Space–Time Communication for OFDM With Implicit Channel Feedback

Gwen Barriac, *Student Member, IEEE,* and Upamanyu Madhow, *Senior Member, IEEE*

***Abstract*—We consider wideband communication (e.g., using orthogonal frequency-division multiplexed (OFDM) systems) over a typical cellular "downlink," in which both the base station and the mobile may have multiple antennas, but the number of antennas at the mobile is assumed to be small. Implicit channel feedback can play a powerful role in such systems, especially for outdoor channels, which typically exhibit narrow spatial spreads. A summary of our findings is as follows.**

a) Implicit channel feedback regarding the covariance matrix for the downlink space–time channel can be obtained, without any power or bandwidth overhead, by suitably averaging uplink channel measurements across frequency. Since this approach relies on *statistical* reciprocity, it applies to both time-division duplex (TDD) and frequency-division duplex (FDD) systems. Using such covariance feedback yields significantly better performance at lower complexity than conventional space–time or space–frequency codes, which do not employ feedback.

b) We provide guidelines for optimizing antenna spacing in systems with covariance feedback. Theoretical investigation of a hypothetical system with completely controllable channel eigenvalues shows that the optimal number of channel eigenmodes is roughly matched to the (small) number of receive antenna elements. Thus, while antenna elements in conventional systems without feedback should be spaced far apart in order to ensure uncorrelated responses, the optimal antenna spacing with covariance feedback is much smaller, thereby concentrating the channel energy into a small number of eigenmodes.

***Index Terms*—Diversity methods, fading channels, feedback communication, information rates, multiple-input multiple-output (MIMO) systems.**

I. Introduction

MUCH of the literature on space–time communication deals with narrowband communication over a rich scattering environment typical of indoor channels. In contrast, the focus of this paper is on wideband communication over outdoor channels, which have higher delay spreads but much narrower spatial spreads. As shown in Fig. 1, we consider a base station (BS) with N_T antennas and a mobile with N_R antennas, where $N_T \gg N_R$. The receiver is assumed to have perfect channel state information, while the transmitter has no direct channel feedback. We will show that, under our assumptions, it is possible to design transceivers that yield better performance at lower complexity than conventional space–time coding and multiple-input multiple-output (MIMO) strategies. Our aim is to optimize the downlink from BS to mobile, which is the bottleneck for many asymmetric applications, such as data downloads or audio/video delivery. While we consider Shannon-theoretic performance measures, our design prescriptions apply directly to orthogonal frequency-division multiplexed (OFDM) systems. Some of our ideas may also apply to wideband direct sequence spread-spectrum signaling, but this setting is not considered in detail here.

A wideband channel sees three kinds of correlation: spatial, spectral, and temporal. For typical outdoor channels with small spatial spreads, the BS antennas see correlated channel responses from the mobile, unless the spacing between the antenna elements is very large. The thrust of this paper is to exploit such *spatial correlations* at the BS transmitter. Specifically, a transmission strategy which sends power along the dominant channel eigenmodes produces large performance gains while reducing complexity at the transmitter and receiver. We also explicitly account for the effect of the channel power–delay profile on the *spectral correlations* among the channel realizations in different frequency bins. We limit attention to transmission strategies in which the BS sends independent and identically distributed (i.i.d.) symbols in each frequency bin, which implies that the ergodic capacity is unaffected by spectral correlations, but the outage rate (considered in Section V) depends on the system bandwidth, relative to the channel coherence bandwidth. In order to focus on the effect of multiple antennas in wideband systems, we consider an idealized model for *temporal correlations*. Specifically, we assume that the channel does not vary with time over a codeword, but that the channel realization is independent from codeword to codeword. However, in related recent work [1] on wideband *uplink* space–time communication, we have considered noncoherent reception techniques for time-varying channels that exploit temporal correlations as well as spatial and spectral correlations.

Contributions: We find that covariance feedback is robust and "free" in a wideband system. As long as a mobile sends to the BS over multiple frequencies which are sufficiently spaced apart, information regarding the covariance can be obtained from uplink measurements without additional overhead. This method for generating implicit feedback relies on the fact that the space–time channels at different frequencies can be model-

Manuscript received December 18, 2003; revised July 26, 2004. This work was supported by Motorola, with matching funds from the University of California Industry–University Cooperative Research program, and by the National Science Foundation under Grants EIA–0080134 and ANI -0220118 (ITR).

G. Barriac was with the University of California at Santa Barbara, Santa, Barbara, CA 93106 USA. She is now with Qualcomm, San Diego, CA 92121 USA (e-mail: barriac@engineering.ucsb.edu).

U. Madhow is with the University of California at Santa Barbara, Santa Barbara, CA 93106 USA (e-mail: madhow@ece.ucsb.edu).

Communicated by R. Müller, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2004.838379

0018-9448/04$20.00 © 2004 IEEE

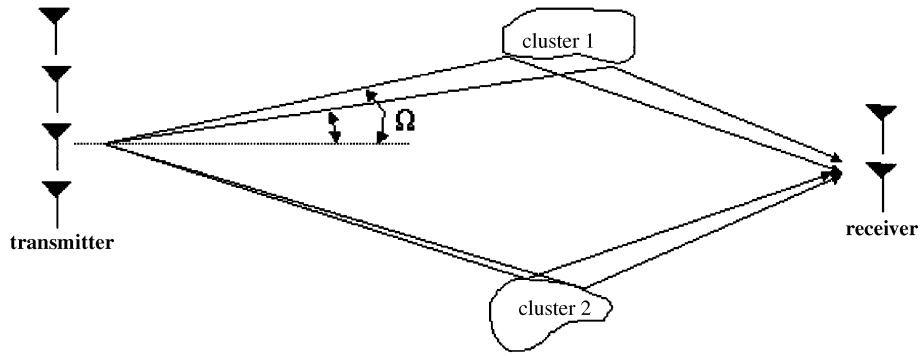


Fig. 1. System setup: the base station has N_T antennas, the mobile N_R.

ed as identically distributed random vectors (whose dependence becomes weaker as the frequency spacing increases). Thus, averaging over the frequencies employed by the mobile on the uplink provides a robust estimate of the covariance matrix which applies to both frequency-division duplex (FDD) and time-division duplex (TDD) systems. Also, the covariance matrix varies very slowly due to mobility, even if the channel itself exhibits fast fading.

We investigate the robustness and performance gains of covariance feedback using outdoor channel models employed in our prior work [2], [3] on wideband systems *without* feedback. These simple, yet accurate, channel models are distilled from the literature on propagation measurements, and have the advantage of yielding analytical insight not possible using typical channel simulators. Our numerical results show that estimates of the channel covariance matrix can be used to significantly increase the capacity relative to conventional space–time or space–frequency coding (which does not utilize feedback). We also show that these performance gains are robust to estimation errors. In other words, nearly all of the gains possible with perfect covariance information at the BS can be attained using an estimate of the covariance from uplink measurements. These gains come without the capacity and complexity costs required for explicit feedback. Explicit feedback would typically require the transmission of orthogonal training sequences from the BS transmit antennas, computation of the spatial correlations by the mobile, and the transmission of (possibly a summarized version) of the correlations from the mobile to the BS This requires overhead on both the downlink and uplink, as well as additional computational complexity at mobiles. On the other hand, compared to a conventional system with no feedback, our methods actually provide performance enhancement while reducing transceiver complexity, since the optimal transmit strategy is to send along the dominant channel eigenmodes (with the most extreme example being pure transmit beamforming).

Next, we provide guidelines for optimizing the transmit antenna spacing for systems with covariance feedback. Capacity changes with antenna spacing because of the corresponding changes in the channel eigenvalues. It is usually possible to space the antennas such that a specified number of eigenmodes are dominant (the less the spacing, the fewer the number). In order to get detailed theoretical insight, therefore, we consider a hypothetical system in which the covariance matrix has K equal, nonzero, eigenvalues, with the remaining eigenvalues set to zero. The goal of this thought experiment is to choose K $(1 \leq K \leq N_T)$ to maximize capacity. This in turn guides the choice of antenna spacing to obtain the appropriate number of dominant eigenmodes, and simulation results are provided to confirm that this strategy does indeed lead to significant gains in both the ergodic capacity and the outage rates. (The outage rate is the maximum rate the BS can send given a constraint on the probability of outage.) Note that it may not be possible to make the eigenvalues of the dominant eigenmodes equal, and in some cases, it may not be possible to attain the optimal number of eigenmodes. However, the results of the thought experiment still provide valuable guidance on shaping the channel covariance and the choice of optimal transmit strategy.

Roughly speaking, our theoretical investigation of the preceding thought experiment shows that, for $N_T \gg N_R$, the optimal number of dominant eigenmodes is matched to the number of receive elements. While there are a number of refinements and caveats to this statement (see Section IV for details), depending on the operating signal-to-noise ratio (SNR) and the values of N_T and N_R, the broad implication is that the optimal transmit strategy is to create a virtual $N_R \times N_R$ MIMO system by sending along N_R dominant eigenmodes. The transmitter's encoder is only required to do $N_R \times N_R$ MIMO processing, with the output of the encoder being sent to an $N_T \times N_R$ beamformer. The receiver is oblivious to the number of transmit elements N_T, and does standard $N_R \times N_R$ MIMO decoding. Thus, the number of transmit elements can be scaled up to provide better beamforming gains, without adversely impacting the encoder or decoder complexity.

Related Work The notion of using covariance information on the uplink to optimize downlink transmission has previously been applied in the context of FDD systems using time-division multiple access (TDMA) or direct-sequence code-division multiple access (DS-CDMA) [4]–[7]. Our work differs from this body of literature not only because of our information-theoretic perspective, but also because we rely on averaging uplink channel responses across *frequency*, whereas previous work relies on averaging uplink channel responses across *time*, or on estimates of the directions of arrival (DOA) of the incoming paths. DOA estimation is known to be computationally intensive [6], and may be infeasible if there are too many multipath components. On the other hand, time averaging has the disadvantage that the amount of time necessary to construct an accurate estimate of the covariance matrix may exceed the allotted uplink transmit time.

