Performance Limits of Coded Diversity Methods for Transmitter Antenna Arrays

Aradhana Narula, *Member, IEEE*, Mitchell D. Trott, *Member, IEEE*, and Gregory W. Wornell, *Member, IEEE*

***Abstract*— Several aspects of the design and optimization of coded multiple-antenna transmission diversity methods for slowly time-varying channels are explored from an information-theoretic perspective. Both optimized vector-coded systems, which can achieve the maximum possible performance, and suboptimal scalar-coded systems, which reduce complexity by exploiting suitably designed linear precoding, are investigated. The achievable rates and associated outage characteristics of these spatial diversity schemes are evaluated and compared, both for the case when temporal diversity is being jointly exploited and for the case when it is not. Complexity and implementation issues more generally are also discussed.**

***Index Terms*— Antenna arrays, antenna precoding, capacity, diversity, fading channels, Gaussian channels, outage probability, wireless communication.**

I. Introduction

IT is well known [10] that multiple-element antenna arrays can improve the performance of a wireless communication system in a fading environment. These antenna arrays may be employed either at the transmitter or the receiver. In a mobile radio system, it is generally most practical to employ an antenna array at the base station rather than at the mobile units. Then, in transmitting from the mobile to the base station, diversity is achieved through a multiple-element receive antenna array ("receiver diversity"), while in transmitting from the base station to the mobiles, diversity is achieved through a multiple-element transmit antenna array ("transmitter diversity"). In this paper, we focus on transmitter diversity, the achievable performance limits of which are comparatively less well understood.

Transmitter diversity has traditionally been viewed as more difficult to exploit than receiver diversity, in part because the transmitter is assumed to know less about the channel than the receiver, and in part because of the challenging signal design problem: the transmitter is permitted to generate a different signal at each antenna element. Unlike the receiver diversity case, where independently faded copies of a single transmitted signal may be combined optimally to achieve a performance gain, for transmitter diversity the many transmitted signals are already combined when they reach the receiver. We assume throughout that the combining coefficients are unknown to the transmitter.

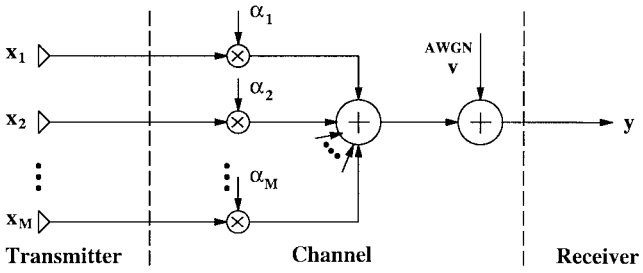

Fig. 1. Diversity channel with an M-element transmit antenna array.

We model the M-element antenna transmitter diversity channel as shown in Fig. 1. The complex baseband received signal

$$y_k = \sum_{i=1}^{M} \alpha_i x_{i,k} + v_k \tag{1}$$

at time k is the superposition of the M transmitted symbols $x_{1,k}, \cdots, x_{M,k}$, each scaled and phase-shifted by a complex coefficient α_i that represents the aggregate effect of the channel encountered by antenna element i. As (1) reflects, the channel is frequency-nonselective, i.e., the delay spread of the channel is small compared to the symbol duration. The additive noise v_k is a white circularly symmetric Gaussian random process with variance (for each real and imaginary component) $N_0/2$, and the transmitted energy over a block of N symbols is limited to

$$\frac{1}{N} \sum_{k=1}^{N} \sum_{i} |x_{i,k}|^2 \leq \mathcal{E}_s$$

per symbol.[1] We focus on diversity methods that conserve bandwidth; the symbol rate of the channel is, therefore, fixed and independent of the number of antenna elements.

We model the coefficient vector $\boldsymbol{\alpha} = [\alpha_1 \ \alpha_2 \ \cdots \ \alpha_M]^T$ as effectively constant over a long block of symbols. This model is appropriate when the transmitter, receiver, and all reflecting

Manuscript received November 19, 1996; revised May 19, 1999. This work was supported in part by NSF under Grants NCR-9314341 and MIP-9502885, by ONR under Contract N00014-96-1-0930, and by the Department of the Air Force under Contract F19628-95-C-0002. The material in this paper was presented in part at ISITA'96.

A. Narula is with MIT Lincoln Laboratory, Lexington, MA, 02420 USA.

M. D. Trott is with ArrayComm, Inc., San Jose, CA, 95134 USA.

G. W. Wornell is with the Department of Electrical Engineering and Computer Science, and the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Communicated by N. Seshadri, Associate Editor for Coding Techniques.

Publisher Item Identifier S 0018-9448(99)07445-3.

[1] We keep $\mathcal{E}_s$ independent of M, so that as M increases, the fixed energy $\mathcal{E}_s$ must be distributed more thinly among the antenna elements. This allows us to distinguish the impact of varying the number of antenna elements from the impact of varying the total transmitted power.

surfaces are either stationary or moving slowly relative to the carrier wavelength and symbol rate.

We restrict our attention to the case in which the transmitter has no knowledge of these channel coefficients but the receiver has perfect knowledge. The lack of knowledge of $\boldsymbol{\alpha}$ at the transmitter represents either a lack of feedback from the receiver to the transmitter, or a broadcast scenario where the transmitter must send the same information to many receivers with different locations and hence different $\boldsymbol{\alpha}$'s. In practice, the receiver can estimate the channel parameters $\boldsymbol{\alpha}$ quite accurately if the channel varies slowly enough. More generally, the scenario with perfect receiver knowledge provides a useful bound on the performance of systems where only estimates are available (cf. Section VI-A).

Our results on transmitter diversity may be summarized as follows.

- We classify existing transmitter diversity methods into two approaches, which we term *vector coding* and *scalar coding*. The information-theoretic commonalities between the scalar-coding methods have not been previously recognized, nor have the fundamental performance differences between the scalar- and vector-coding approaches. These differences are not exposed, for example, by prior analyses based on uncoded bit-error rate.
- We compare the performance of transmitter diversity schemes in terms of both outage regions—a nonstochastic concept—and outage probabilities. Here, an *outage* occurs when mutual information falls below a prescribed threshold. Power savings at a given outage probability are determined for both finite number of antennas and in the limit as $M \to \infty$. We show that scalar-coded systems can come quite close to the performance of vector-coded ones at a fraction of complexity for typical array sizes, and that both are vastly superior to systems that do not exploit transmitter diversity.
- We show that optimum vector coding achieves the same performance as repetition diversity, without the factor of M increase in bandwidth.
- We introduce new vector- and scalar-coded diversity schemes with performance and complexity advantages. Our vector-coded construction, which exploits multiple-access coding to solve a single-user problem, is a useful benchmark against which to compare promising alternatives emerging in the literature.
- We compare temporal and spatial diversity and show that in the large-diversity limit, vector-coded spatial diversity can provide up to a roughly 2.51-dB performance advantage over either temporal diversity or scalar-coded spatial diversity.

The detailed organization of the paper is as follows. In Section II, we introduce the coded antenna systems evaluated in the remainder of the paper and determine the mutual information achieved by each. The mutual informations are compared in the form of outage regions in Section III, where we establish the fundamental performance gap between vector and scalar coding. In Section IV, we let the channel parameters vary with time according to a Rayleigh model, then evaluate and compare the associated rate and outage characteristics of the different coded antenna systems by examining the distributions of their respective mutual informations. In Section V, code design issues and implementation of both scalar- and vector-coded systems are discussed, as are ways of upgrading existing systems to at least partly realize the potential gains of transmitter diversity. Finally, to allow the results of this paper to be appreciated in context of a broader array processing literature, Section VI discusses how still further gains are achievable with transmitter diversity when side information about the propagation channel is available at the transmitter via feedback. In such scenarios, we discuss how transmitter diversity schemes behave more like beamforming systems or directive arrays from some key perspectives.

II. Diversity Methods and Mutual Information

In the absence of complexity and delay constraints at the transmitter and receiver, the performance limits of transmitter diversity can be determined by examining the information-theoretic characteristics of the multiple-input, single-output channel (1). These performance limits can be approached arbitrarily closely through the use of suitably designed coding across the antenna array. We refer to systems that use such vector-valued codebooks as "vector-coded antenna systems."

Another approach for constraining complexity in transmitter diversity systems uses more conventional codes in conjunction with linear preprocessing at the antenna array that converts the multiple-input, single-output channel into a single-input, single-output channel. We refer to such structures as "scalar-coded antenna systems."

In Section II-A, we use the above taxonomy to classify the transmitter diversity methods that have appeared in the literature. Additional assumptions and notation are introduced in Section II-B. The mutual informations achieved by vector and scalar coding are computed in Sections II-C and II-D, respectively. We interpret several prototypical scalar-coded methods as time–frequency duals, and introduce a new randomized scalar-coded technique with performance advantages. Finally, in Section II-E, we compute the mutual information achieved by repetition diversity and show that its use of channel resources is comparatively wasteful.

A. Prior and Concurrent Work

The simplest form of scalar-coded transmitter diversity uses repetition coding to transmit orthogonal versions of a signal from each antenna element. The M-fold repetition causes a factor of M in bandwidth expansion. An early example of repetition diversity using disjoint frequency bands may be found in Brinkely [2]; see also the summary in Jakes [10]. More recent examples include the rapid phase sweeping arrangements of Hattori and Hirade [8] and Weerackody [26]. The latter method is applied to a direct sequence spread-spectrum system, where (in a certain sense) no additional bandwidth expansion occurs beyond that already caused by the spreading.

Scalar-coded methods that conserve bandwidth were first introduced by Wittneben [27], who proposed linear time-

invariant precoding combined with maximum-likelihood detection. This method is further analyzed and extended by Winters [25], and Seshadri and Winters [22]. The advantageous combination of error-control coding and linear precoding was recognized by Hiroike, Adachi, and Nakajima [9], who achieve diversity through block coding and slow phase sweeping. Kuo and Fitz [12] analyze the performance of the phase-sweeping approach in detail. Finally, in [28], [29], and [20, Ch. 1], broader classes of linear precoding methods with attractive, low-complexity linear receiver structures are developed for use with scalar-coded antenna systems.

In addition to our own work [15], the vector-coding approach, with extensions to both multiple transmit and receive antennas, is fully embraced in the recent work of Foschini [4], Foschini and Gans [5], and Tarokh, Seshadri, and Calderbank [23]. These papers aim to realize much of the potential gains of transmitter diversity with computationally efficient encoding and decoding strategies.

