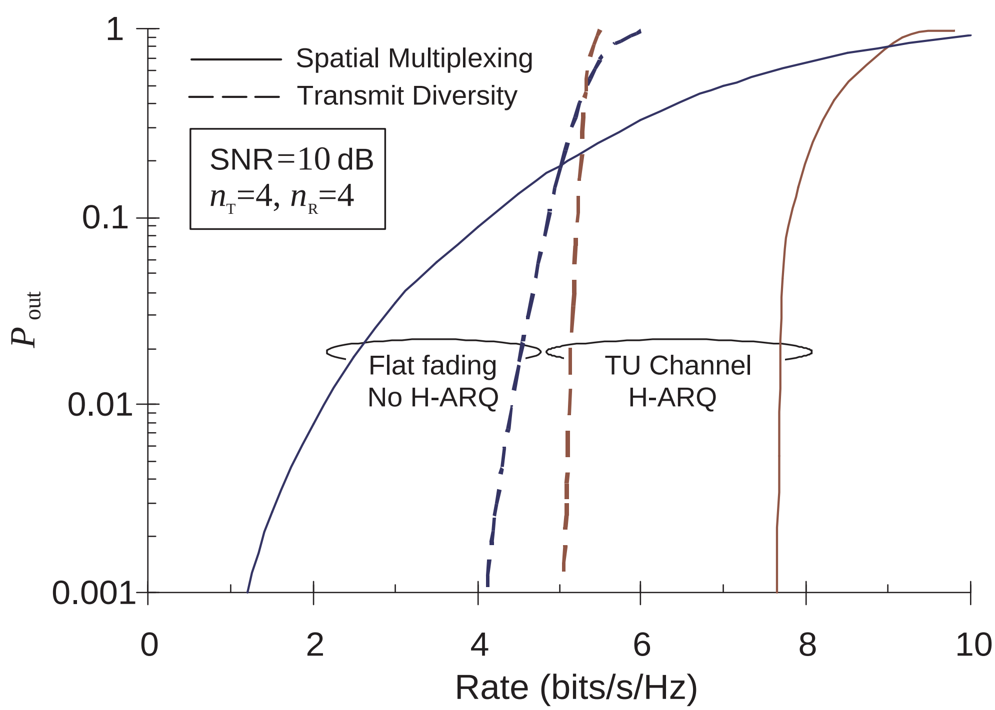

Fig. 4. Outage probability as function of $\mathcal{R}_\epsilon$ for transmit diversity and for MMSE-SIC spatial multiplexing at $\mathsf{SNR} = 10$ dB, for both the frequency-flat channel without H-ARQ and for the channel described in Tables I–II.