When the uplink and downlink bands are very far apart, the covariance feedback method proposed here does not apply directly, since the uplink and downlink covariance matrices may differ significantly. This is due to the disparity between uplink and downlink antenna array responses for a given DOA. We do not address this issue here, since we consider contiguous uplink and downlink bands where the difference between the uplink and downlink covariance matrices is negligible. However, we note that the issue of converting the uplink covariance matrix to the downlink covariance matrix has been addressed before in the literature: proposals include the use of a frequency calibration matrix [6], or of a clever antenna configuration [5] that attains identical beam patterns at both uplink/downlink wavelengths. Any of the proposed conversion methods can be directly applied to the OFDM-type cellular systems we consider.

Our work leverages the seminal work of Telatar [8], [9] on the Shannon theory of MIMO systems. More recent information-theoretic analyses of space–time communication with covariance feedback includes [10]–[16]. It is shown in [10] that, for a system with multiple transmit antennas and a single receive antenna, the optimal strategy given covariance feedback is to send along the eigenvectors of the covariance matrix. This result is extended to systems with multiple receive antennas in [12].The work in [13], [14], [16] considers the problem of optimal power allocation for a given channel covariance matrix. In particular, the ergodic capacity for a narrowband multiple-input single-output (MISO) system is shown in [13], [14] to be Schur-convex with respect to the correlation properties of the channel. While [13], [14], [16] do not consider shaping the channel correlation properties by varying the antenna spacing, these results can actually be used to provide alternative proofs of our finding that, for a MISO system, it is optimal to have one dominant channel eigenmode. However, there appears to be no literature that is relevant to our results of optimizing the channel eigenmode distribution for more general MIMO systems. Moreover, unlike our work, prior information-theoretic treatments of covariance feedback [10]–[14], [16] do not consider mechanisms for obtaining such feedback, and do not relate the channel covariance to physical parameters such as the channel propagation model and the antenna array geometry. Finally, while most of this prior work focuses on ergodic capacity for narrowband systems, we also investigate outage rates for wideband systems, exploiting a framework similar to that in our previous work on wideband systems which do not employ feedback [2], [3]. Note that the ergodic capacity of wideband OFDM systems without feedback was previously considered in Bolcskei *et al.* [17].

Our analysis suggests the use of $N_R \times N_R$ space–frequency codes in conjunction with a $N_T \times N_R$ beamforming matrix. Implicit in this, as well as in prior information-theoretic studies such as [10]–[12], is the combined use of space–time/frequency codes and "beamforming." Thus, the $N_R \times N_R$ space–frequency code can be chosen to attain any desired point on the diversity–multiplexing tradeoff curve described in [18]. Examples of codes optimized for spatial diversity are the classic space–time codes of Alamouti [19] and Tarokh *et al.* [20]. Examples of codes optimized for spatial multiplexing include the Bell Labs layered space–time (BLAST) system [21] and its variants [22].

Explicit constructions for space–time communication with feedback are considered in [23]–[25]. Giannakis *et al.* [23] propose a two-dimensional eigen-beamformer along with the well-known Alamouti space–time code [19], while Jongren *et al.* [25] modify more general orthogonal space–time block codes with $M_T \times M_T$ precoding matrices. The performance measure in [25] is the pairwise error probability between codewords, and that in [23] is the symbol error rate. While we do not consider specific code constructions in this work, we conjecture the following: a) the ergodic capacity promised by our information-theoretic analysis should be attainable by appropriate use of powerful turbo-like codes, coding across frequencies, and eigenmodes; b) for the wideband systems considered here, there is enough frequency diversity available that space–frequency codes with high multiplexing gains are expected to yield better performance than codes with high diversity gains.

Finally, we note that independent work on covariance feedback for OFDM systems has been reported recently by Vook *et al.* [26]. This work reports on simulation-based results for specific code constructions, unlike the information-theoretic analysis in the present paper.

Paper Layout The remainder of the paper is organized as follows. Section II describes the channel model, and the relevant background information. Section III describes how implicit covariance feedback is obtained from uplink measurements, verifies its robustness for FDD and TDD systems for typical outdoor channel models, and briefly discusses the geometry of the channel eigenmodes. In Section IV, we consider optimization of ergodic capacity by choice of antenna spacing, and obtain theoretical insight by investigation of the thought experiment mentioned previously. Section V discusses the choice of antenna spacing for optimization of outage rates. Channels with multiple clusters are discussed in Section VI, and concluding remarks are given in Section VII.

II. Preliminaries

In this section, we discuss the system model and the information-theoretic performance metrics employed in our work.

A. System Model

The system considered is as in Fig. 1. The BS is assumed to be far away from the mobile, and at high enough altitude that there is little to no local scattering around it. Thus, both uplink and downlink signals for a given mobile are restricted to a fairly narrow spatial cone, from the viewpoint of the BS antenna array. In contrast, the mobile is assumed to be in a rich scattering environment.

We assume a typical space–time channel model based on the superposition of specular rays, which are grouped into clusters as in the classic Saleh–Valenzuela model [27]. Each ray is parameterized by its delay, angle of departure (AOD), and amplitude. The channel impulse response can then be written as

$$\boldsymbol{h}(\tau,\Omega)=\sum_{m=1}^{M}\sum_{n=1}^{N_m}\alpha_{nm}e^{j\Theta_{nm}}\boldsymbol{a}(\Omega_{nm})\delta(\tau-\tau_{nm})\delta(\Omega-\Omega_{nm}) \quad (1)$$

where M is the number of clusters and N_m is the number of waves in the mth cluster. Here, Ω_{nm} denotes the angle of departure, τ_{nm} the delay, and α_{nm} the amplitude of the nth wave of the mth cluster. The phase shifts Θ_{nm} for these paths are modeled as i.i.d. uniform random variables over $[0, 2\pi]$. For our running example of a linear array, the BS array response as a function of angle of departure is given as follows:

$$\begin{aligned} \boldsymbol{a}(\Omega) &= [a_1 \dots a_l \dots a_{N_T}]^T, \\ a_l(\Omega) &= e^{j(l-1)2\pi\frac{d}{\lambda}sin(\Omega)}, \qquad l = 1, \dots, N_T \end{aligned} \tag{2}$$

where d is the antenna array spacing, and λ the carrier wavelength.

Equation (1) is the BS channel response from a particular antenna at the mobile. It is assumed that there is enough scattering at the mobile so that the BS responses from different mobile antennas are i.i.d.

Experimental measurements of outdoor channels [28]–[30] indicate that the number of clusters is small, usually one or two, and that the power delay profile (PDP) and power angle profile (PAP) for each cluster as seen by the BS can be modeled as exponential and Laplacian, respectively. For propagation environments with moderate scattering, both single- and double-cluster channels are common. We first restrict our attention to one-cluster channels, postponing discussion of multicluster channels to Section VI. The power profiles we use throughout this paper are given as follows:

$$P_\tau(\tau') = \frac{1}{\tau_{rms}} e^{-\frac{\tau'-\tau_e}{\tau_{rms}}}, \qquad \text{for } \tau' \geq \tau_e \tag{3}$$

$$P_\Omega(\Omega') = \frac{1}{2\Omega_{\text{spread}}} e^{-\left|\frac{\Omega'-\overline{\Omega}}{\Omega_{\text{spread}}}\right|}. \tag{4}$$

The notation $L(a, b)$ will be used to refer to a Laplacian profile with $\overline{\Omega} = a$ and $\Omega_{\text{spread}} = b$.

In [3], we showed that the discrete ray channel model can be substituted by a vector tap delay line (TDL) model with complex Gaussian taps without any loss of generality. This model comes about because paths which are spaced less than $\frac{1}{W}$ apart (where W is the bandwidth), cannot be distinguished and hence sum together to form taps. Since the paths are independent, with uniformly distributed phases and zero mean complex Gaussian amplitudes, their sum is complex Gaussian (as long as there are enough paths for the central limit theorem to apply). Thus, the TDL model is given as [3]

$$\boldsymbol{h}_W(\tau) = \sum_{l=0}^{\infty} A_l \boldsymbol{v}_l \delta\left(\tau - \frac{l}{W}\right) \tag{5}$$

where the tap weights A_l are proportional to the square root of the power delay profile

$$A_l \propto \sqrt{P_\tau\left(\frac{l}{W}\right)} \tag{6}$$

and are normalized such that

$$\sum_{l=0}^{\infty} A_l^2 = 1. \tag{7}$$

The vectors $\boldsymbol{v}_l$ are i.i.d. zero-mean complex normal random vectors

$$\boldsymbol{v}_l \sim CN(0, \boldsymbol{C}) \tag{8}$$

where $\boldsymbol{C}$ is the expected value of the outer product of the antenna array response, where the expectation is taken over the power angle profile

$$\boldsymbol{C} = E[\boldsymbol{a}(\Omega')\boldsymbol{a}(\Omega')^H] = \int_{-\pi}^{\pi} \boldsymbol{a}(\Omega')\boldsymbol{a}(\Omega')^H P_\Omega(\Omega') d\Omega'. \tag{9}$$

Now, consider an OFDM system with N_T transmit antennas, N_R receive antennas, and N frequency bins. One OFDM symbol consists of N symbol vectors $\boldsymbol{s}_i$, $i = 1 \dots N$ of length N_T, each transmitted at a different frequency. Assuming negligible intercarrier interference (ICI), we can write

$$\hat{\boldsymbol{s}}_i = \boldsymbol{H}_i \boldsymbol{s}_i + \boldsymbol{n}_i \tag{10}$$

where $\hat{\boldsymbol{s}}_i$ is the received data vector for the ith tone, $\boldsymbol{H}_i$ is the $N_R \times N_T$ channel frequency response at the ith tone, and $\boldsymbol{n}_i$ is additive white Gaussian noise satisfying

$$E\left[\boldsymbol{n}_i \boldsymbol{n}_j^H\right] = \sigma_n^2 \boldsymbol{I}_{N_R} \delta(i-j).$$

It follows [2] from the statistics of the channel model that the rows of the channel matrix for different frequencies (denoted $\boldsymbol{H}_i(k,:)$) are well modeled as identically distributed $1 \times N_T$ proper complex Gaussian random vectors with zero mean and covariance matrix $\boldsymbol{C}$

$$\boldsymbol{H}_i(k,:) \sim CN(0, \boldsymbol{C}), \qquad i = 1 \dots N, \ k = 1 \dots N_R. \tag{11}$$

Because the channel responses are identically distributed, if the BS measures the channel from uplink measurements spaced much further apart than the coherence bandwidth, it can reconstruct an estimate of $\boldsymbol{C}$. This method of obtaining implicit covariance information is discussed in Section III.