B. Assumptions

To analyze the performance limits of both vector-coded and scalar-coded transmitter antenna systems, we examine the mutual information between input and output over a long block of symbols, following the approach of Ozarow, Shamai, and Wyner [18], which corresponds in an approximate sense to the maximum achievable rate of reliable communication. Formally, we define this mutual information as

$$I = \frac{1}{N} \lim_{N\to\infty} I_N \tag{2}$$

where $I_N = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_N; Y_1, Y_2, \cdots, Y_N)$ is the mutual information between a block of N input and output symbols. The existence of this limit is straightforward to establish in all cases we consider.

We restrict our attention to input codebooks that are distributed according to complex circularly symmetric white Gaussian random processes. Gaussian codebooks are appropriate because the receiver knows $\boldsymbol{\alpha}$, which reduces the multiple-element antenna channel to an additive white Gaussian noise (AWGN) channel; the use of codebooks with independent and identically distributed (i.i.d.) components follows from standard arguments. We emphasize that beamforming and waterfilling methods cannot be applied because the transmitter does not have knowledge of the channel parameters.

All analysis is done using a discrete-time channel model rather than the more physically correct continuous-time, strictly time-limited, approximately bandlimited model adopted in Gallager [6]. A development based on the continuous-time model resolves a number of technical problems in what follows, such as the use of a finite block-length channel code with infinite impulse response bandpass filters, but at the expense of a considerably more cumbersome analysis. Moreover, the conclusions are effectively the same; for a comparison, see [16].

To make our development as broadly applicable as possible, our initial results below and in Section III assume no stochastic model for $\boldsymbol{\alpha}$. That is, we express performance as a function of the coefficients of the realized channel. As we will see, the performance measures of interest depend on these coefficients through their individual magnitudes $|\alpha_1|, |\alpha_2|, \cdots, |\alpha_M|$, not on their relative phases. Furthermore, in some cases the dependence on these parameters is through the magnitude of the coefficient vector alone, i.e.,

$$||\boldsymbol{\alpha}|| = \sqrt{|\alpha_1|^2 + |\alpha_2|^2 + \cdots + |\alpha_M|^2}.$$

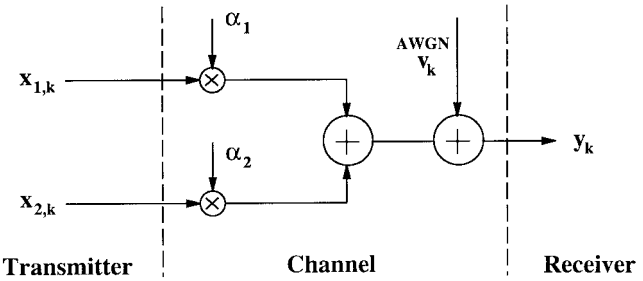

Fig. 2. Channel corresponding to vector-coded antenna array; $M = 2$ case.

In later sections of the paper we adopt a Rayleigh model, where the components of $\boldsymbol{\alpha}$ are i.i.d. zero-mean, complex Gaussian random variables, i.e., the components of $\boldsymbol{\alpha}$ have uniform phase and Rayleigh magnitude. This model is generally appropriate when there is no line of sight to the receiver, when there are a large number of reflected paths, and when the antenna elements are at least $1/2$ wavelength apart.[2] The stochastic model allows performance to be expressed probabilistically in terms of the Rayleigh statistics.

C. Vector-Coded Transmitter Antenna Systems

In this section we consider fully unconstrained signaling schemes for the memoryless vector-input scalar-output power-limited Gaussian channel, which is depicted in Fig. 2 for the case $M = 2$. With an M-element transmit antenna array, the complex baseband received signal at time k is

$$Y_k = \boldsymbol{\alpha}^T \boldsymbol{X}_k + V_k \tag{3}$$

where $\boldsymbol{X}_k = [X_{1,k}, \cdots, X_{M,k}]^T$ is the input vector and V_k is complex white Gaussian noise with variance N_0.

When the components of $\boldsymbol{X}_k$ are independent, zero-mean complex-valued, circularly symmetric Gaussian random variables each with variance $E|X_{i,k}|^2 = \mathcal{E}_s/M$, the output Y_k is zero-mean Gaussian with variance $||\boldsymbol{\alpha}||^2(\mathcal{E}_s/M) + N_0$. The mutual information is then

$$\begin{aligned} I_{\mathrm{OPT}} &= h(Y) - h(Y|\boldsymbol{X}) \\ &= \log\left(1 + \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s}{M N_0}\right) \end{aligned} \tag{4}$$

which we emphasize depends on the channel parameters only through the antenna gain $||\boldsymbol{\alpha}||$.

Selecting the components of $\boldsymbol{X}$ to be i.i.d. is not "optimal" in some global sense. Deviating the covariance matrix $E[\boldsymbol{X}\boldsymbol{X}^T]$ from a scaled identity matrix increases the mutual information

[2] If the antenna elements are extremely widely distributed relative to the propagation distance to the receiver, as would be the case if they were placed throughout a building, then a more complex model must be adopted that accounts for the strong attenuation of more distant antenna elements. We do not consider this case here.

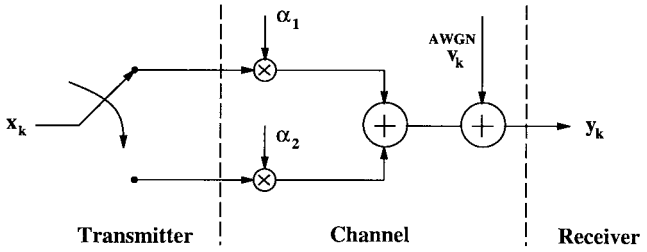

Fig. 3. Scalar-coded antenna system via time division.

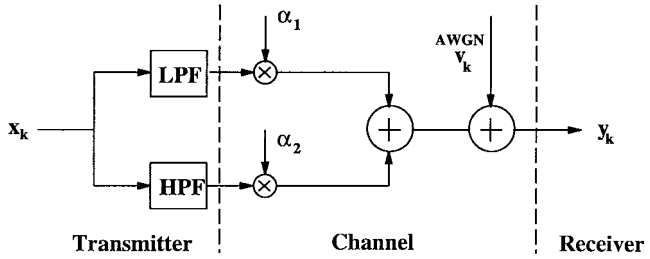

Fig. 4. Scalar-coded antenna system via frequency division.

for some values of $\boldsymbol{\alpha}$ and decreases it for others. However, codebooks for which the components of $\boldsymbol{X}$ are i.i.d. achieve the saddle point solution to a max-min problem in which nature chooses a distribution on $\boldsymbol{\alpha}$ to minimize the rate of reliable communication [16].

An approach to the design of specific codes that asymptotically achieve this mutual information is described in Section V-A. However, in general, the decoding of the transmitted message from the received signal can be computationally demanding when such vector codes are used. Therefore, there is considerable interest in efficient antenna coding strategies for which low-complexity decoding algorithms are available. The scalar-coded methods we consider next are one such class.

D. Scalar-Coded Transmitter Antenna Systems

Scalar-coded antenna systems can be viewed as coded antenna systems that are used in conjunction with linear "precoding." The vector input to the channel is transformed into a scalar input through a suitably designed single-input, multiple-output linear system; a good code is then designed for the associated single-input, single-output composite channel. In this section, we investigate some important prototypical examples of such systems.

As in the vector-coded case, for each scalar-coded method, we evaluate performance with coding under the constraint that the input sequence $\{X_k\}$ be i.i.d. complex circularly symmetric Gaussian with energy $E|X_k|^2 = \mathcal{E}_s$ per symbol.

1) Time- and Frequency-Division Systems: Time and frequency division, which are duals of one another, exploit transmit antenna diversity by using linear precoding to generate orthogonal signals for each antenna element (which remain orthogonal at the receiver, in the absence of intersymbol interference or Doppler spread in the channel). As a result of this orthogonality, the multiple-element antenna channel can then be analyzed as a set of independent parallel channels.

With the time-division approach [22], the input to the antenna array is formed by time-multiplexing a scalar-coded symbol stream $\{X_k\}$ across the antenna elements. Symbols are "dealt" to the antenna elements in a periodic manner, so that X_k is transmitted using antenna element i when $k \equiv i \,(\mathrm{mod}\, M)$. For example, when $M = 2$, odd-time inputs are transmitted on antenna element 1 and even-time inputs are transmitted on antenna element 2, as depicted in Fig. 3.

Time division converts spatial (antenna) diversity into time diversity: the time-invariant vector-input channel is transformed into a periodically time-varying scalar-input channel. In particular, the output of this channel employing time division is

$$Y_k = h_k X_k + V_k \tag{5a}$$

where

$$h_k = \alpha_{(k \bmod M)+1}. \tag{5b}$$

The mutual information between input and output of this channel, a measure of the achievable rate of reliable communication of this coded system, is the average of the mutual informations achieved by each antenna element, i.e.,

$$I_{\mathrm{TD}} = \frac{1}{M}\sum_{i=1}^{M} \log\left(1 + \frac{|\alpha_i|^2 \mathcal{E}_s}{N_0}\right). \tag{6}$$

Analogously, with frequency-division systems, the input process $\{X_k\}$ is frequency-multiplexed across the antenna elements by bandpass filtering into M disjoint bands, each of width π/M. This is illustrated in Fig. 4 for the case $M = 2$.

Frequency-division systems convert spatial diversity into frequency diversity: the memoryless vector-input channel is transformed into a scalar-input channel with intersymbol interference. In particular, the output of this channel is

$$Y_k = \sum_n h_n X_{k-n} + V_k \tag{7}$$

where the unit-sample response h_k has Fourier transform

$$H(\omega) = \begin{cases} \alpha_1, & 0 \le |\omega| < \pi/M \\ \alpha_2, & \pi/M \le |\omega| < 2\pi/M \\ & \vdots \\ \alpha_M, & (M-1)\pi/M \le |\omega| < \pi. \end{cases} \tag{8}$$

In the frequency-division case, the mutual information between input and output follows as

$$\begin{aligned} I_{\mathrm{FD}} &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(1 + \frac{\mathcal{E}_s |H(\omega)|^2}{N_0}\right) d\omega \\ &= \frac{1}{M}\sum_{i=1}^{M} \log\left(1 + \frac{|\alpha_i|^2 \mathcal{E}_s}{N_0}\right). \end{aligned} \tag{9}$$

Comparing (6) with (9) reveals that time division and frequency division yield the same mutual information. More generally, many linear methods that generate orthogonal signals (that remain orthogonal after passing through the channel) have the same characteristic behavior as time and frequency division.

The performance of such orthogonal systems can achieve that of optimized vector-coded antennas only in special cases.