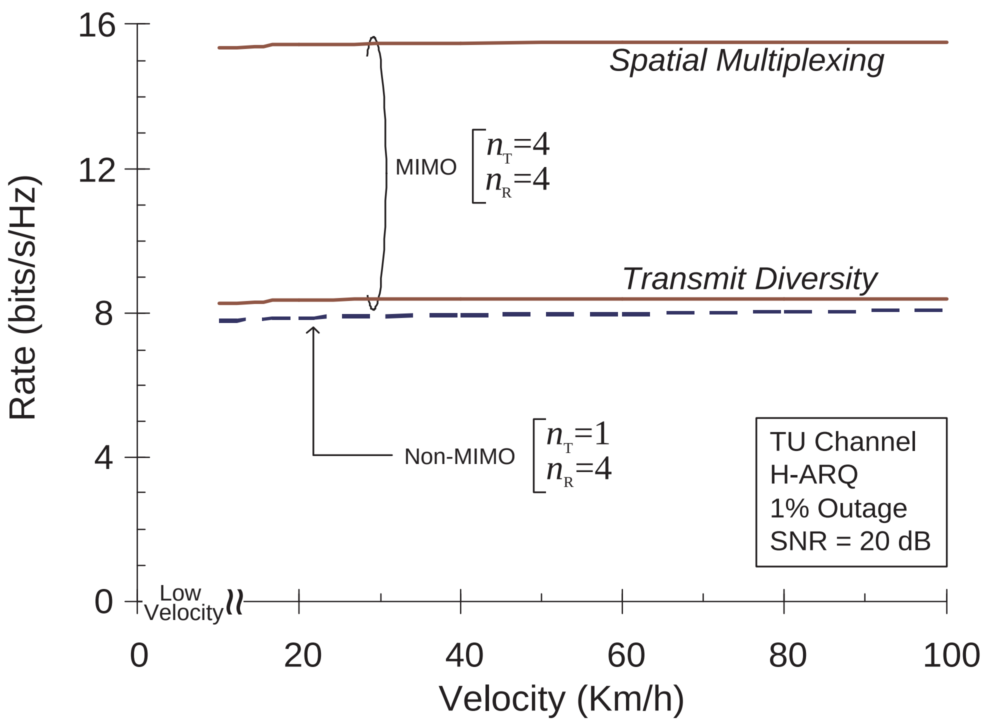

Fig. 5. MMSE-SIC spatial multiplexing, transmit diversity and non-MIMO transmission as function of velocity for the channel described in Tables I–II at SNR $= 20$ dB. (Below some point, the system transitions to the low-velocity regime and thus the curves are no longer meaningful.)

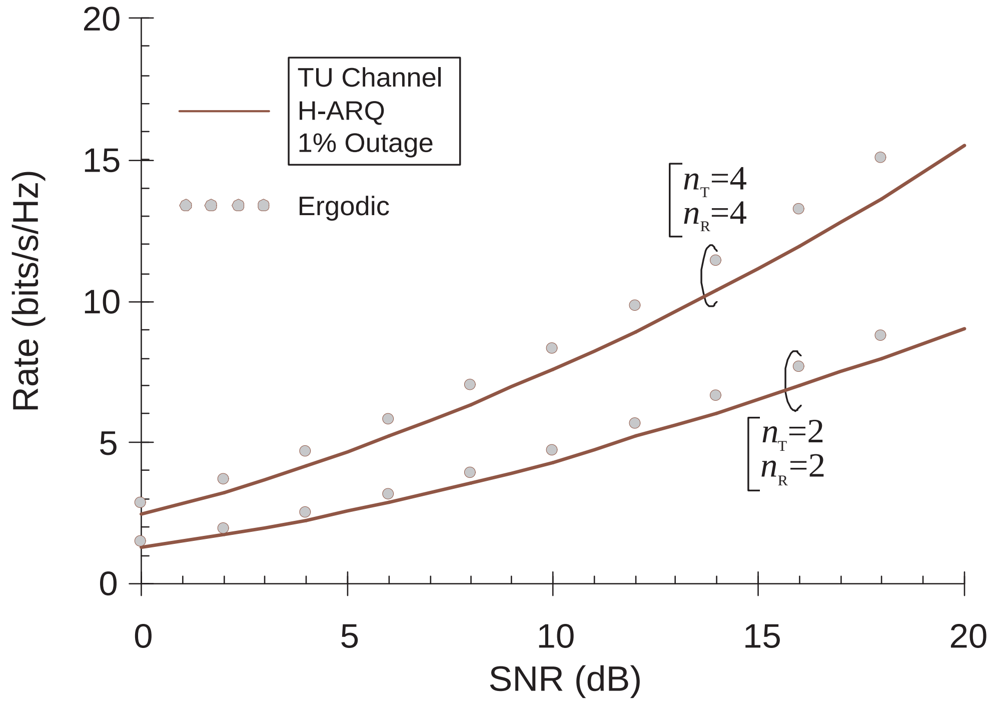

Fig. 6. In solid lines, 1%-outage rate achievable with MMSE-SIC spatial multiplexing in the channel described in Tables I–II. In circles, corresponding ergodic rate for the same numbers of antennas.