B. Ergodic Capacity and Outage Rates

Throughout this paper, we use either ergodic capacity or outage rates as a measure of performance. We now summarize the relevant definitions relating to these performance metrics.

The spectral efficiency of an OFDM system, denoted I_N, is given as

$$I_N = \frac{1}{N} \sum_{i=1}^{N} \log\left(\left|\boldsymbol{I}_{N_R} + \frac{1}{\sigma_n^2} \boldsymbol{H}_i \boldsymbol{Q} \boldsymbol{H}_i^H\right|\right) \tag{12}$$

where $\boldsymbol{s}_i \sim CN(0, \boldsymbol{Q})$, $\boldsymbol{I}_{N_R}$ denotes the $N_R \times N_R$ identity matrix, and the trace of $\boldsymbol{Q}$ is constrained to be less than the total transmit power P. The distribution of $\boldsymbol{s}_i$ is assumed to be the same for all i. That is, we constrain the input symbol vectors $\boldsymbol{s}_i$ to be i.i.d. across frequency.

The outage rate $R(\epsilon)$ is defined as the largest transmission rate R such that the following condition holds:

$$P[I_N < R] \neq \epsilon. \tag{13}$$

In other words, "outage" occurs when the spectral efficiency falls below the transmission rate, and sending at rate $R(\epsilon)$ ensures that the probability of outage is below the specified level ϵ.

Ergodic capacity is the maximum average spectral efficiency achievable, as follows:

$$C = \max_{Q:\mathrm{trace}(Q)=P} E\left[\frac{1}{N}\sum_{i=1}^{N}\log\left(\left|\boldsymbol{I}_{N_R} + \frac{1}{\sigma_n^2}\boldsymbol{H}_i\boldsymbol{Q}\boldsymbol{H}_i^H\right|\right)\right] \tag{14}$$

where $\boldsymbol{Q}$ is the input covariance matrix (and is therefore constrained to positive semi-definite). When the transmitter knows the channel covariance $\boldsymbol{C}$, the optimal $\boldsymbol{Q}$ [12], [10] is given by

$$\boldsymbol{Q}_0 = \boldsymbol{U}_C\boldsymbol{\Lambda}_0\boldsymbol{U}_C^H \tag{15}$$

where $\boldsymbol{U}_C$ is the unitary eigenvector matrix of $\boldsymbol{C}$, i.e.,

$$\boldsymbol{C} = \boldsymbol{U}_C\boldsymbol{\Lambda}_C\boldsymbol{U}_C^H. \tag{16}$$

The optimal transmit strategy is thus to send along the eigenmodes of the channel covariance matrix $\boldsymbol{C}$. The optimal power to send along each eigenmode (the total power is constrained to be less than P) can be found numerically. Note that outage rates are defined as a function of the transmit covariance matrix $\boldsymbol{Q}$, whereas ergodic capacity is defined as a maximization over all possible $\boldsymbol{Q}$.

III. Implicit Covariance Feedback

In this section, we show how covariance information can be obtained from uplink measurements for both FDD and TDD systems. For the large delay spreads typical of outdoor environments, the coherence bandwidth is small, and the correlation between the channel responses at different frequencies dies out quickly with their separation. Thus, by measuring the channel over a rich enough set of frequencies on the uplink, the BS can accurately estimate the channel covariance matrix $\boldsymbol{C}$, and employ this knowledge on the downlink, regardless of whether the system is TDD or FDD. We term this concept *statistical reciprocity*, to distinguish it from deterministic reciprocity, which states that the channel response at a given frequency and time is the same in both directions. Statistical reciprocity leads to implicit covariance feedback which is robust under a wide variety of conditions.

We assume, for simplicity, that the mobiles have only one antenna. (Since the responses from the BS array to different antenna elements at the mobile are modeled as i.i.d., more mobile antennas would provide even more averaging when estimating the covariance matrix on the uplink.)

The BS estimates the covariance matrix for a particular user as follows:

$$\hat{\boldsymbol{C}}_k = \frac{1}{|J_k|}\sum_{i\in J_k}\hat{\boldsymbol{H}}_i^H\hat{\boldsymbol{H}}_i \tag{17}$$

where J_k is the set of subcarriers the kth mobile uses on its uplink, and $\hat{\boldsymbol{H}}_i$ is the base station's estimate of $\boldsymbol{H}_i$ obtained from the kth user's uplink measurement at frequency f_i.

There are two key issues regarding the efficacy of estimating the covariance matrix $\boldsymbol{C}$ in this way.

i) The covariance must vary slowly enough such that, when the BS sends to user k, the estimate $\hat{\boldsymbol{C}}_k$ is still valid.
ii) There must be enough subcarriers in J_k for $\hat{\boldsymbol{C}}_k$ to well approximate $\boldsymbol{C}_k$.

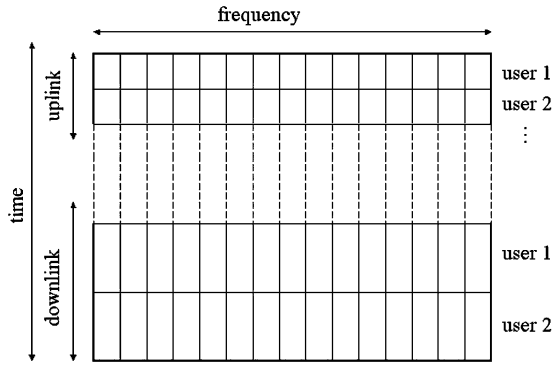


Fig. 2. A TDD system with TDMA on the uplink and TDM on the downlink.

Both of these conditions are met for a wide variety of resource-sharing models. For illustrative purposes, we focus on two extreme examples. The first is a TDD system with TDMA on the uplink and TDM on the downlink. This particular setup requires the longest time for which feedback from the mobiles needs to remain accurate. The second example is an FDD system with FDM on the uplink and TDM on the downlink—a scenario which allows each user a limited amount of uplink frequency bins and therefore tests the limits of condition ii).

In both examples, we consider an OFDM system with 1024 subcarriers spaced 25 kHz apart. The PAP is initially $L(0°, 5°)$($L(M°, \alpha°)$ denotes a Laplacian distribution with mean M and variance $2\alpha^2$) and the PDP is exponential with a root mean square (rms) value of 0.5 μs. SNR is set to 10 dB. The BS has six antennas, with a typical antenna spacing of $d/\lambda = 0.5$. At this spacing, beamforming is the optimal transmit strategy for the given PAP.

A. Example 1: TDD System

A TDD system with TDMA/TDM on the uplink/downlink is shown in Fig. 2. Each user sends to the BS using the entire frequency band for a certain amount of time, and subsequently the BS takes turns sending to the mobiles over the whole band. For such a system, J_k in (17) equals the entire set of frequency bins for all k. If the bandwidth is large, $\hat{\boldsymbol{C}}_k$ is clearly a good approximation for $\boldsymbol{C}_k$, but the question remains as to whether this covariance will remain valid until the BS is ready to reply to that mobile on the downlink. The longest a user will have to wait until it hears back from the BS is approximately the number of users in the system multiplied by the time the BS sends to each user. For a rate of 20 Mb/s and 10 packet payloads of 10 000 bits each, the time the BS sends to each mobile is approximately 5 ms. If there are 10 users, this means the total delay is around 50 ms. However, even if the channel is fast fading, the covariance need not change much in this length of time, since it depends only on the power-angle profile, which, in general, is slowly varying. It is shown in [31] that for a mobile 500 m from the BS traveling less than 1000 km/h, and a BS station with eight antennas spaced half a wavelength apart, the channel statistics can be considered stationary for around 100 ms. Thus, the PAP, and hence the covariance would also be stationary for that time interval.

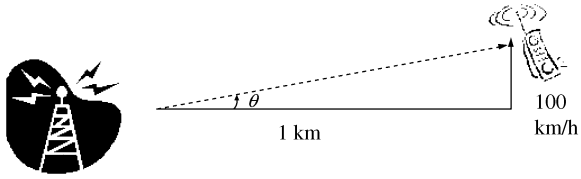


Fig. 3. For a mobile 1 km away moving directly up at 100 km/h, θ changes 0.08° in 50 ms.