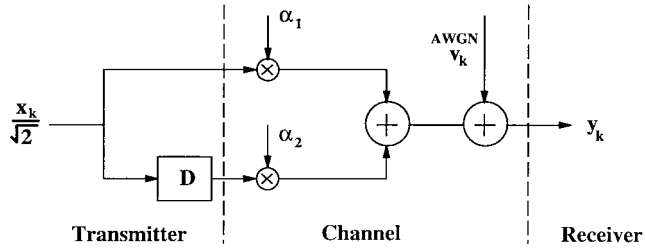

Fig. 5. Scalar-coded antenna system by time-shifting; $M = 2$ case.

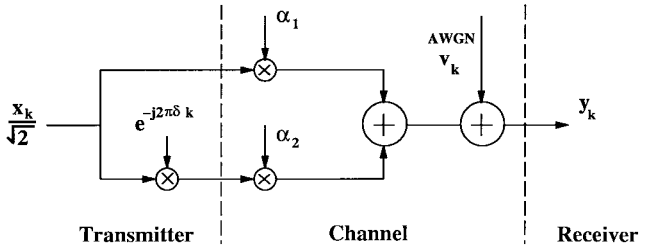

Fig. 6. Scalar-coded antenna system by frequency-shifting; $M = 2$ case.

To see this, apply Jensen's inequality to (6) or (9) and compare to (4):

$$I_{\mathrm{TD}} = I_{\mathrm{FD}} \leq I_{\mathrm{OPT}}$$

with equality if and only if all antenna element gains are equal, i.e.,

$$|\alpha_1| = |\alpha_2| = \cdots = |\alpha_M|.$$

2) Time- and Frequency-Shifting Systems: Another class of scalar-coded systems are obtained by a time- or frequency-shifting approach. In contrast to time- and frequency-division systems, these diversity methods do not generate orthogonal signals at the antenna elements.

The time-shifting diversity method [27] sends delayed versions of a common input signal over the constituent transmit antenna elements: the ith antenna element carries the input signal $\{X_k\}$ delayed by $i-1$ time steps. This system is illustrated for the case $M = 2$ in Fig. 5.

Like the frequency-division system, time-shifting transforms spatial diversity into frequency diversity, turning the memoryless vector-input channel into a scalar-input channel with intersymbol interference: the output of the channel is

$$Y_k = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} \alpha_i X_{k-i+1} + V_k, \tag{10}$$

so that the Fourier transform of the unit-sample response of the associated channel is

$$H(\omega) = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} e^{-j(i-1)\omega} \alpha_i. \tag{11}$$

For the time-shifting method, the mutual information between input and output of the scalar channel can be computed from the frequency response (11) as

$$I_{\mathrm{TS}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(1 + \frac{\mathcal{E}_s}{MN_0} \left| \sum_{i=1}^{M} e^{-j(i-1)\omega} \alpha_i \right|^2 \right) d\omega. \tag{12}$$

The dual of time shifting is frequency shifting [9], [26], which sends modulated versions of a common input signal over the different elements of the transmit antenna array: the ith antenna element carries the signal $\{e^{j\pi\delta(i-1)k} X_k\}$, where $\delta > 0$ is an arbitrary modulation parameter. This system is illustrated in Fig. 6 for the case $M = 2$.

Frequency shifting, like time division, transforms spatial diversity into time diversity. The output of the channel is

$$Y_k = h_k X_k + V_k \tag{13a}$$

where

$$h_k = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} e^{j\pi\delta(i-1)k} \alpha_i. \tag{13b}$$

Computing the mutual information between input and output for the frequency-shifting diversity method is slightly more involved. From (13), over a block of N symbols, the average mutual information is

$$\frac{1}{N} I_N = \frac{1}{N} \sum_{k=1}^{N} \log \left(1 + \frac{\mathcal{E}_s}{N_0} |h_k X_k|^2 \right), \tag{14}$$

which under the mild technical condition that δ is irrational [19], converges to

$$I_{\mathrm{FS}} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left(1 + \frac{\mathcal{E}_s}{MN_0} \left| \sum_{i=1}^{M} e^{j(i-1)\omega} \alpha_i \right|^2 \right) d\omega \tag{15}$$

as $N \to \infty$.

Comparing (15) to (12), we see that time- and frequency-shifting methods achieve the same performance when used in conjunction with suitably designed codes. Comparing to (6) and (9), however, we see that the two methods perform differently than time and frequency division for any particular realized channel. To find the cases where time and frequency shifting meet the performance of the optimal vector-coded channel, apply Jensen's inequality to (12) or (15) and compare to (4):

$$I_{\mathrm{TS}} = I_{\mathrm{FS}} \leq I_{\mathrm{OPT}}$$

with equality if and only if at most a single component α_i of $\boldsymbol{\alpha}$ is nonzero.

3) Randomized Systems: While the performance of optimum vector-coded antenna systems depends on the channel coefficients only through $||\boldsymbol{\alpha}||$, the performance of the four scalar-coded antenna systems developed thus far depends not simply on $||\boldsymbol{\alpha}||$, but on the magnitudes of the constituent coefficients themselves. In this section, we develop a randomized antenna precoding strategy for which the associated mutual information depends on the channel coefficients only through

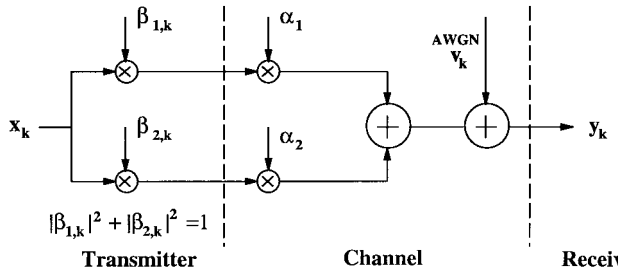

Fig. 7. Scalar-coded antenna system by random time weighting; $M = 2$ case.

$||\boldsymbol{\alpha}||$. As we will see in Section V, this randomization simplifies the code design problem for the scalar-coded antenna system.

The specific system we consider implements a random time-weighting strategy to transform the vector-input channel into a scalar-input channel. The vector input is generated by multiplying the scalar input by a complex-valued, unit-magnitude random vector $\boldsymbol{\beta}_k$, i.e.,

$$\boldsymbol{X}_k = \boldsymbol{\beta}_k X_k \tag{16}$$

where $\boldsymbol{\beta}_k = [\beta_{1,k}\ \beta_{2,k}\ \cdots\ \beta_{M,k}]^T$ is chosen randomly and uniformly over the surface of the M-dimensional complex unit sphere. We make the weighting vector known at the receiver, which is readily achieved in practice by selecting $\boldsymbol{\beta}_k$ pseudo-randomly according to a prearranged scheme. The scheme is illustrated in Fig. 7 for the case $M = 2$.

Like time-division and frequency-shifting methods, random weighting transforms spatial diversity into time diversity; the output of the channel is

$$Y_k = h_k X_k + V_k \tag{17a}$$

where

$$h_k = \boldsymbol{\alpha}^T \boldsymbol{\beta}_k. \tag{17b}$$

To evaluate the mutual information associated with this randomized scheme, we first show that the channel (17) can be conveniently rewritten in the form

$$Y_k = ||\boldsymbol{\alpha}||\mu_k X_k + V_k \tag{18}$$

where μ_k denotes a random variable that is the upper left entry of a matrix drawn from what is referred to as the "circular unitary ensemble." The ensemble is defined by the unique distribution on unitary matrices that is invariant under left and right unitary transformation [14]. That is, given a random matrix U drawn from the circular unitary ensemble, for any unitary matrix Q, both QU and UQ are equal in distribution to U.

Equation (18) is obtained as follows. First, we interpret $\boldsymbol{\beta}_k$ as the first column of a random unitary matrix U_k drawn from the circular unitary ensemble, i.e.,

$$\boldsymbol{\beta}_k = U_k[1 \quad 0 \quad \cdots \quad 0]^T. \tag{19}$$

Next, we similarly write the vector of coefficients $\boldsymbol{\alpha}$ as an appropriately normalized unit vector multiplied by a unitary matrix $\hat{U}(\boldsymbol{\alpha})$. This rearrangement of (17) yields

$$Y_k = [||\boldsymbol{\alpha}|| \quad 0 \quad \cdots \quad 0]\hat{U}(\boldsymbol{\alpha})U_k \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} X_k + V_k. \tag{20}$$

Finally, because the unitary matrix $\hat{U}(\boldsymbol{\alpha})U_k$ is equal in distribution to U_k, (18) follows.

With the channel expressed in the form (18), the mutual information achieved by random time weighting follows readily as

$$I_{\mathrm{RAN}} = \int_0^1 \log\left(1 + \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s \eta}{N_0}\right) f_{|\mu|^2}(\eta)\, d\eta \tag{21a}$$

where $f_{|\mu|^2}(\cdot)$ is the probability density function of the squared magnitude of any entry μ of a matrix drawn from the circular unitary ensemble. As developed in Appendix A, this density takes the form

$$f_{|\mu|^2}(\eta) = \begin{cases} (M-1)(1-\eta)^{M-2}, & 0 < \eta \le 1 \\ 0, & \text{otherwise} \end{cases} \tag{21b}$$

for $M \ge 2$.

The randomized time-weighting system developed here is optimum over a broad class of scalar-coded antenna systems whose performance depends on the channel only through $||\boldsymbol{\alpha}||$. For example, choosing $\boldsymbol{\beta}$ according to any other distribution that is left-invariant under unitary transformation also gives a mutual information that depends only on $||\boldsymbol{\alpha}||$. However, the first column of the circular unitary ensemble is the only such distribution with $||\boldsymbol{\beta}||$ identically 1, and thus it can be verified using Jensen's inequality that all other such ensembles—which necessarily allow $||\boldsymbol{\beta}||$ to vary—achieve strictly lower mutual information.

Nevertheless, the randomized scalar-coded antenna system cannot achieve the performance of the optimum vector-coded antenna system on any nontrivial channel. To see this, apply Jensen's inequality to (21) and compare to (4):

$$I_{\mathrm{RAN}} \le I_{\mathrm{OPT}}$$

with equality if and only if $||\boldsymbol{\alpha}|| = 0$.

Other randomized scalar-coded systems can be developed that convert spatial diversity to time diversity, though their mutual informations do not in general depend on the channel parameters only through $||\boldsymbol{\alpha}||$. For example, a randomized time-division system selects an antenna element randomly and uniformly for the transmission of each symbol. A randomized frequency-shifting system selects the phases of the antenna elements independently and uniformly over $[0, 2\pi)$ for each symbol. It is straightforward to verify that the mutual informations associated with these schemes are identical to their deterministic counterparts. Finally, one can define a random frequency-weighting scheme dual to and with similar properties as the random time-weighting scheme developed in this section.