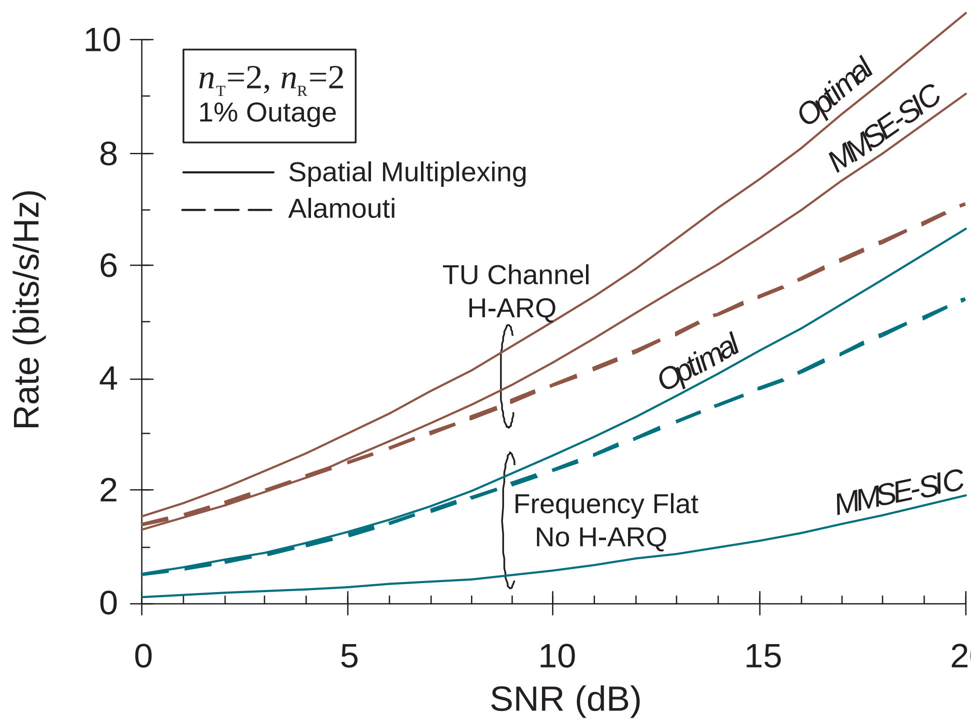

Fig. 7. Spectral efficiencies achievable with Alamouti transmission and with spatial multiplexing (optimal and MMSE-SIC) for $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$. The comparisons are shown for both a frequency-flat channel without H-ARQ and for the channel described in Tables I–II.

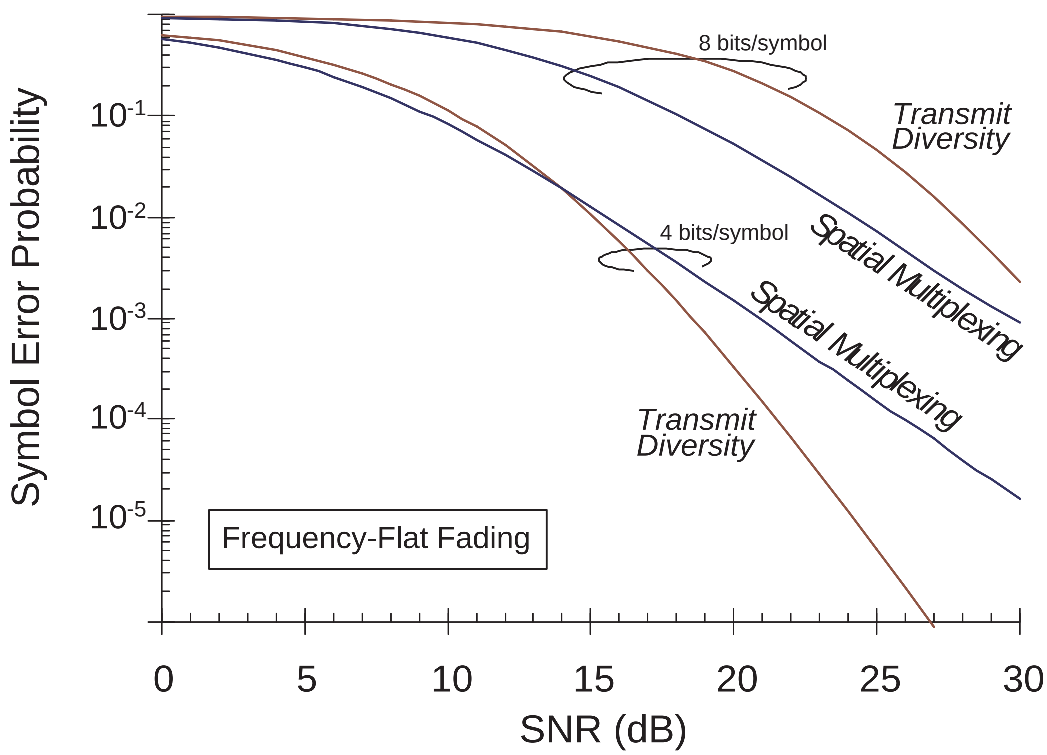

Fig. 8. Uncoded symbol error probability for transmit diversity and spatial multiplexing with $n_{\mathrm{T}} = n_{\mathrm{R}} = 2$. The comparisons are shown for both 4 bits/symbol and 8 bits/symbol.