TABLE I
THE ERGODIC CAPACITY (C) AND 1% OUTAGE RATE (R_o) IN B/S/HZ WHEN THE BS BEAMFORMS TO THE DOMINANT EIGENMODE OF $\hat{\boldsymbol{C}}_k$, COMPUTED FOR $\Omega \sim L(0°, 5°)$. RESULTS ARE ALSO SHOWN FOR THE CASE WHERE NO FEEDBACK IS AVAILABLE, AND THE BS THUS SENDS EQUAL POWER IN ORTHOGONAL DIRECTIONS

Actual PAP	Feedback PAP	C	R_o
$\Omega \sim L(0.0^o, 5^o)$	no feedback	3.12	2.70
$\Omega \sim L(0.0^o, 5^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.83	4.13
$\Omega \sim L(0.6^o, 5^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.83	4.14
$\Omega \sim L(2.9^o, 5^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.76	4.09
$\Omega \sim L(0.0^o, 9^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.83	4.13
$\Omega \sim L(0.0^o, 1^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.82	4.13
$\Omega \sim L(2.9^o, 9^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.77	4.10
$\Omega \sim L(2.9^o, 1^o)$	$\Omega \sim L(0.0^o, 5^o)$	4.76	4.08

We now consider how variations (we can assume they are small) in the PAP would affect system performance. For a mobile moving away from the BS at 100 km/h as pictured in Fig. 3, the angle θ will change approximately 0.08° in 50 ms. If the center angle of the PAP changes a corresponding amount, we would like to know how this impacts performance results. Table I gives the 1% outage rate and ergodic capacity of a wideband system when the actual PAP differs from the PAP used to estimate the covariance. The outage rates are computed using the transmit strategy which maximizes ergodic capacity. It is assumed that the power angle profile remains Laplacian, and that only the mean and/or angular spread change with time. The first row shows the capacity and outage rate when there is no feedback and the transmitter employs a full blown space–time code (the optimal transmit strategy when no feedback is available). The second row shows the capacity and outage rate when the BS has perfect covariance feedback information and beamforms in the direction of the covariance's dominant eigenmode (beamforming is the optimal strategy in this scenario for the given parameters). The following rows display the resulting capacity when the BS beamforms using imperfect covariance information. It can be seen that even if the BS uses covariance information obtained from a Laplacian whose mean has since shifted 2.9°, and whose variance has doubled, deleterious effects on performance are minimal. Even in this case, where the changes in the PAP are much larger than one might expect, both the ergodic capacity and outage rate are much higher than the corresponding quantities when no feedback is available.

B. Example 2: FDD System

We next consider an FDD system that uses FDM on the uplink and TDM on the downlink. Since the mobiles send using only part of the frequency spectrum, it is possible that the BS may not be able to get a good estimate of the channel covariance matrix. Table II shows how the mobile's uplink bandwidth affects performance. Simulation parameters are the same as in the last example. We assume that the mobile transmits using only a fixed set of frequency bins which are equally spaced throughout the uplink spectrum so as to minimize the correlations between the channel responses. Half of the available spectrum (the left half) is reserved for the downlink. The BS forms $\hat{\boldsymbol{C}}_k$ from the feedback information in one OFDM uplink symbol and beamforms in the direction of the dominant eigenmode. The statistics of $\boldsymbol{C}$ are kept stationary for simplicity.

TABLE II
ERGODIC CAPACITY AND 1% OUTAGE RATE IN B/S/HZ WHEN THE BS BEAMFORMS TO THE DOMINANT EIGENMODE OF $\hat{\boldsymbol{C}}_k$ COMPUTED USING UPLINK MEASUREMENTS FROM THE SPECIFIED NUMBER OF FREQUENCY BINS

# subcarriers on user's uplink	C	R_o
no feedback	3.12	2.69
500	4.80	3.87
100	4.80	3.87
50	4.80	3.87
20	4.80	3.88
10	4.78	3.82

As can be seen in Table II, multiplexing mobiles onto different subcarriers causes a negligible decrease in capacity. For instance, the difference between one mobile using the 500 subcarriers on the uplink, and 50 mobiles sharing this spectrum is merely a 1% loss in outage rate. Thus, the system can support at least 50 users without incurring any performance degradation.

Using FDD instead of TDD reduces the time delay between the uplink and downlink of a particular user, and therefore may be more desirable if the speed at which the channel statistics vary is of concern. It is clear that covariance feedback information from the mobile can be successfully exploited in both cases, and that the system design can be tailored for specific scenarios.

C. The Geometry of Covariance Feedback

In this subsection, we consider the physical interpretation of the channel covariance and its relationship to both the power angle profile and the antenna spacing. Specifically, we are interested in how the N_T-dimensional eigenvectors relate to physical two-dimensional directions, and in turn, how the eigenvalues relate to these physical directions. By plotting the correlation between a particular eigenvector and the BS antenna response as a function of Ω, we get a visual representation of the antenna pattern created when the BS transmits along that eigenmode. Such graphs, which we dub eigen-patterns, illustrate the two-dimensional "eigenmode directions." The degree of overlap between these "eigenmode directions" and the PAP is reflected in the eigenvalues, with greater overlap resulting in larger eigenvalues. We find the following general trends.

i) For a given antenna spacing, if the PAP is Laplacian with fixed mean and variable covariance, increasing the angular spread (assuming it remains less than 90°), causes little change in the eigen-patterns, though the eigenvalues can change considerably depending on the overlap between the lobes of the eigen-pattern and the PAP.

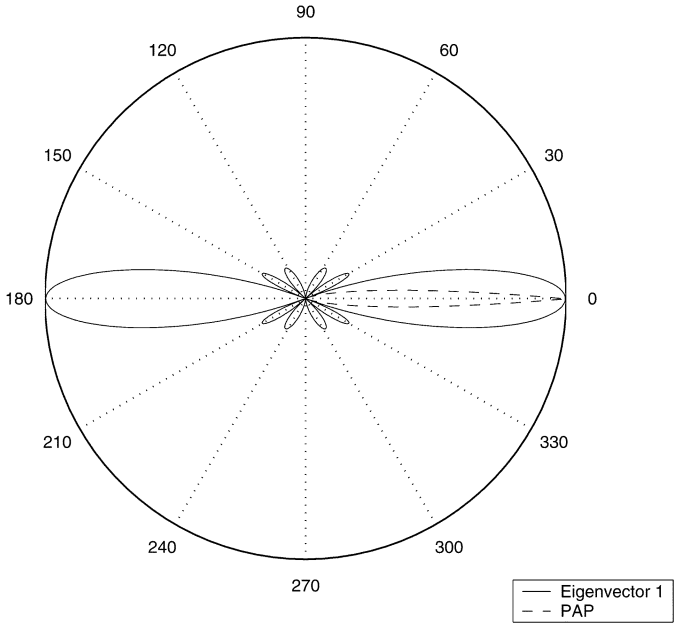


Fig. 4. The correlation ρ_1 between the first (dominant) eigenvector and the normalized array response as a function of Ω. $N_T = 6$, $d/\lambda = 0.5$, and $\Omega \sim L(0^\circ, 5^\circ)$.

ii) For a given Laplacian PAP, decreasing the antenna spacing causes the channel power to be concentrated in fewer eigenvalues.

We now illustrate these trends with some examples.

In the two examples of the previous section (where $N_T = 6$, $N_R = 1$, $d/\lambda = 0.5$, and $\Omega \sim L(0^\circ, 5^\circ)$), there is a single dominant eigenmode and beamforming along this eigen-direction is the optimal transmit strategy. (The dominant eigenmode is simply the eigenvector of the channel covariance matrix that corresponds to the largest eigenvalue. We assume that the eigenvectors are ordered according to the magnitude of their eigenvalues, so that the first eigenvector has the largest eigenvalue and the last eigenvector has the smallest.) Fig. 4 shows the correlation between the array response $\boldsymbol{a}(\Omega)$ and the dominant eigenvector of $\boldsymbol{C}$ as a function of Ω (plotted as a solid line). More formally, the correlation ρ_j is defined as

$$\rho_j \equiv \left\langle \frac{\boldsymbol{a}(\Omega)}{||\boldsymbol{a}(\Omega)||}, \boldsymbol{e}_j \right\rangle \tag{18}$$

where $\boldsymbol{e}_j$ is the jth eigenvector. This eigen-pattern represents the antenna pattern if the transmitter were to beamform along the direction of the first eigenmode. The PAP, normalized so the maximum value is equal to one, is plotted as a dashed line in Fig. 4. It can be seen that the dominant eigenmode can project power along all incoming directions from the mobile. It is also clear that the dominant eigenvector lies approximately in the array manifold, corresponding to the array response at $\Omega = 0^\circ$. Thus, beamforming along the dominant eigenvector sends power in the expected direction of the mobile.

Fig. 5(a) shows all three eigen-patterns for a system where $N_T = 3$, $d/\lambda = 0.5$, and the PAP is $L(0^\circ, 5^\circ)$, as above (shown in Fig. 4). The eigenvalues for this system are $\lambda_1 = 2.75$, $\lambda_2 = 0.23$, and $\lambda_3 = 0.01$. Thus, almost all the channel power is concentrated in the first eigenmode. This is reasonable since there is little overlap between the PAP and the eigen-patterns for the second and third eigenmodes. For a fixed PAP, changing the antenna spacing changes both the eigenvalues and the eigenvectors of $\boldsymbol{C}$. Fig. 5(b) shows the three eigen-patterns for the same parameters used in Fig. 5(a), except that $d/\lambda = 1$. The eigenvalues are now $\lambda_1 = 2.33$, $\lambda_2 = 0.55$, and $\lambda_3 = 0.11$. When $d/\lambda = 0.5$, all the channel energy is concentrated in the first eigenmode, whereas when $d/\lambda = 1$, some of the energy has shifted into the second eigenmode. As expected, the second eigenmode's lobes overlap more with the PAP when $d/\lambda = 1$.

The eigen-patterns remain essentially unchanged for fixed d/λ as the angular spread of the Laplacian power angle profile increases (assuming it remains within 90°), as can be seen in Fig. 6. However, the corresponding eigenvalue distribution becomes less concentrated in the dominant eigenvalue since more overlap occurs between the PAP and the second and third eigenmodes.