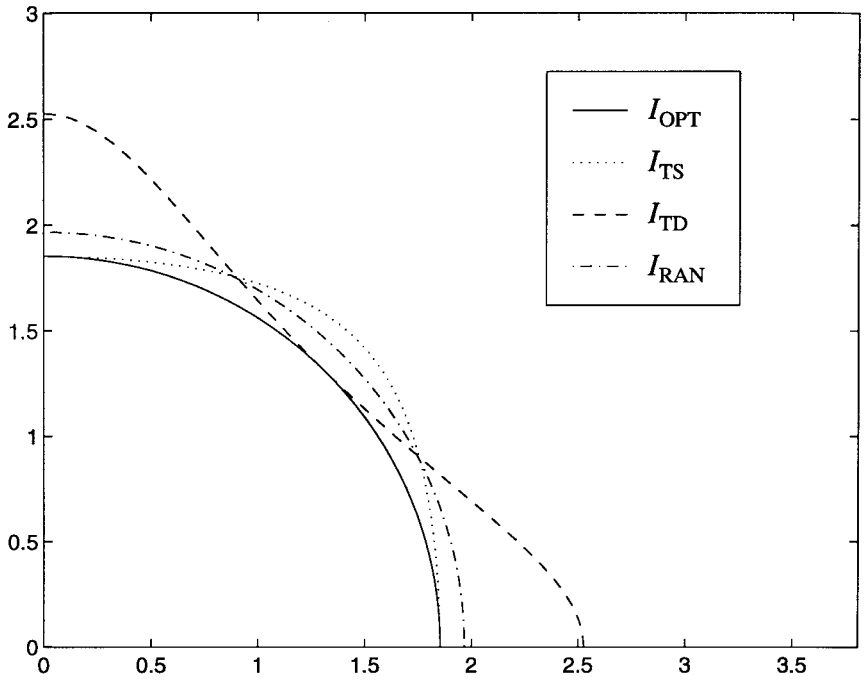

Fig. 8. Representative outage region boundaries: values of $|\alpha_1|\sqrt{\mathcal{E}_s/N_0}$ and $|\alpha_2|\sqrt{\mathcal{E}_s/N_0}$ that correspond to a mutual information of 1 nat/complex symbol. For values of (α_1, α_2) outside these curves, reliable communication is possible at rate 1; for values inside, attempting communication at rate 1 will result in outage.

E. Repetition-Coded Transmitter Diversity

Perhaps the oldest form of scalar-coded transmitter diversity is repetition coding across orthogonal carriers [2], [10]. Unlike the frequency-division method described in Section II-D1, repetition expands bandwidth by a factor of M. Thus for this section only, we deviate from our assumption that bandwidth is conserved.

We assign each antenna element a separate carrier and repeat a common scalar-valued symbol across the carriers. Each antenna receives $1/M$th the total power. The effective channel model becomes

$$\boldsymbol{Y}_k = \boldsymbol{\alpha} X_k/M + \boldsymbol{V}_k \tag{22}$$

where $\boldsymbol{Y}_k$ is an M-vector of channel output symbols at time k, one per carrier, and $\boldsymbol{V}_k$ is an i.i.d. complex Gaussian vector with variance N_0 in each component.

The mutual information between X and $\boldsymbol{Y}$ is easily computed as

$$I_{\mathrm{REP}} = \log\left(1 + \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s}{MN_0}\right) \tag{23}$$

which, interestingly, is identical to the performance I_{OPT} achieved by the optimum vector-coded antenna system (4).

From this perspective repetition coding is quite inefficient—no performance benefit is obtained from the M-fold increase in bandwidth. On the other hand, the complexity of "minimum bandwidth" vector-coding as described subsequently in Section V-A2 is comparatively high: it requires the use of error-correction codes designed for time-varying channels and an M-user decoder. As a result, in applications where ultra-low computational complexity is important but bandwidth is plentiful, simple multiband repeated-transmission methods are quite attractive.

III. Outage Regions

For the system designer, the performance characteristics of the methods developed in Section II may be usefully compared in terms of their associated outage regions in the space of channel gains $|\alpha_1|, |\alpha_2|, \cdots, |\alpha_M|$. The outage region has the following defining property: within the outage region, reliable communication is not possible at the desired rate; outside the outage region, reliable communication is possible at or above the desired transmission rate. Such outage regions are delimited by the surface of constant mutual information corresponding to the target transmission rate.

As a representative example, the outage region boundaries for transmission at $I = 1$ nat per symbol using $M = 2$ antenna elements are depicted in Fig. 8. The solid innermost quarter circle[3] corresponds to the optimum vector-coded antenna system whose performance is given by (4). The dashed curve tangent to the quarter circle at $|\alpha_1| = |\alpha_2|$ corresponds to the scalar-coded antenna systems obtained via the time- and frequency-division (orthogonal signaling) approaches, whose performance is given by (6). The dotted curve tangent to the quarter circle at both $|\alpha_1| = 0$ and $|\alpha_2| = 0$ corresponds to the scalar-coded antenna systems obtained via the time- and frequency-shifting approaches, whose performance is given by (12). Finally, the dash-dotted outer quarter circle corresponds to the scalar-coded antenna

[3] As noted in Section II-C, deviating from the assumption that each antenna receives equal power changes the outage region; for the vector-coded case the quarter circle becomes an ellipse.

system obtained via the randomized time-weighting method, whose performance is given by (21).

As Fig. 8 reflects, the outage region for the optimum vector-coded antenna system is strictly smaller than the outage regions for the scalar-coded antenna systems. However, none of the scalar-coded antenna systems dominates the others for all $\boldsymbol{\alpha}$.

A useful measure of the benefit of vector coding over scalar coding is in terms of the additional rate that can be supported at a given signal-to-noise ratio (SNR), or equivalently in terms of the additional power required to support a target transmission rate. For most scalar-coding methods, these measures depend on the individual realized channel parameters. However, for scalar coding based on randomized time weighting, these measures are conveniently parameter-independent. This is because the performance of both vector coding and scalar coding via randomized time weighting depend on the channel parameters only through $||\boldsymbol{\alpha}||$.

To determine the additional rate that can be supported using vector coding over scalar coding in this case, we subtract (21a) from (4), insert (21b) into the result, and set $z = M\eta$, yielding

$$I_{\mathrm{OPT}} - I_{\mathrm{RAN}} = \int_0^M \log\left(\frac{1+\mathrm{SNR}}{1+z\,\mathrm{SNR}}\right)\frac{M-1}{M} \cdot \left(1-\frac{z}{M}\right)^{M-2} dz \tag{24}$$

where

$$\mathrm{SNR} = \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s}{MN_0} \tag{25}$$

is the per antenna element SNR. For large M and $\mathrm{SNR} \gg 1$, the difference is

$$I_{\mathrm{OPT}} - I_{\mathrm{RAN}} \approx \int_0^\infty \log\left(\frac{1}{z}\right) e^{-z}\, dz = \gamma \tag{26}$$

where γ is Euler's constant—approximately 0.833 bits per symbol. Hence, when the SNR per antenna element is large, scalar-coded systems require roughly $10\gamma \log_{10} e \approx 2.51$ dB more signal power to achieve the same outage region as vector-coded ones. As the number of antenna elements decreases, so does the performance of both vector-coded and scalar-coded systems, and in turn the gap between them. For example, with an $M = 8$ element array the high-SNR gap is 0.740 bit/symbol (2.23 dB), with an $M = 4$ element array it is 0.649 bit/symbol (1.94 dB), and with an $M = 2$ element array it is 0.443 bit/symbol (1.33 dB).

IV. Time-Varying Channels

In this section we consider a scenario in which the coefficient vector $\boldsymbol{\alpha}$ has i.i.d. Rayleigh components (see Section II-B) and varies in time according to a stationary ergodic fading model. In this case, both spatial and temporal diversity can, in principle, be jointly exploited through the use of suitably designed coding.

We focus on the case in which the coherence time of the fading process is finite but large, so that $\boldsymbol{\alpha}$ is effectively constant over a block.[4] To incorporate the effects of temporal diversity we allow coding to span a sequence of blocks $k = 1, \cdots, L$, where $\boldsymbol{\alpha}_k$ represents the coefficient vector for block k. This model is useful for slow frequency-hopped systems in which each block of symbols is transmitted on a different band.

In Section IV-A, we consider spatial diversity without temporal diversity, and demonstrate that both vector- and scalar-coded systems with even modest sized arrays offer dramatic performance enhancements over systems without such arrays. In particular, the use of transmitter diversity allows a target rate to be achieved at a given outage probability with substantially less signal power. In Section IV-B, we compare spatial to temporal diversity in the limit of a large number of antennas M and a large number of blocks L. Vector and scalar coding behave quite differently in these limits: while in all cases the outage probability drops to zero, scalar coding and temporal diversity achieve only the capacity of a Rayleigh fading channel, while vector coding achieves a larger capacity equal to that of an AWGN channel with the same average SNR.

A. Exploiting Spatial Diversity Without Temporal Diversity

In the absence of temporal diversity, spatial diversity, in the form of either vector or scalar coding, has a large impact on system performance. We measure this impact by the additional power—for a given outage probability—that must be transmitted in order to achieve the same rate over an AWGN channel having the same average channel gain and noise power. We compute this performance loss as a function of the number of antenna elements M and the target outage probability. Measuring performance in terms of outage probability requires that we know not only the expected value of mutual information, but the degree of variation of mutual information about the mean. Specifically, the outage probability associated with any particular transmission rate depends on the tail behavior of the mutual information distribution.

Note that the channels we consider have no capacity in the usual sense. Indeed, when finitely many antenna elements M are used and the message is described by finitely many blocks L, mutual information takes values arbitrarily close to zero, so there is always some nontrivial probability of decoding error regardless of rate. (See Ozarow, Shamai, and Wyner [18] and references therein for a more thorough discussion.)

For vector-coded antenna systems, it is straightforward to determine the transmission rate R at which outage occurs with some prescribed probability ϵ. We assume without loss of generality that $E[|\alpha_i|^2] = 1$ for $i = 1, 2, \cdots, M$. Let $\delta_{M,\epsilon}$ denote the threshold on $||\boldsymbol{\alpha}||^2$ at which outage occurs with probability ϵ for an M-element antenna array, i.e.,

$$P(||\boldsymbol{\alpha}||^2 \le \delta_{M,\epsilon}) = \epsilon. \tag{27}$$

We obtain, via (4), that the vector-coded system achieves the target outage probability of ϵ at a maximum rate of

$$R = \log\left(1 + \frac{\delta_{M,\epsilon}\mathcal{E}_s}{MN_0}\right). \tag{28}$$

[4]This assumption may be relaxed if we exclude the scalar-coded diversity methods that use temporal filtering.