The more channel energy there is in a particular eigenmode, the more power the mobile will receive if the transmitter sends along that direction. Thus, the optimal transmit strategy is highly contingent upon the eigenvalue distribution and hence the antenna spacing. In Section IV, we show how smaller antenna spacing can lead to both simplified transmit strategies and increased capacity.

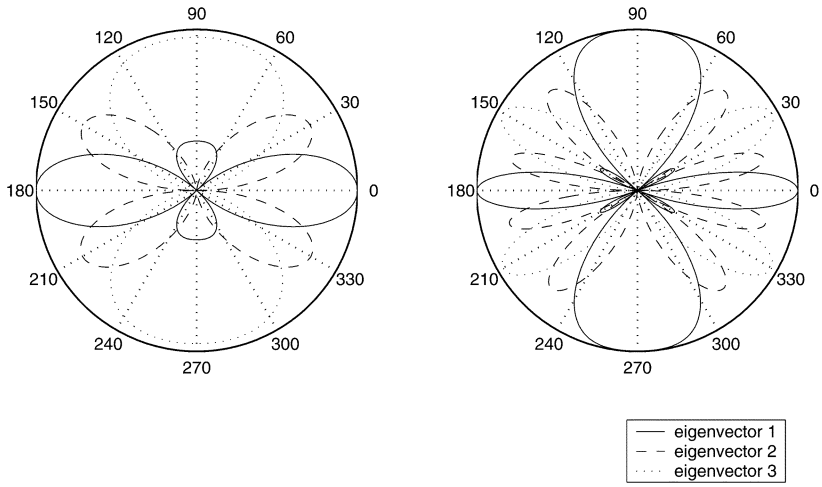
IV. Optimal Antenna Spacing for Ergodic Capacity

When there is no feedback, a reasonable strategy is to send i.i.d. Gaussian input from each transmit antenna. The mutual information thus attained is maximized when the spatial covariance matrix is white. That is, with no feedback, the best performance is attained by spacing the antennas far enough apart that they see uncorrelated responses [3]. However, when the BS knows the channel covariance, the optimal antenna spacing can be much smaller. In this subsection, we focus on finding the antenna spacing that maximizes the ergodic capacity of a single frequency bin, given that accurate covariance feedback is available. This spacing also maximizes the ergodic capacity of the wideband channel, since the covariance statistics are frequency invariant. Furthermore, as shown in Section V, the antenna spacing which maximizes ergodic capacity approximately maximizes outage rates as well.

The ergodic capacity when the transmitter knows the channel covariance, (see (14)), can be rewritten as [12]

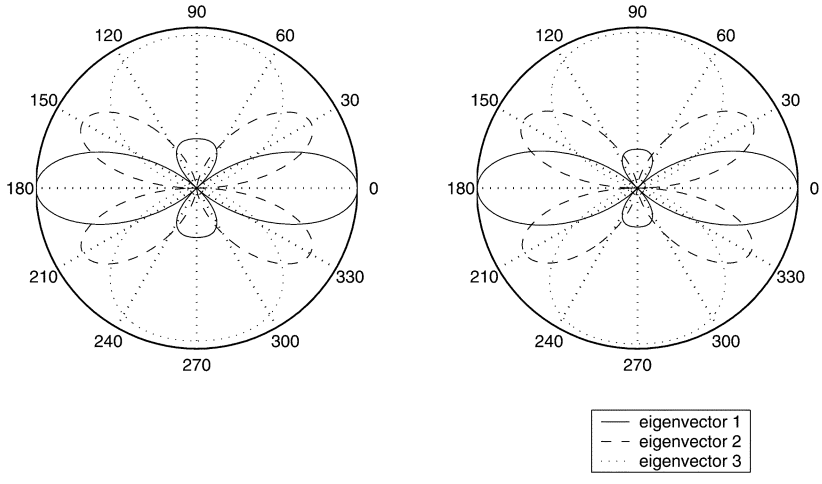
$$C = \max_{p_i : \sum_{i=1}^{N_T} p_i = P, p_i \geq 0} E\left[\log\left|\boldsymbol{I}_{N_R} + \frac{1}{\sigma_n^2}\sum_{i=1}^{N_T} \boldsymbol{z}_i \boldsymbol{z}_i^H p_i \lambda_i\right|\right] \tag{19}$$

where the $\boldsymbol{z}_i$ are independent $N_R \times 1$ vectors whose elements are i.i.d. $CN(0,1)$. The λ_i are the eigenvalues of the channel covariance matrix $\boldsymbol{C}$ (the diagonal elements of $\boldsymbol{\Lambda}_C$ in (16)), and the p_i are the powers the transmitter sends along the corresponding channel eigenmodes. (Recall that sending in the directions of the channel eigenmodes is optimal when covariance feedback is available.) In the following, we set $\sigma_n^2 = 1$ without loss of generality, so that capacity is a function of the power P and the channel eigenvalues $\{\lambda_i\}$, normalized such that $\sum_{i=1}^{N_T} \lambda_i = N_T$.



(a) (b)

Fig. 5. The eigen-patterns for all three eigenvectors ($N_T = 3$). In (a), $d/\lambda = 0.5$, and in (b), $d/\lambda = 1$.



(a) (b)

Fig. 6. The eigen-patterns for all three eigenvectors ($N_T = 3$) where $d/\lambda = 0.5$ In (a), $\Omega \sim L(0°, 5°)$ and in (b), $\Omega \sim L(0°, 15°)$.

Changing the antenna spacing changes the eigenvalues λ_i in (19), and hence the capacity. In order to find the optimal spacing, we would first like to understand how the eigenvalue distribution affects performance. To this end, we consider a hypothetical system in which we can freely manipulate the channel eigenvalues. In practice, it is not possible to exactly match the channel eigenvalues to the optimal eigenvalues obtained in the hypothetical system. For instance, suppose that it is optimal in the hypothetical system to have exactly two equal, nonzero, eigenvalues. This is difficult to achieve (at least with a moderate number of antenna elements) if the physical channel consists of a single cluster with a narrow PAP. Each eigenmode has fairly broad lobes (see Section III-C), and the main lobes for different eigenmodes (which must be orthogonal) have limited overlap. Thus, it is difficult for two different eigenmodes to have substantial overlap with the spatial channel profile, so that there is typically one eigenvalue that is larger than the rest. However, it is still usually possible to manipulate the antenna spacing such that the number of dominant eigenmodes in the physical system approximately matches the number prescribed by optimizing the hypothetical system. Thus, the optimal eigenvalue distribution for the hypothetical system is a valuable guide for optimizing the physical antenna spacing.

A good rule of thumb is to space the antennas such that N_R eigenmodes are dominant. In the ensuing subsections, we justify this statement, while providing a number of caveats.

A. A Thought Experiment

We now consider how to maximize capacity when the channel eigenvalues are parameters under our control. An upper bound on the capacity, denoted C_{ub}, can be found by maximizing the ergodic capacity C (19) over the set of possible channel eigenvalues:

$$C_{ub} = \max_{p_i, \lambda_i} E\left[\log\left|\boldsymbol{I}_{N_R} + \sum_{i=1}^{N_T} \boldsymbol{z}_i \boldsymbol{z}_i^H p_i \lambda_i\right|\right] \tag{20}$$

subject to

$$\sum_{i=1}^{N_T} p_i = P, \quad \sum_{i=1}^{N_T} \lambda_i = N_T, \qquad \lambda_i, p_i \geq 01 \leq i \leq N_T. \tag{21}$$

Using standard Lagrange multiplier theory, it can be shown that $\lambda_{i,\text{opt}} \propto p_{i,\text{opt}} \; \forall \; i$, where the subscript *opt* stands for optimal. Invoking the constraints that $p_i : \sum_{i=1}^{N_T} p_i = P$, and $\lambda_i : \sum_{i=1}^{N_T} \lambda_i = N_T$, we get that

$$\lambda_{i,\text{opt}} = \kappa p_{i,\text{opt}}, \quad \forall \, i \tag{22}$$

where $\kappa = \frac{N_T}{P}$. We can then rewrite (20) as follows:

$$C_{ub} = \max_{p_i : \sum_{i=1}^{N_T} p_i = P,\, p_i \geq 0} E\left[\log\left|\boldsymbol{I}_{N_R} + \kappa \sum_{i=1}^{N_T} \boldsymbol{z}_i \boldsymbol{z}_i^H p_i^2\right|\right]. \tag{23}$$

Letting $\boldsymbol{p}$ denote the length N_T vector of the p_i, we can write

$$\boldsymbol{p}_{\text{opt}} \equiv \arg \max_{p_i : \sum_{i=1}^{N_T} p_i = P,\, p_i \geq 0} C_{ub}(\boldsymbol{p}). \tag{24}$$

The $\{p_{i,\text{opt}}\}$ and $\{\lambda_{i,\text{opt}}\}$ can then be numerically calculated using (23), (24), and (22).

We first consider a system with two transmit antennas and two receive antennas, where $P = 10$. Numerical optimization shows that $p_{1,\text{opt}} = p_{2,\text{opt}} = 5 = P/2$, and $\lambda_{1,\text{opt}} = \lambda_{2,\text{opt}} = 1$. Hence, in this instance, ergodic capacity is maximized when the channel eigenvalues are equal, and hence an antenna spacing large enough to ensure uncorrelated responses is necessary for optimal performance.

When there are three transmit antennas and two receive antennas (keeping $P = 10$), $p_{1,\text{opt}} = p_{2,\text{opt}} = \frac{P}{2}$ and $p_{3,\text{opt}} = 0$. This means $\lambda_{1,\text{opt}} = \lambda_{2,\text{opt}} = \frac{3}{2}$ and $\lambda_{3,\text{opt}} = 0$, indicating that in the best possible scenario, maximum capacity is achieved by transmitting along only two of the available three eigenmodes.