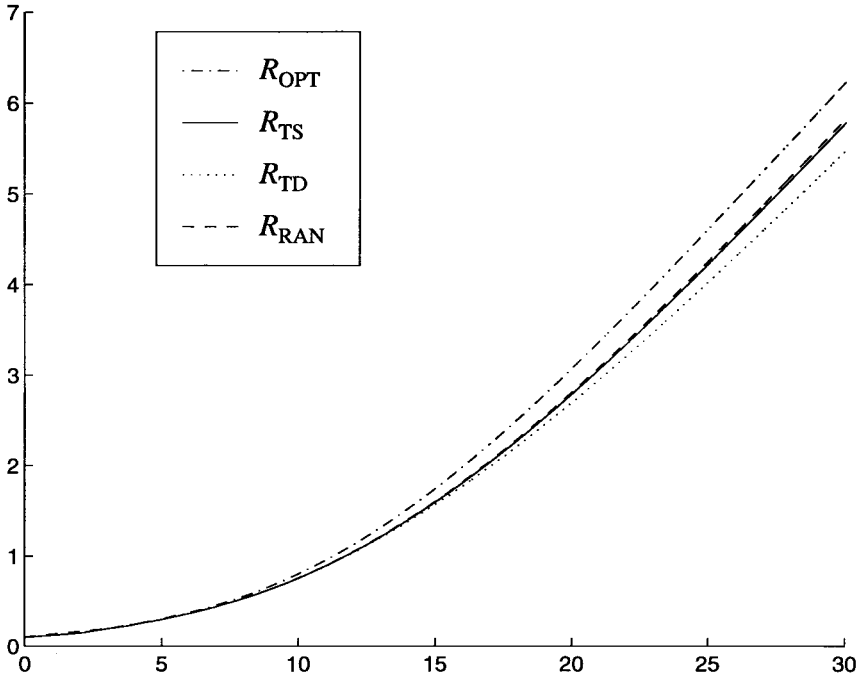

Fig. 9. Rate in bits per second versus expected SNR at an outage probability of $\epsilon = 0.01$ with an $M = 2$ element array, for both vector-coded antenna systems and scalar-coded antenna systems obtained by time-shifting, time-division, and random weighting.

TABLE I
ADDITIONAL SIGNAL POWER $\delta_{M,\epsilon}/M$ IN DECIBELS FOR VECTOR-CODED ANTENNA SYSTEMS TO ACHIEVE RATE OF AWGN CHANNEL WITH OUTAGE PROBABILITY ϵ

M	$\epsilon = 10^{-2}$	$\epsilon = 10^{-4}$	$\epsilon = 10^{-6}$
1	19.98	40.00	60.00
2	11.29	21.48	31.50
4	6.87	12.37	17.51
8	4.40	7.61	10.43
16	2.91	4.91	6.57
256	0.65	1.06	1.36

The additional signal power in decibels required to achieve the rate of the corresponding AWGN channel is, therefore, $10\log_{10}(\delta_{M,\epsilon}/M)$. This additional power is tabulated for various values of M and ϵ in Table I. Calculating these quantities is straightforward: it suffices to recognize that since $||\boldsymbol{\alpha}||^2$ is the sum of M independent exponential random variables, it has an Mth-order Erlang cumulative distribution function that can be approximated according to

$$\begin{aligned} P(||\boldsymbol{\alpha}||^2 \le \delta) &= \int_0^{\delta} \frac{x^{M-1}}{(M-1)!} e^{-x}\, dx \\ &= e^{-\delta} \sum_{r=M}^{\infty} \frac{\delta^r}{r!} \\ &\approx \frac{1}{M!}\delta^M e^{-\delta} \end{aligned} \tag{29}$$

for small δ.

As Table I indicates, even an array with a small number of elements dramatically reduces the power needed to achieve reliable communication. As is typical with other forms of diversity, the amount of additional power required approximately halves when M is doubled. Moreover, as we will confirm in Section IV-B, the loss approaches 0 dB as $M \to \infty$ because the mutual information for the vector-coded system converges almost surely to the AWGN channel capacity.

Similar behavior is achieved through the use of scalar-coded systems, as we show in the remainder of this section. Outage probabilities for scalar-coded systems are invariably larger than for optimized vector-coded ones—regardless of the stochastic model on $\boldsymbol{\alpha}$. This follows from our result in Section III that the outage region of the optimized vector-coded antenna system lies wholly inside the outage regions of the scalar-coded antenna systems. However, as we now illustrate in the case of a simple two-element $(M = 2)$ antenna array, the reductions achievable using scalar-coded systems are generally comparable to those achievable using vector-coded systems.

To facilitate the comparison, we begin by noting that Ozarow *et al.* [18] have calculated the probabilities of outage—or equivalently, the cumulative distribution functions of mutual information—that correspond to time-shifting [18, eq. (2.26b)] and time-division [18, eq. (3.4)]. We compute the cumulative distribution function of the mutual information achieved by the random time-weighting method using Monte Carlo simulation. Using these distributions, the maximum rate of reliable communication can be determined as a function of expected SNR with the outage probability held fixed at any desired level. The resulting curves are depicted in Fig. 9 for the case of an outage probability of 1%.

As Fig. 9 reflects, the gap between random weighting and time shifting is nearly zero, while time division performs slightly worse. The optimum vector-coded antenna system is somewhat better than random weighting at high SNR; as derived in Section III, with $M = 2$ antenna elements the gap is about 0.443 bit/symbol or 1.33 dB. Using this result with

the corresponding entry in Table I we see, e.g., that the scalar-coded antenna system based on randomized time-weighting requires at most $11.29 + 1.33 = 12.62$ dB more power than the corresponding AWGN channel to support the same rate at an outage probability of 0.01. Similar calculations can be used to determine how other entries in Table I change when vector coding is replaced with scalar coding.

B. Jointly Exploiting Spatial and Temporal Diversity

The mutual information achieved by vector coding (4) immediately generalizes to the case of temporal diversity spanning L blocks:

$$I_{\mathrm{OPT},L} = \frac{1}{L}\sum_{k=1}^{L} \log\left(1 + \frac{||\boldsymbol{\alpha}_k||^2 \mathcal{E}_s}{MN_0}\right). \tag{30}$$

The mutual information is a function of the random process $\{\boldsymbol{\alpha}_k\}$ and is therefore a random variable itself.

Consider first the limit of infinite temporal diversity and fixed spatial diversity. As the number of blocks L spanned by the code becomes arbitrarily large, for a stationary ergodic fading process $\{\boldsymbol{\alpha}_k\}$ the mutual information (30) converges to the expected value of (4), namely,

$$C_{\mathrm{OPT}} = E\left[\log\left(1 + \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s}{MN_0}\right)\right]. \tag{31}$$

Equation (31) may be interpreted as the Shannon capacity of the channel in the usual sense.

With $M = 1$ there is no spatial diversity, so that

$$Y_k = \alpha_{1,k} X_k + V_k. \tag{32}$$

Now (31) specializes to

$$C_{\mathrm{RAY}} = E\left[\log\left(1 + \frac{|\alpha_1|^2 \mathcal{E}_s}{N_0}\right)\right] \tag{33}$$

which is the capacity of an i.i.d. Rayleigh fading channel with perfect channel state information at the receiver originally derived in [3].

Consider next the limit of infinite spatial diversity and fixed temporal diversity. As the number of antennas M grows to infinity, $||\boldsymbol{\alpha}||^2/M$ converges almost surely to its expected value $E|\alpha_1|^2$. Then, for any amount L of temporal diversity, $I_{\mathrm{OPT},L}$ and hence its expected value C_{OPT} (31) converge almost surely to

$$C_{\mathrm{AWGN}} = \log\left(1 + \frac{E[|\alpha_1|^2]\mathcal{E}_s}{N_0}\right), \tag{34}$$

the capacity of an AWGN channel with the same average SNR.

As is well known (see, e.g., [13]), the gap between (33) and (34) increases with SNR and at high SNR is a maximum of $10\gamma \log_{10} e \approx 2.51$ dB. More generally, expected mutual information (31) increases monotonically within this range with increasing M. Hence, when temporal diversity is fully exploited, spatial diversity in the form of vector-coded antenna systems can further improve performance by up to roughly 2.51 dB. Such a gain is small but often worth pursuing.

The behavior of scalar-coded systems is quite different. We show next that as $L \to \infty$, mutual information converges to C_{RAY} (33) for every scalar-coded method described in Section II-D, regardless of the number of antennas. Thus augmenting full temporal diversity with spatial diversity yields *no* additional performance benefit. We similarly show that as $M \to \infty$, mutual information also converges to C_{RAY},[5] confirming that the performance gap between the vector and scalar approaches persists in the large-array limit.

Consider first the limit of infinite temporal diversity and fixed spatial diversity. Given a stationary ergodic fading process $\{\boldsymbol{\alpha}_k\}$, with infinite temporal diversity the scalar-coded methods have a Shannon capacity equal to the expected value of their respective mutual informations. The mutual informations achieved by time division and frequency division are equal for all $\boldsymbol{\alpha}$, hence their expected values are equal. Similarly, the expected mutual informations achieved by time and frequency shifting are equal. To show that the expected value of the respective mutual informations (6), (15), and (21) of time division, frequency shifting, and random weighting are equal, we make the following direct argument from the channel model.

As discussed in Section II-D3, time division, frequency shifting, and random weighting all have the same general form. The channel output is

$$Y = \boldsymbol{\alpha}^T \boldsymbol{\beta} X + V \tag{35}$$

where $\boldsymbol{\beta}$ is chosen either randomly or according to some deterministic pattern. Since in all cases $||\boldsymbol{\beta}|| \equiv 1$

$$\boldsymbol{\alpha}^T \boldsymbol{\beta} = \boldsymbol{\alpha}^T U(\boldsymbol{\beta})[1 \quad 0 \quad \cdots \quad 0]^T \tag{36}$$

for some suitable unitary matrix $U(\boldsymbol{\beta})$. The components of $\boldsymbol{\alpha}$ are i.i.d. complex circularly symmetric Gaussian random variables, hence the distribution of $\boldsymbol{\alpha}$ is invariant under unitary transformation. This implies that $\boldsymbol{\alpha}^T\boldsymbol{\beta}$ is equal in distribution to

$$\boldsymbol{\alpha}^T[1 \quad 0 \quad \cdots \quad 0]^T = \alpha_1.$$

Thus as $L \to \infty$, the scalar-coded multiple-antenna systems described by (35) have the same capacity (33) as the single-antenna channel (32).

Consider next the limit of infinite spatial diversity and fixed temporal diversity. We argue in the remainder of this section that for all scalar-coded methods we have considered, mutual information again converges to the capacity (33) of a Rayleigh fading channel.

The result is straightforward to prove for scalar-coded systems obtained by time division: by the strong law of large numbers, the mutual information (6) achieved by time-division systems with an M-element array converges almost surely to its expected value.