For these particular values of N_R, N_T, and P, matching the number of dominant eigenmodes to the number of receive antennas is optimal. However, we would like to know whether "matching" is also optimal for a more general set of parameters. Numerical optimizations like the latter were done by recursively quantizing the search space and performing an exhaustive search for the maximum. Such procedures become prohibitively complex for larger values of N_T and N_R, and standard convex optimization techniques cannot be used since the right-hand side (RHS) of (23) is not, in general, concave (or convex). (The presence of the expectation in (23) also makes numerical optimization difficult.) In addition, we would like to have analytical results on the optimality of "matching" the number of channel eigemodes to the number of receive antennas. We therefore consider the following simplification which allows us to derive useful analytical results.

B. Simplified Thought Experiment

Now, let us assume that there are exactly K equal nonzero eigenmodes ($K \leq N_T$), so that each nonzero eigenvalue equals N_T/K. In this case, dividing the available power equally among the eigenmodes achieves capacity, which can then be written as

$$C(K) = E\left[\log\left|\boldsymbol{I}_{N_R} + \frac{PN_T}{K^2} \sum_{i=1}^{K} \boldsymbol{z}_i \boldsymbol{z}_i^H\right|\right]. \tag{25}$$

We define the optimal K as follows:

$$K_{\text{opt}} = \arg \max_K C(K). \tag{26}$$

Note that K_{opt} is a function of P, N_T, and N_R. Its dependence on one or more of these parameters will be made explicit in the notation as needed. Prior to a formal statement of our results, we highlight some key conclusions.

Low SNR regime: At low SNR, $K_{\text{opt}} = 1$ regardless of N_T and N_R. Thus, the channel should be shaped such that all the energy is focused along a single eigenmode.

High SNR regime: At high SNR, K_{opt} is of the order of N_R. That is, the channel should be shaped such that the number of dominant eigenmodes matches the number of receive elements.

One or two receive antennas: This covers most downlink scenarios of practical interest. For $N_R = 1$, we have $K_{\text{opt}} = 1$ regardless of SNR. For $N_R = 2$, $K_{\text{opt}} = 2$ except at low SNR, where $K_{\text{opt}} = 1$.

Since we consider a hypothetical system with K equal eigenmodes, our mathematical setting is analogous to that of Telatar's information-theoretic analysis of a MIMO system with independently fading paths [8]. We therefore summarize some results from [8], using our notation, before proceeding further.

Define the $N_R \times N_R$ matrix

$$\boldsymbol{W}_K \equiv \sum_{i=1}^{K} \boldsymbol{z}_i \boldsymbol{z}_i^H \tag{27}$$

and denote its eigenvalues by $\nu_{K,i}$ $i = 1 \ldots N_T$. Denote a typical eigenvalue by ν_K. The distribution of ν_K is given by

$$f_{\nu_K}(\nu) = \frac{1}{N_R} \sum_{l=0}^{N_R - 1} \frac{l!}{l + K - N_R!} \left[L_l^{K - N_R}(\nu)\right]^2 \nu^{K - N_R} e^{-\nu} \tag{28}$$

where L_l^k is the associated Laguerre polynomial of order l.

Theorem 1: At low SNR, $K_{\text{opt}} = 1$.

Proof: Equation (25) can be written as

$$C(K) = E\left[\log\left(\prod_{i=1}^{N_R}\left(1 + \frac{PN_T}{K^2}\nu_{K,i}\right)\right)\right] \tag{29}$$

where the $\nu_{K,i}$ $i = 1 \ldots N_T$ are the eigenvalues of $\boldsymbol{W}_K$ (27). Thus, we have

$$C(K) = E\left[\sum_{i=1}^{N_R} log\left(1 + \frac{PN_T}{K^2}\nu_{K,i}\right)\right] \tag{30}$$

$$\approx E\left[\sum_{i=1}^{N_R} \frac{PN_T}{K^2}\nu_{K,i}\right] \tag{31}$$

$$= E\left[\frac{PN_T}{K^2}\text{trace}(\boldsymbol{W}_K)\right] \tag{32}$$

$$= \frac{PN_T}{K^2} N_R K \tag{33}$$

$$= \frac{PN_T N_R}{K} \tag{34}$$

where (31) holds for small P. From (31), it is clear that $C(K)$ is maximized for $K = 1$. □.

Note that this theorem also follows from the work of [16], [11], which shows that at low SNR, beamforming is optimal.

We now consider the high SNR case. When P is large, i.e., $\frac{P}{N_T} \gg 1$, we can approximate (25) as [32]

$$C(K) \approx \begin{cases} E\left[\log\left|\frac{PN_T}{K^2}\sum_{i=1}^{K} \boldsymbol{z}_i \boldsymbol{z}_i^H\right|\right], & K \geq N_R \\ E\left[\log\left|\frac{PN_T}{K^2}\sum_{i=1}^{N_R} \tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^H\right|\right], & K < N_R \end{cases} \tag{35}$$

where the $\tilde{\boldsymbol{z}}_i$ are independent $K \times 1$ vectors whose elements are i.i.d. $CN(0,1)$. (The $\boldsymbol{z}_i$ are independent $N_R \times 1$ vectors whose elements are i.i.d. $CN(0,1)$.)

Theorem 2: For high SNR, when $N_R = 1$, $K_{\text{opt}} = 1$.

Proof: At high SNR, when $N_R = 1$, $C(K)$ is approximated by the top equation in (35). When $N_R = 1$, the $\boldsymbol{z}_i$ in (35) are scalar complex Gaussian random variables with unit power, so that $w_i = \boldsymbol{z}_i \boldsymbol{z}_i^H = |\boldsymbol{z}_i|^2$ are i.i.d. exponential random variables with mean one. We can therefore rewrite (35) to read

$$C(K) \approx E\left[\log\left(\frac{PN_T}{K^2}\sum_{i=1}^{K} w_i\right)\right] \tag{36}$$

$$= \log(PN_T) + f(K) \tag{37}$$

where

$$f(K) = E\left[\log\left(\frac{1}{K^2}\sum_{i=1}^{K} w_i\right)\right]. \tag{38}$$

We would like to show that $f(K)$, and hence $C(K)$ at high SNR, is maximized at $K = 1$. Using Jensen's inequality

$$f(K) \leq \log\left(\frac{1}{K^2}\sum_{i=1}^{K} E[w_i]\right) = \log\left(\frac{1}{K}\right). \tag{39}$$

For $K \geq 2$

$$f(K) \leq \log\left(\frac{1}{K}\right) \leq -\log(2) = -0.69. \tag{40}$$

However,

$$f(1) = E[\log(w_1)] = -\gamma \tag{41}$$

where $\gamma \approx 0.5772$ is Euler's constant. This shows that

$$f(1) \geq f(K) \quad K \geq 2. \tag{42}$$

□

A similar result, which was derived independently, can be found in [13]. We now consider the high-SNR regime when $N_R > 1$.

Theorem 3: Let K^* denote

$$K^* = \lim_{P\to\infty} K_{\text{opt}}(P). \tag{43}$$

a) For $N_R \geq 1$, K^* is given as

$$K^* = \arg\max_{K \geq N_R} -2N_R \log(K) + \sum_{j=1}^{N_R} \Psi(K-j+1) \tag{44}$$

where $\Psi(x)$ is the digamma function [32], and may be expressed, for integer x, as

$$\Psi(x) = -\gamma + \sum_{s=1}^{x-1} \frac{1}{x}. \tag{45}$$

When $x = 1$, $\Psi(x) = -\gamma$, where $\gamma \approx 0.5772$ (Euler's constant). K^* can then be found numerically using (44) and (45), and is a function of N_R.

b) As both N_T and N_R tend to ∞, but the ratio $\eta \equiv \frac{N_T}{N_R}$ remains fixed, $C(K)$ is maximized when $K \approx 1.26 N_R$.

Proof: To prove part a), we start with (35)

$$C(K) \approx \begin{cases} E\left[\log\left|\frac{PN_T}{K^2}\sum_{i=1}^{K} \boldsymbol{z}_i \boldsymbol{z}_i^H\right|\right], & K \geq N_R \\ E\left[\log\left|\frac{PN_T}{K^2}\sum_{i=1}^{N_R} \tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^H\right|\right], & K < N_R. \end{cases} \tag{46}$$

We can rewrite this using results from [32] as

$$C(K) \approx \begin{cases} 2N_R \log\left(\frac{PN_T}{K}\right) + \sum_{j=1}^{N_R} \Psi(K-j+1), & K \geq N_R \\ 2K \log\left(\frac{PN_T}{K}\right) + \sum_{j=1}^{K} \Psi(N_R-j+1), & K < N_R \end{cases} \tag{47}$$

where $\Psi(x)$ is the digamma function [32], and may be expressed, for integer x, as

$$\Psi(x) = -\gamma + \sum_{s=1}^{x-1} \frac{1}{x}.$$

When $x = 1$, $\Psi(x) = -\gamma$. Rewriting (47) as (48) at the bottom of the page, it is clear that as $P \to \infty$, the $\log(P)$ term becomes dominant in both the top and bottom parts of (48), and hence it is always better to have $K \geq N_R$. (If $K < N_R$, the $\log(P)$ term grows as K instead of N_R, so the bottom of (48) will be less than the top.) We can thus write that $K^* = \arg\max_{K \geq N_R} C(K)$, which gives, after eliminating the terms that do not depend on K

$$K^* = \arg\max_{K \geq N_R} -2N_R \log(K) + \sum_{j=1}^{N_R} \Psi(K-j+1) \tag{49}$$

as desired.