A proof for scalar-coded systems obtained by time shifting, given in Appendix B, is based on the idea that as M gets large, $H(\omega)$ in (11) becomes statistically independent at distinct values of ω. Integrating $\log(1 + |H(\omega)|^2 \mathcal{E}_s/N_0)$ over ω then corresponds to adding up an infinite number

[5]This result was conjectured in [18] for a system equivalent to the scalar-coded system obtained via time shifting.

of independent random variables, which by the law of large numbers converges to the expected mutual information.

In turn, since the mutual information of the channels using frequency division and time shifting have a form equivalent to time division and frequency shifting, respectively, the arguments above also establish that the mutual informations achieved by these methods converge to their expected value as $M \to \infty$. Finally, in Appendix C, we prove this result for the random time-weighting technique.

V. Code Design and Integration Issues

In this section, we discuss implementation aspects of the coded antenna systems evaluated in this paper. Coding strategies for the vector- and scalar-coded diversity methods are considered in Section V-A; we argue that scalar methods should use channel codes designed for Rayleigh fading, while vector methods require more sophisticated techniques. We introduce one such technique, based on a virtual multiuser approach, in Section V-A2.

A. Channel-Coding Strategies

Among the diversity methods considered here, vector-coded antenna systems and scalar-coded antenna systems obtained via random time weighting seem most appealing in terms of robustness and performance. We therefore focus our code design discussion on these two systems, beginning with the latter. We consider only spatial and not temporal diversity; the channel parameters $\boldsymbol{\alpha}$ are viewed as deterministic and time-invariant.

1) Scalar-Coded Systems: Although they cannot achieve the full performance of optimum vector-coded systems, scalar-coded systems have straightforward practical implementations. In particular, as we now show, the scalar-coded system derived from random time weighting creates a channel with synthetic i.i.d. fading that approximates Rayleigh fading as the number of antenna elements increases, and so a codebook designed for a Rayleigh fading channel can be used efficiently for at least moderately large arrays.

To see that random weighting asymptotically transforms the vector-input time-invariant channel into a (frequency-nonselective) Rayleigh fading channel, rewrite (18) from Section II-D3 as

$$Y_k = \frac{||\boldsymbol{\alpha}||}{\sqrt{M}} Z_k X_k + V_k \tag{37}$$

with $Z_k = \sqrt{M}\mu_k$. The process Z_k is i.i.d., zero mean, and unit variance. As $M \to \infty$, Z_k converges in distribution to a circularly symmetric complex Gaussian random variable (cf. (45) in Appendix A), as required. The SNR of the synthetic fading channel is determined by $\mathcal{E}_s/N_0$ and the average antenna gain $||\boldsymbol{\alpha}||^2/M$.

2) Vector-Coded Systems: Practical vector codes that approach the optimum achievable performance are more challenging to design than the scalar-coding method described above. Using, for example, standard quadrature phase-shift keying (QPSK) signal sets on each antenna element in a straightforward manner results in a received waveform that is computationally expensive to decode after the constituent signal sets are rotated, scaled, and summed.

We provide a conceptually useful approach to this code design problem by converting the vector-input time-invariant channel into M (approximately) Rayleigh fading channels driven by M virtual users. Standard coding methods for Rayleigh channels can then be directly applied.

Signaling reliably at the vector-coded optimum rate I_{OPT} (4) requires a coded system whose outage region depends on the channel parameters only through the antenna gain $||\boldsymbol{\alpha}||$. Directly assigning the ith virtual user to the ith antenna element is, therefore, a bad strategy, for the rate achievable by user i depends on $|\alpha_i|$.

We remove this dependence by homogenizing the antenna array before transmission. Specifically, we premultiply the vectors $\boldsymbol{X}_k$ of coded virtual user symbols by matrices U_k drawn pseudorandomly—in a manner known to the receiver—from the circular unitary ensemble. The output of the channel is then

$$Y_k = \boldsymbol{\alpha}^T U_k \boldsymbol{X}_k + V_k \tag{38}$$

$$= ||\boldsymbol{\alpha}|| \boldsymbol{\beta}_k^T \boldsymbol{X}_k + V_k \tag{39}$$

where $\boldsymbol{\beta}_k$ as given by (19) is the first column of a matrix from the circular unitary ensemble. The channel remains i.i.d. and memoryless, but, because the components of $\boldsymbol{\beta}_k$ have the same marginal densities, each virtual antenna element (controlled by a component of $\boldsymbol{X}_k$) now "looks" the same. We assign one virtual user to each virtual antenna element, sharing transmit power evenly, and apply multiple-access coding with a "stripping"-style decoder. The virtual users operate in a time-synchronized but otherwise noncooperative fashion.

It remains to prove that the achievable rate of each virtual user depends on $\boldsymbol{\alpha}$ only through $||\boldsymbol{\alpha}||$ and that the sum of the achievable rates equals I_{OPT}. We consider the case of a two-element $(M = 2)$ antenna array; generalization to $M > 2$ is straightforward.

With $M = 2$, successive decoding (stripping) achieves mutual informations for the first and second virtual users equal to

$$I_{\mathrm{OPT1}} = E\left[\log\left(1 + \frac{||\boldsymbol{\alpha}||^2 |\beta_{1,k}|^2 \mathcal{E}_s/2}{N_0 + ||\boldsymbol{\alpha}||^2 |\beta_{2,k}|^2 \mathcal{E}_s/2}\right)\right] \tag{40}$$

and

$$I_{\mathrm{OPT2}} = E\left[\log\left(1 + \frac{||\boldsymbol{\alpha}||^2 |\beta_{2,k}|^2 \mathcal{E}_s/2}{N_0}\right)\right] \tag{41}$$

respectively, where expectations are taken over $\boldsymbol{\beta}_k$. The first user has lower mutual information due to interference from the second. As required, (40) and (41) depend on $\boldsymbol{\alpha}$ only through its magnitude $||\boldsymbol{\alpha}||$, hence the rates of the two virtual users can be selected so that decoding simultaneously fails when $||\boldsymbol{\alpha}||$ drops below a prescribed outage threshold.

The sum of the achievable rates is

$$I_{\mathrm{OPT1}} + I_{\mathrm{OPT2}} = E\left[\log\left(1 + \frac{||\boldsymbol{\alpha}||^2 ||\boldsymbol{\beta}_k||^2 \mathcal{E}_s}{2N_0}\right)\right]$$

$$= \log\left(1 + \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s}{2N_0}\right)$$

$$= I_{\mathrm{OPT}}$$

where the first equality results from combining (40) and (41), the second follows from the fact that $||\boldsymbol{\beta}_k|| \equiv 1$, and the third from (4). Thus regardless of the realized channel, performance is never sacrificed by either the randomization (homogenization) process nor the noncooperative nature of the coding by virtual users.[6]

In terms of its implementation, complexity for this virtual M-user system is roughly M times larger than that of a single-element antenna system.[7] This means that for a price of roughly twice the complexity, a vector-coded antenna system can achieve the performance of the simpler randomized scalar-coded system with roughly 1.33 dB less power (see Section III) when an $M = 2$ element array is used. By using progressively larger arrays, one can reduce the power requirements up to roughly an additional $2.51 - 1.33 = 1.18$ dB using such vector coding, but the cost in complexity per decibel of gain grows steeply in this regime.

B. Adding Diversity to Existing Systems

When upgrading existing wireless communication systems to take advantage of transmitter antenna arrays, the specific performance benefits of vector- and scalar-coded systems must be weighed against the cost of the system modifications required to realize these enhancements. In this section, we comment on some of the tradeoffs involved.

To begin, it is important to recognize that the scalar-coding techniques we have described are considerably easier to integrate into existing systems than vector-coding techniques. Moreover, while our development of scalar-coded antenna systems has emphasized discrete-time implementations, implementations in continuous time at passband are a convenient way to upgrade a system designed for a single-element antenna to use transmitter diversity via a multiple-element array.

To realize the full performance gain possible with such an array requires that the transmitter and receiver processing be subsequently redesigned for the (artificially created) time-varying channel. In principle, this need not be too difficult: the time variation follows a pattern known to the receiver, so channel identification is no harder than learning the slowly varying coefficient vector $\boldsymbol{\alpha}$. Moreover, more modest gains are possible even without such redesign. Likewise, for best performance, the error-correction portion of the system needs to be redesigned for use in conjunction with the associated precoding, although in practice suboptimal coding can be used to realize more limited but still significant benefits. In fact, the linear precoding systems described in this paper, or more generally those developed in [29] and [20, Ch. 1], provide a substantial performance benefit even when used without coding.

[6] Given the convexity of mutual information it may seem paradoxical that converting a time-invariant channel into a time-varying one does not degrade performance. The resolution to this conundrum is that since $||\boldsymbol{\beta}_k|| \equiv 1$, the two virtual channels do not fade independently: when one is good the other is bad, and the full antenna gain is used at all times.

[7] Conveniently, the channel identification and power-control problems normally associated with stripping are not as severe as with a true M-user system because the multiple virtual users arise from the coordinated action of a single user.

VI. Discussion and Concluding Remarks

We close with some additional insights and perspectives on transmit diversity in general and the results of this paper in particular.

A. Ideal Beamforming: Transmitter Arrays with Side Information

Our results on vector- and scalar-coded antenna systems are equally applicable to a single-user scenario with an unknown channel coefficient vector $\boldsymbol{\alpha}$ and a broadcast scenario where the number of receivers is much larger than the number of array elements M. If side information is considered, however, fundamental differences between these two scenarios emerge. In particular, when the goal is to ensure a uniform SNR among a large number of receivers, the benefits of side information are inherently rather limited. By contrast, with a single receiver, power efficiency can be enhanced by a factor of up to M by using feedback to send information about the channel to the transmitter, as we now show.

Consider the case of a single receiver with $\boldsymbol{\alpha}$ known perfectly at the transmitter. The mutual information achievable with such side information follows as a special case of the results of Salz and Wyner [21]; mutual information is maximized by what can be viewed as *beamforming*, i.e., by setting

$$\boldsymbol{X} = \frac{\boldsymbol{\alpha}^*}{||\boldsymbol{\alpha}||} X \tag{42}$$

where X is a zero-mean complex Gaussian random variable with variance $E|X|^2 = \mathcal{E}_s$. The channel output is then

$$Y = ||\boldsymbol{\alpha}||X + V \tag{43}$$

and the resulting mutual information is

$$I_{\mathrm{BMF}} = \log\left(1 + \frac{||\boldsymbol{\alpha}||^2 \mathcal{E}_s}{N_0}\right). \tag{44}$$

Comparing with (4), we see that when the transmitter knows the coefficient vector $\boldsymbol{\alpha}$, a target level of mutual information can be achieved with a factor of M less power than required by a vector-coded antenna system not having such transmitter side information. For large arrays this difference is dramatic. Substantial gains are still possible even when only partial side information is available, as shown in [17].

The factor of M SNR enhancement provided by feedback can be understood as follows. With ideal beamforming, the covariance matrix of $\boldsymbol{X}$ has rank 1, with the principal component "steered" in the direction of $\boldsymbol{X}$ as (42) indicates. By contrast, instead of steering the antenna beampattern in a particular direction, vector-coded antenna systems create a field that is spatially white: the covariance matrix of $\boldsymbol{X}$ is a scaled identity matrix; the curves of constant likelihood are spheres, and the fraction of energy in $\boldsymbol{X}$ in the direction of $\boldsymbol{\alpha}$ is always $1/M$.

Frequency shifting, time division, and random time weighting implement a form of time-varying beamforming—in particular, for each symbol a rank 1 covariance matrix Λ is chosen for $\boldsymbol{X}$, where the principal component of Λ follows a pattern that depends on the method. When the principal component

happens to be closely aligned with the direction of $\boldsymbol{\alpha}$ (or a phase-shifted version thereof), then essentially optimum SNR is achieved. For other principal component choices, the signal is highly attenuated. This results in loss in performance compared to ideal beamforming; that scalar coding performs worse than vector coding is due to the higher variability.

B. Diversity Versus Directive Arrays

In the preceding section, we developed a transmission strategy we found convenient to refer to as "beamforming," terminology that arises more frequently in the directive array literature than in the diversity literature. More generally, the communication and signal processing literature loosely uses terms such as "diversity," "directive arrays," and "beamforming" to describe the ways in which a multiple-element antenna—at either the transmitter or receiver—can be used to improve communication. There appears to be no widely accepted definitions of these terms, and, interestingly, the literature on diversity is often conceptually orthogonal to that on beamforming; compare, for example, Jakes [10] to Johnson and Dudgeon [11]. To allow our results to be appreciated in the context of this broader literature, we discuss in this section how the difference between diversity and beamforming is not so much in their methods, but rather in their modeling assumptions for the channel.

In the traditional directive array or beamforming scenario, there is no stochastic model for the channel parameters $\boldsymbol{\alpha}$. Rather, there is a dominant direction of arrival for the signal, which in turn implies that the components of $\boldsymbol{\alpha}$ have a particular parameterized structure, and the antenna is assumed to know the arrival angle almost perfectly. The array pattern is adjusted by manipulating the gains and phases of the antenna elements, or, more generally, by combining the array with a multiple-input or multiple-output linear filter. For a transmit array, the channel knowledge is used to focus as much energy in the direction of the receiver as possible. For a receive array, the gain of the antenna is maximized in the direction of the transmitter. By reciprocity, transmitter and receiver beamforming have the same performance, e.g., the same SNR or mutual information (44).

In traditional diversity scenarios, on the other hand, a stochastic model for the channel parameters $\boldsymbol{\alpha}$ is inherently assumed. This stochastic model is typically derived assuming a large number of multipath reflections, so that signals do not have a meaningful notion of direction of arrival and the components of $\boldsymbol{\alpha}$ are comparatively much less correlated. As with beamforming, the array pattern is adjusted by manipulating the gains and phases of the antenna elements. However, most typically the array *does not* have full knowledge of the channel: the phases are generally unknown while the gains may be unknown or partly known through a statistical characterization. In other words, the array is free to choose an antenna directivity pattern, but it does not know where the peaks and nulls will fall. The receiver usually has more information about the channel than the transmitter, which means that transmit and receive diversity may behave quite differently.

From these perspectives, the results of Section VI-A help bridge the gap between results in the diversity and directive array literature. More generally, the distinction between diversity and directive arrays blurs in many applications. For example, in a diversity setting, as bandwidth increases, multiple paths may begin to resolve, rendering a stochastic model invalid. In a beamforming setting, as the number of transmit antenna elements increases and the power per antenna element decreases, channel identification becomes more difficult, making the assumption of perfect channel knowledge unrealistic. Ultimately, the achievable performance in such cases depends critically on the channel model.

C. Open Problems

As an emerging research area, many important and interesting aspects of the design of coded antenna systems remain to be explored. We close by summarizing just two of the more obvious issues that warrant further investigation.

First, our work has not considered the effects of a channel with limited coherence time or coherence bandwidth. It would be worth exploring the possibility of a general theory that allows time diversity, frequency diversity, and spatial diversity to be jointly exploited through a simultaneous conversion of these sources of variability into something resembling an i.i.d. Rayleigh fading channel with slowly varying SNR.

Second, the problem of channel identification at the receiver needs to be explored in more detail. For a time-varying channel, the identification problem will become increasingly difficult as the number of antenna elements increases and power is spread more thinly among these elements. Diminishing returns will eventually set in; an incorrectly designed system might even deteriorate with additional antenna elements. To fully understand this issue would require a theory of communication over channels on the margin of identifiability. While this theory does not presently appear to exist, recent preliminary work [7], [24] suggests a framework that might be adapted to exploring these issues. We anticipate results in this area having important implications for future system design.

Appendix A
Density of Matrix Element from Circular Unitary Ensemble

To evaluate $f_{|\mu|^2}(\cdot)$ in (21b), note that μ is a component of a vector chosen randomly and uniformly over the surface of the complex unit sphere. Such a vector can be generated taking a vector with zero-mean i.i.d. complex Gaussian components and scaling it by its norm. Thus μ may be written as

$$\mu = \frac{G_1}{\sqrt{|G_1|^2 + |G_2|^2 + \cdots + |G_M|^2}} \tag{45}$$

where $G_1, G_2, \cdots, G_M$ are zero-mean i.i.d. complex circularly symmetric Gaussian random variables. The squared magnitude of a complex circularly symmetric Gaussian is exponentially distributed, and hence $|\mu|^2$ may be written as

$$|\mu|^2 = \frac{E_1}{E_1 + E_2 + \cdots + E_M} \tag{46}$$

where $E_1, E_2, \cdots, E_M$ are i.i.d. exponential random variables. The density function (21b) follows in a straightforward manner.

APPENDIX B
TIME-SHIFTING DIVERSITY AS $M \to \infty$

In this appendix we show that for the time-shifting diversity scheme, as the number of transmitters M increases to infinity, mutual information converges in mean square to its expected value.

The mutual information of an M transmitter channel employing time shifting diversity is given in (12). Without loss of generality we will assume that $\mathcal{E}_s/N_0 = 1$. Then

$$I_{\mathrm{TS}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(1+|H_M(\omega)|^2\right) d\omega \tag{47}$$

where

$$H_M(\omega) = \sum_{i=1}^{M} \frac{\alpha_i}{\sqrt{M}} e^{-j\omega(i-1)}. \tag{48}$$

Recall that the coefficients α_i are i.i.d. zero-mean complex Gaussian random variables. This implies that $H_M(\omega)$ is also a complex Gaussian random variable with mean zero and variance $E|\alpha_1|^2$.

We will show that as the number of transmitters increases to infinity, the mutual information converges to its expected value

$$\lim_{M\to\infty} \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(1+|H_M(\omega)|^2\right) d\omega = E[\log\left(1+|\alpha_1|^2\right)]. \tag{49}$$

It is sufficient to show that the variance of I_{TS} goes to zero:

$$\lim_{M\to\infty} \sigma^2_{I_{\mathrm{TS}}} = 0. \tag{50}$$

To calculate the variance of I_{TS} as $M \to \infty$, we first calculate

$$E[I^2_{\mathrm{TS}}] = \frac{1}{4\pi^2} E\left[\int_{-\pi}^{\pi} \log\left(1+|H_M(\omega_1)|^2\right) d\omega_1 \cdot \int_{-\pi}^{\pi} \log\left(1+|H_M(\omega_2)|^2\right) d\omega_2\right] \tag{51}$$

$$= \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} E\Big[\log\left(1+|H_M(\omega_1)|^2\right) \cdot \log\left(1+|H_M(\omega_2)|^2\right)\Big] d\omega_1\, d\omega_2. \tag{52}$$

The integrand in (52), i.e.,

$$E[\log\left(1+|H_M(\omega_1)|^2\right)\log\left(1+|H_M(\omega_2)|^2\right)]$$

is bounded. To see this, we use the Schwartz Inequality and the fact that $\log(1+x^2) \le x^2$.

$$E[\log\left(1+|H_M(\omega_1)|^2\right)\log\left(1+|H_M(\omega_2)|^2\right)] \le \{E[\log\left(1+|H_M(\omega_1)|^2\right)]^2\}^{1/2} \cdot \{E[\log\left(1+|H_M(\omega_2)|^2\right)]^2\}^{1/2} \tag{53}$$

$$\le \{E|H_M(\omega_1)|^4\}^{1/2}\{E|H_M(\omega_2)|^4\}^{1/2} \tag{54}$$

$$= 3[E|\alpha_1|^2]^2 < \infty. \tag{55}$$

Therefore, by the Lebesgue Dominated Convergence Theorem

$$\lim_{M\to\infty} E[I^2_{\mathrm{TS}}] = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \lim_{M\to\infty} E\Big[\log\left(1+|H_M(\omega_1)|^2\right) \cdot \log\left(1+|H_M(\omega_2)|^2\right)\Big] d\omega_1\, d\omega_2. \tag{56}$$

It remains only to establish the following limit:

$$\lim_{M\to\infty} E[\log\left(1+|H_M(\omega_1)|^2\right)\log\left(1+|H_M(\omega_2)|^2\right)] = E[\log\left(1+|H_1|^2\right)]E[\log\left(1+|H_2|^2\right)] \tag{57}$$

where H_1 and H_2 are independent Gaussian random variables with mean zero and variance $E[|\alpha_1|^2]$. Indeed, using (57) with (56) we obtain

$$\lim_{M\to\infty} E[I^2_{\mathrm{TS}}] = (E[\log\left(1+|H_1|^2\right])^2 = (E[I_{\mathrm{TS}}])^2 \tag{58}$$

and thus

$$\lim_{M\to\infty} \sigma^2_{I_{\mathrm{TS}}} = 0. \tag{59}$$

To establish (57), we first show that in the limit as $M \to \infty$, $H_M(\omega_1)$ and $H_M(\omega_2)$ are uncorrelated for all $\omega_1 \ne \omega_2 \in [-\pi, \pi)$.

$$E[H_M(\omega_1)H^*_M(\omega_2)] = E\left[\sum_{i=1}^{M}\frac{\alpha_i}{\sqrt{M}}e^{-j\omega_1(i-1)}\sum_{k=1}^{M}\frac{\alpha_k}{\sqrt{M}}e^{-j\omega_2(k-1)}\right] \tag{60}$$

$$= \sum_{i=1}^{M}\sum_{k=1}^{M}\frac{E[\alpha_i\alpha_k^*]}{M}e^{-j\omega_1(i-1)}e^{-j\omega_2(k-1)} \tag{61}$$

$$= \sum_{i=1}^{M}\frac{E[|\alpha_i|^2]}{M}e^{-j(\omega_1-\omega_2)(i-1)} \tag{62}$$

$$= \frac{e^{-j(\omega_1-\omega_2)(M+1)}-1}{M(e^{-j(\omega_1-\omega_2)}-1)}. \tag{63}$$

Therefore,

$$\lim_{M\to\infty} E[H_M(\omega_1)H^*_M(\omega_2)] = 0 \tag{64}$$

for all ω_1, ω_2 such that $\omega_1 \not\equiv \omega_2 \pmod{2\pi}$. Thus as $M \to \infty$, $H_M(\omega_1)$ and $H_M(\omega_2)$ converge in distribution to independent Gaussian random variables.