To prove part b), we have that, as N_T and $N_R \to \infty$, but the ratio $\eta \equiv \frac{N_T}{N_R}$ remains fixed, (see [8])

$$\frac{C(\tau)}{N_R} \to \frac{1}{2\pi}\int_{\upsilon-}^{\upsilon+} \log\left(1+\frac{P\eta}{\tau^2}\upsilon\right) \times \sqrt{\left(\frac{\upsilon+}{\upsilon}-1\right)\left(1-\frac{\upsilon-}{\upsilon}\right)}\, d\upsilon \tag{50}$$

$$C(K) \approx \begin{cases} 2N_R \log(P) + 2N_R \log\left(\frac{N_T}{K}\right) + \sum_{j=1}^{N_R} \Psi(K-j+1), & K \geq N_R \\ 2K \log(P) + 2K \log\left(\frac{N_T}{K}\right) + \sum_{j=1}^{K} \Psi(N_R-j+1), & K < N_R \end{cases} \tag{48}$$

where $\tau \equiv \frac{K}{N_R}$, and $\upsilon\pm = (\sqrt{\tau} \pm 1)^2$. In the limit of large P

$$\frac{C(\tau)}{N_R} \approx \log(P) + \log(\eta) - 2\log(\tau) + \frac{1}{2\pi}\int_{\upsilon-}^{\upsilon+} \log(\upsilon)\sqrt{\left(\frac{\upsilon+}{\upsilon} - 1\right)\left(1 - \frac{\upsilon-}{\upsilon}\right)}d\upsilon \quad (51)$$

which is maximized when $\tau \approx 1.26$.[1] □

When there are two receive antennas, $K^* = 2$, so at high SNR, capacity is not increased by having more than two eigenmodes. When $N_R = 3$, $K^* = 4$, and when $N_R = 4$, $K^* = 5$, indicating that for larger numbers of receive antennas and high SNR, $K^* > N_R$. Even as the number of receive and transmit antennas approach ∞ (with fixed ratio), in the limit of large SNR, $K^* \approx 1.26N_R$, indicating that the optimal number of eigenmodes is of the same order as the number of receive antennas.

We have now seen that the low SNR approximation (34) to the capacity is monotonically decreasing in K, and that the high SNR approximation (35) has a maximum at K^*. Note that K^* is a function of N_R alone as N_T gets large, showing that the optimal number of eigenmodes, and hence encoder and decoder complexity, is limited even as we increase the beamforming gain per eigenmode.

If the behavior of the actual capacity as a function of K were sandwiched between that of the high and low SNR approximations, we would expect that $K_{\text{opt}}(P)$ grows from $K_{\text{opt}} = 1$ to $K_{\text{opt}} = K^*$ monotonically as P gets large. While this property is seen to hold in all our simulations, we have not been able to prove it, and state it as a conjecture below.

Conjecture: $K_{\text{opt}}(P)$ is a nondecreasing function of P, increasing from 1 to K^* as P varies over $(0, \infty)$.

If this conjecture were true, it would imply that the optimal number of eigenmodes is bounded by K^*, which scales with N_R, even as N_T gets arbitrarily large. This allows large beamforming gains from increasing the number of transmit antennas, while limiting the complexity of MIMO processing to that of a virtual MIMO system whose size scales with N_R rather than N_T (see discussion in Section VII). For $1 \leq N_R \leq 3$, we have been able to prove a weaker form of the conjecture which has the same practical consequence. This result is stated in the following theorem.

Theorem 4: Let $1 \leq N_R \leq 3$. Then $K_{\text{opt}}(P) \leq K^*$, for $\alpha \leq P \leq \infty$, and for all N_T, where α is close to 0. Of course, for $P < \alpha$, and values of N_T that are not very large, $K_{\text{opt}}(P) = 1$ by Theorem 1, and so $K_{\text{opt}}(P)$ is necessarily smaller than or equal to K^*. (Note that we now explicitly denote the dependence of K_{opt} on P as $K_{\text{opt}}(P)$.)

Proof: The proof is given in the Appendix.

A consequence of this theorem is that when $N_R = 1$, it is best (in terms of ergodic capacity) when the channel has one dominant eigenmode. This result is further confirmed by [13], [14]. The numerical results obtained for the proof also indicate that when $N_R = 2$, it is best for the channel to have two dominant eigenmodes, unless P is small, in which case it is best when the channel has only one dominant eigenmode.

[1]Maximization is done numerically over the interval $1 \leq \tau \leq 10$.

C. Practical Implications

To explore the practical implications of our thought experiment, consider a system with a single receive antenna element. Theorem 4 says that, for $N_R = 1$, having a single-channel eigenmode with all the channel energy maximizes capacity. This can be confirmed using the results in [13], [14]. In practice, one can only control the channel eigenvalues via the antenna spacing. The antenna spacing should then be such that, if possible, there is only a single dominant eigenmode. This is demonstrated in the following simulation results.

As a practical example, we consider optimizing the antenna spacing for our running example of a BS with six antennas transmitting to a mobile whose PAP is $L(0°, 5°)$. The SNR, P/σ_n^2, is set to 10 dB. As the antenna spacing changes, so does $\mathbf{\Lambda}_C$, and hence the optimal values of p_i, which can be solved for numerically. Fig. 7 shows how the ergodic capacity changes for different values of d/λ (the antenna spacing over the wavelength). At $d/\lambda = 8$, all channel eigenvalues are equal and hence the capacity at this point corresponds to the maximum capacity attainable when there is no feedback. As d/λ decreases, the channel energy becomes concentrated in fewer eigenmodes, until only one eigenmode is dominant. Below $d/\lambda = 0.5$, beamforming is optimal. (See [11] for the necessary and sufficient conditions for the optimality of beamforming.) We do not consider values of d/λ smaller than 0.4 because at very close spacing, the different antennas can no longer be treated as separate elements due to electromagnetic coupling.

It is evident that beamforming with the BS antennas spaced at 0.4λ is superior to using a full blown space–time code with $d/\lambda = 8$, giving a gain of over 1.5 bits/s/Hz. Not only is capacity increased by using a smaller spacing, but complexity is decreased dramatically by using beamforming instead of space–time codes.

Now, suppose that there are two receive elements. Considering Theorem 1, and noting that $K^* = 2$ when $N_R = 2$, it is reasonable to suppose that for a system with $N_R = 2$ and moderate SNR, spacing the antennas such that there are two dominant eigenmodes should give the best performance. We once again look at our running example with $N_T = 6$ and the PAP $\sim L(0°, 5°)$, but with N_R now equal to 2. For different values of d, the optimal powers p_i are calculated numerically by approximating derivatives by differentials and using the projected gradient descent algorithm. Values for the ergodic capacity are plotted versus d/λ in Fig. 8. Below $d/\lambda = 0.82$, sending along two eigenmodes is optimal. As expected, the maximum capacity occurs when two eigenmodes are dominant, at $d/\lambda = 0.7$.

V. Antenna Spacing and Outage Rates

We have seen how smaller antenna spacing, because it leads to fewer eigenmodes, can lead to larger ergodic capacity in wideband systems. However, smaller spacing also reduces the spatial diversity, and this reduction could potentially lead to increased outages. In this section, we consider how the antenna spacing affects the outage rate $R(\epsilon)$, which is defined to be the

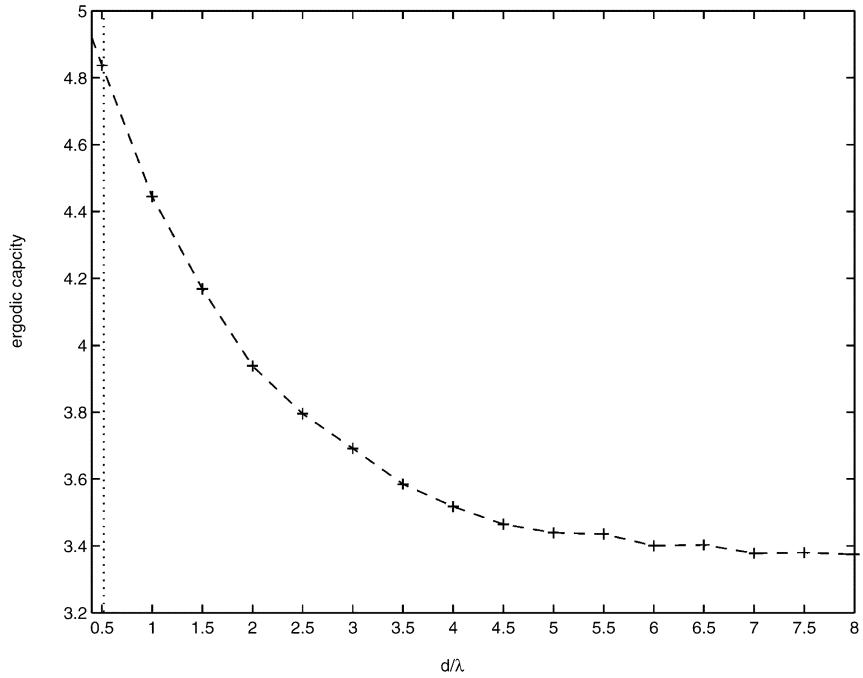


Fig. 7. Ergodic capacity versus d/λ when $N_T = 6$, $N_R = 1$, and $\Omega \sim L(0^\circ, 5^\circ)$. To the left of the dotted line, beamforming is optimal.