Convergence in distribution is not enough for our needs. Let $X_M = H_M(\omega_1)$, $Y_M = H_M(\omega_2)$, $X = H_1$, and $Y = H_2$. We have shown that in the limit as $M \to \infty$, $(X_M, Y_M) \to (X, Y)$ in distribution. By Skorohod's Theorem [1, Theorem 25.6], we can construct random variables $(\tilde{X}_M, \tilde{Y}_M)$ and $(\tilde{X}, \tilde{Y})$ such that: (X_M, Y_M) and $(\tilde{X}_M, \tilde{Y}_M)$ have the same distribution for each M, (X, Y) and $(\tilde{X}, \tilde{Y})$ have the same distribution, and $(\tilde{X}_M, \tilde{Y}_M) \to (\tilde{X}, \tilde{Y})$ with probability 1. Therefore,

$$\lim_{M\to\infty} \log\left(1+|\tilde{X}_M|^2\right)\log\left(1+|\tilde{Y}_M|^2\right) = \log\left(1+|\tilde{X}|^2\right)\log\left(1+|\tilde{Y}|^2\right) \tag{65}$$

with probability 1. If we can take expectations, (57) is proven. From [1, Theorem 25.12], if $\log(1+|\tilde{X}_M|^2)\log(1+|\tilde{Y}_M|^2)$ is uniformly integrable for $M \geq 1$, then

$$\lim_{M\to\infty} E[\log(1+|\tilde{X}_M|^2)\log(1+|\tilde{Y}_M|^2)] = E[\log(1+|\tilde{X}|^2)]E[\log(1+|\tilde{Y}|^2)]. \quad (66)$$

From [1, p. 338], random variables Z_M are uniformly integrable if

$$\lim_{\alpha\to\infty} \sup_M \int_{[|Z_M|\geq\alpha]} |Z_M|\, dP = 0. \quad (67)$$

If $\sup_M E[|Z_M|^{1+\epsilon}] < \infty$ for some positive ϵ, then the Z_M are uniformly integrable because

$$\int_{[|Z_M|\geq\alpha]} |Z_M|\, dP \leq \frac{1}{\alpha^\epsilon} E[|Z_M|^{1+\epsilon}] \quad (68)$$

which goes to zero as $\alpha \to \infty$. Let

$$Z_M = \log(1+|\tilde{X}_M|^2)\log(1+|\tilde{Y}_M|^2)$$

and $\epsilon = 1$. Then using the Schwartz inequality

$$\sup_M E\{[\log(1+|\tilde{X}_M|^2)]^2[\log(1+|\tilde{Y}_M|^2)]^2\} \leq \sup_M E[|\tilde{X}_M|^4|\tilde{Y}_M|^4] \quad (69)$$

$$\leq \sup_M (E[|\tilde{X}_M|^8])^{1/2}(E[|\tilde{Y}_M|^8])^{1/2} \quad (70)$$

$$< \infty. \quad (71)$$

Appendix C
Random Time Weighting as $M \to \infty$

In this appendix we show that for the random time-weighting diversity scheme, as the number of transmitters M increases to infinity, mutual information converges in mean square to its expected value. Without loss of generality we assume that $\mathcal{E}_s/N_0 = 1$. The mutual information achieved by the scalar-coded antenna system that employs random time weighting is given in (21), which we rewrite as the conditional expectation

$$I_{\mathrm{RAN}} = E[\log(1+X_M Y_M)|X_M] \quad (72)$$

where $X_M = |\boldsymbol{\alpha}_M|^2/M$, Y_M has density

$$f_{Y_M}(y) = \begin{cases} \dfrac{M-1}{M}\left(1-\dfrac{y}{M}\right)^{M-2}, & 0 < y \leq M \\ 0, & \text{otherwise} \end{cases} \quad (73)$$

for $M \geq 2$, and X_M and Y_M are independent.

The second moment of I_{RAN} is

$$E[I_{\mathrm{RAN}}^2] = E[E[\log(1+X_M Y_M)|X_M]^2] \quad (74)$$

$$= E[E[\log(1+X_M Y_M) \cdot \log(1+X_M Y'_M)|X_M]] \quad (75)$$

$$= E[\log(1+X_M Y_M) \cdot \log(1+X_M Y'_M)] \quad (76)$$

where Y_M and Y'_M are equal in distribution and independent.

Observe that Y_M converges in distribution to an exponential random variable Y with mean 1, Y'_M similarly converges to an exponential Y', and X_M converges to 1 as $M \to \infty$. Paralleling Appendix B, we apply Skorohod's Theorem [1, Theorem 25.6] to construct random variables (Z_M, Z'_M) and (Z, Z') such that: $(X_M Y_M, X_M Y'_M)$ and (Z_M, Z'_M) have the same distribution for each M, $(1\cdot Y, 1\cdot Y')$ and (Z, Z') have the same distribution, and $(Z_M, Z'_M) \to (Z, Z')$ with probability 1.

Then

$$\lim_{M\to\infty} E[I_{\mathrm{RAN}}^2] = \lim_{M\to\infty} E[\log(1+Z_M)\log(1+Z'_M)] \quad (77)$$

$$= E[\lim_{M\to\infty} \log(1+Z_M)\log(1+Z'_M)] \quad (78)$$

$$= E[\log(1+Z)\log(1+Z')] \quad (79)$$

$$= E[\log(1+Z)]^2 \quad (80)$$

$$= \lim_{M\to\infty} E[I_{\mathrm{RAN}}]^2 \quad (81)$$

where (78) follows from uniform integrability and (80) follows because Z and Z' are independent and identically distributed.

Acknowledgment

The authors wish to thank A. Lapidoth for helpful discussions and V. Borkar for detailed assistance with Appendix B.

References

[1] P. Billingsley, *Probability and Measure*. New York: Wiley, 1995.
[2] J. R. Brinkley, "A method of increasing the range of VHF communication systems by multi-carrier amplitude modulation," *J. Inst. Elec. Eng.*, vol. 93, pt. III, pp. 159–176, May 1946.
[3] T. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 353–356, May 1970.
[4] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Lab. Tech. J.*, pp. 41–59, Autumn 1996.
[5] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.
[6] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[7] R. G. Gallager and M. Medard, "Bandwidth scaling for fading channels," in *Proc. Int. Symp. Information Theory* (Ulm, Germany, July 1997), p. 471.
[8] T. Hattori and K. Hirade, "Multitransmitter digital signal transmission by using offset frequency strategy in a land-mobile telephone system," *IEEE Trans. Veh. Technol.*, vol. VT-27, pp. 231–238, Nov. 1978.
[9] A. Hiroike, F. Adachi, and N. Nakajima, "Combined effects of phase sweeping transmitter diversity and channel coding," *IEEE Trans. Veh. Technol.*, vol. 41, pp. 170–176, May 1992.
[10] W. C. Jakes, Ed., *Microwave Mobile Communications*. New York: IEEE Press, 1974.
[11] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1993.
[12] W.-Y. Kuo and M. P. Fitz, "Design and analysis of transmitter diversity using intentional frequency offset for wireless communications," *IEEE Trans. Veh. Technol.*, vol. 46, pp. 871–881, Nov. 1997.
[13] W. C. Y. Lee, "Estimate of channel capacity in Rayleigh fading environment," *IEEE Trans. Veh. Technol.*, vol. 39, pp. 187–189, Aug. 1990.
[14] M. L. Mehta, *Random Matrices and the Statisical Theory of Energy Levels*. New York: Academic, 1967.
[15] A. Narula, M. D. Trott, and G. W. Wornell, "Information-theoretic analysis of multiple-antenna transmission diversity for fading channels," in *IEEE Int. Symp. Information Theory and Its Applications*, Sept. 1996.
[16] A. Narula, "Information theoretic analysis of multiple-antenna transmission diversity," Ph.D. dissertation, MIT, Cambridge, MA, 1997.
[17] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell, "Efficient use of side information in multiple-antenna data transmission over fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1423–1436, Oct. 1998.

[18] L. H. 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.
[19] K. Peterson, *Ergodic Theory*. Cambridge, U.K.: Cambridge Univ. Press, 1983.
[20] H. V. Poor and G. W. Wornell, Eds., *Wireless Communications: Signal Processing Perspectives*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[21] J. Salz and A. D. Wyner, "On data transmission over cross coupled multi-input, multi-output linear channels with applications to mobile radio," AT&T Bell Labs. Tech. Memo., May 1990.
[22] N. Seshadri and J. H. Winters, "Two signaling schemes for improving the error performance of frequency division duplex (FDD) transmission systems using transmitter antenna diversity," *Int. J. Wireless Inform. Networks*, vol. 1, no. 1, pp. 49–60, Jan. 1994.
[23] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 744–765, Mar. 1998.
[24] I. E. Teletar and D. Tse, "Capacity and mutual information of broadband multipath fading channels," in *Proc. Int. Symp. Information Theory* (Cambridge, MA, Aug. 1998). p. 395.
[25] J. H. Winters, "The diversity gain of transmit diversity in wireless systems with Rayleigh fading," in *Proc. Int. Conf. Commununications*, May 1994, pp. 1121–1125.
[26] V. Weerackody, "Diversity for the direct-sequence spread spectrum system using multiple transmit antennas," in *Proc. IEEE ICC'93*, May 1993, pp. 1775–1779.
[27] A. Wittneben, "Basestation modulation diversity for digital SIMULCAST," in *Proc. IEEE VTC'91*, May 1991, pp. 848–853.
[28] G. W. Wornell and M. D. Trott, "Signal processing techniques for efficient use of transmit diversity in wireless communications," in *Proc. ICASSP*, May 1996, vol. 2, pp. 1057–1060.
[29] ———, "Efficient signal processing techniques for exploiting transmit antenna diversity on fading channels," *IEEE Trans. Signal Processing* (Special Issue on Signal Processing Advances in Communications), vol. 45, pp. 191–205, Jan. 1997.