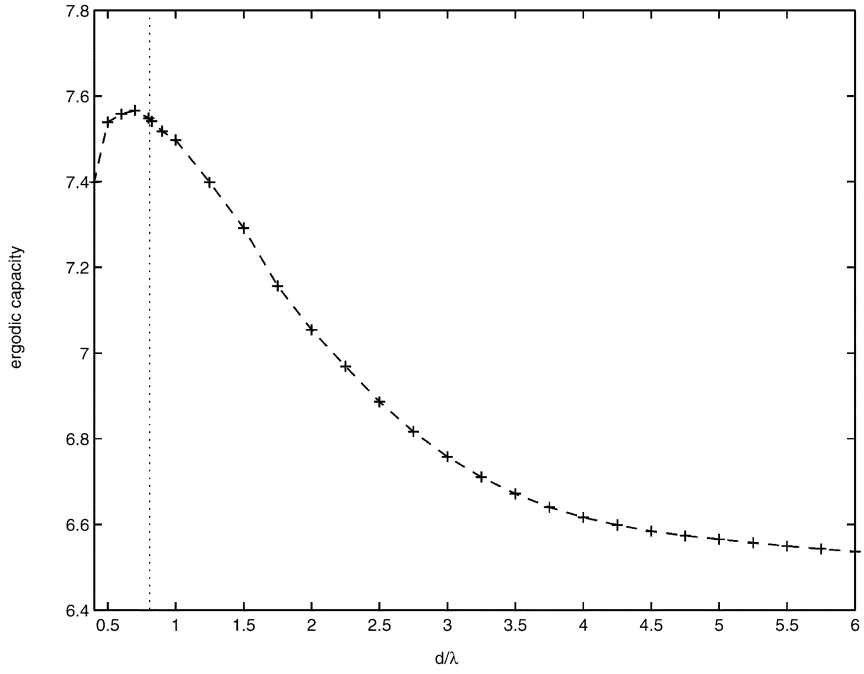


Fig. 8. Ergodic capacity versus d/λ when $N_T = 6$, $N_R = 2$, and $\Omega \sim L(0^\circ, 5^\circ)$. To the left of the dotted line, "two-directional beamforming" is optimal.

largest transmission rate R, such that the following condition holds

$$P[I_W \leq R] \leq \epsilon \tag{52}$$

where I_W is the spectral efficiency of the system, as defined below

$$I_W = \frac{1}{W} \int_{-W/2}^{W/2} \log \left| \boldsymbol{I}_{N_R} + \boldsymbol{H}(\boldsymbol{f}) \boldsymbol{Q} \boldsymbol{H}(\boldsymbol{f})^H \right| df. \tag{53}$$

(The spectral efficiency for a wideband system with a large number of subcarriers is well approximated by the preceding integral, which is the limit when the size of each frequency bin goes to zero.) As can be seen, the spectral efficiency is a function of the transmit covariance matrix $\boldsymbol{Q}$.

In [3], we showed that the spectral efficiency for a wideband system without feedback can be accurately modeled as a Gaussian random variable. We find that the same holds true for the wideband systems with covariance feedback we are presently considering. The channel frequency response decorrelates fast enough so that the summation in the formula for

the spectral efficiency involves sufficiently many independent variables for the central limit theorem to come into effect. Thus, we can approximate $R(\epsilon)$ as

$$\hat{R}(\epsilon) \approx E[I_W] - \sqrt{var[I_W]} * Q^{-1}(\epsilon) \tag{54}$$

where $Q(x)$ is the Gaussian Q function.

We now show that for large enough bandwidths, the variance of the spectral efficiency is fairly insensitive to the transmit strategy, so that the transmit strategy which maximizes ergodic capacity also approximately maximizes the outage rate. We begin by characterizing the variance of the spectral efficiency for an arbitrary transmit covariance matrix $\boldsymbol{Q}$.

Theorem 5: The variance of I_W for a single cluster wideband system using covariance feedback can be approximated as

$$var[I_W] \approx \frac{N_R \sum_{i=1}^{N_T} \gamma_i^2}{\left(1+\sum_{i=1}^{N_T} \gamma_i\right)^2} \sum_{l=0}^{\infty} A_l^4 \tag{55}$$

where the transmit covariance matrix $\boldsymbol{Q} = \boldsymbol{A}^H\boldsymbol{A}$ and the $\{\gamma_i\}$ $i = 1 \ldots N_T$ are the eigenvalues of $\boldsymbol{ACA}^H$. (Recall that $\boldsymbol{C}$ is the channel covariance matrix.) The channel tap weights $\{A_l\}$ are proportional to the square root of the power delay profile, as defined in (6). If the BS sends along the optimal directions for maximizing ergodic capacity, then the variance of the spectral efficiency can be approximated as follows:

$$var[I_W] \approx \frac{N_R \sum\limits_{i=1}^{N_T} (\lambda_i p_i)^2}{\left(1+\sum\limits_{i=1}^{N_T} \lambda_i p_i\right)^2} \sum_{l=0}^{\infty} A_l^4 \tag{56}$$

where, for $1 \le i \le N_T$, λ_i is the ith eigenvalue of the channel covariance matrix, and p_i is the power transmitted along the ith eigenmode.

Proof: We start with the definition of spectral efficiency I_W

$$I_W = \frac{1}{W}\int_{-W/2}^{W/2} \log\left|\boldsymbol{I}_{N_R} + \boldsymbol{H}(\boldsymbol{f})\boldsymbol{Q}\boldsymbol{H}(\boldsymbol{f})^H\right| df \tag{57}$$

$$= \frac{1}{W}\int_{-W/2}^{W/2} \sum_{i=1}^{N_R} \log(1+\mu_i(f)) df \tag{58}$$

where $\mu_i(f)$, $i = 1 \ldots N_R$, are the eigenvalues of $\boldsymbol{M}(f)$, with $\boldsymbol{M}(f) \equiv \boldsymbol{H}(f)\boldsymbol{Q}\boldsymbol{H}^H(f)$. Letting $\bar{\mu}_i(f)$ denote the average value of $\mu_i(f)$, we can then write (59)–(62) at the bottom of the page, using $\log(1+x) \approx x$ in the final step. Since the channel responses at different frequencies are identically distributed, and since, at any particular frequency, the N_R eigenvalues of $\boldsymbol{M}(f)$ have the same mean, we have that $\bar{\mu}_i(f) = \bar{\mu}_j(f)\ \forall\ i, j,\ \forall\ f$. Letting $\bar{\mu} = \bar{\mu}_i(f)\ i = 1 \ldots N_R\ \forall\ f$, the variance of the spectral efficiency is then

$$\text{var}[I_W] \approx \text{var}\left[\frac{1}{W}\int_{-W/2}^{W/2}\left(\sum_{i=1}^{N_R}\frac{\mu_i(f)}{1+\bar{\mu}}\right) df\right] \tag{63}$$

$$= \text{var}\left[\frac{1}{W}\int_{-W/2}^{W/2}\left(\frac{\text{trace}(\boldsymbol{M}(f))}{1+\bar{\mu}}\right) df\right] \tag{64}$$

$$= N_R \text{var}\left[\frac{1}{W}\int_{-W/2}^{W/2}\left(\frac{\boldsymbol{H}_{(1,:)}(f)\boldsymbol{Q}\boldsymbol{H}_{(1,:)}^H(f)}{1+\bar{\mu}}\right) df\right] \tag{65}$$

$$= \frac{N_R}{(1+\bar{\mu})^2}\text{var}\left[\sum_{l=0}^{\infty} A_l^2 \boldsymbol{v}_l^H\boldsymbol{Q}\boldsymbol{v}_l\right] \tag{66}$$

$$= \frac{N_R \text{var}\left[\boldsymbol{v}_1^H\boldsymbol{Q}\boldsymbol{v}_1\right]}{(1+\bar{\mu})^2}\sum_{l=0}^{\infty} A_l^4 \tag{67}$$

where, in (65), $\boldsymbol{H}_{(1,:)}(f)$ is the first row of $\boldsymbol{H}(f)$, and we have used the fact that the rows of $\boldsymbol{H}(f)$ are i.i.d. for all frequencies. Equation (66) is a consequence of Parseval's theorem, and (67) results because the $\{\boldsymbol{v}_l\}$ are i.i.d. (see (8)). We now note that

$$\bar{\mu} = E\left[\boldsymbol{H}_{(1,:)}(f)\boldsymbol{Q}\boldsymbol{H}_{(1,:)}^H(f)\right], \quad \forall\ f \tag{68}$$

$$= E\left[\boldsymbol{v}_1^H\boldsymbol{Q}\boldsymbol{v}_1\right] \tag{69}$$

since the rows of $\boldsymbol{H}$ and $\boldsymbol{v}_1^H$ are identically distributed (see (8) and (11), and note that $\boldsymbol{v}_1$ is a column, not a row.) Writing $\boldsymbol{Q}$ as $\boldsymbol{Q} = \boldsymbol{A}^H\boldsymbol{A}$, and letting $\tilde{\boldsymbol{v}} \equiv \boldsymbol{A}\boldsymbol{v}_1$, we have that $\tilde{\boldsymbol{v}} \sim CN(0, \boldsymbol{ACA}^H)$, and we denote the eigenvalues of $\boldsymbol{ACA}^H$ by $\gamma_1 \ldots \gamma_{N_T}$. It follows that

$$\bar{\mu} = E\left[\boldsymbol{v}_1^H\boldsymbol{Q}\boldsymbol{v}_1\right] \tag{70}$$

$$= E\left[\tilde{\boldsymbol{v}}^H\tilde{\boldsymbol{v}}\right] \tag{71}$$

$$= \sum_{i=1}^{N_T}\gamma_i. \tag{72}$$

Similarly, it can be shown that

$$\text{var}\left[\boldsymbol{v}_1^H\boldsymbol{Q}\boldsymbol{v}_1\right] = \text{var}\left[\tilde{\boldsymbol{v}}^H\tilde{\boldsymbol{v}}\right] \tag{73}$$

$$= \sum_{i=1}^{N_T}\gamma_i^2 \tag{74}$$

$$I_W = \frac{1}{W}\int_{-W/2}^{W/2}\left(\sum_{i=1}^{N_R}\log(1+\bar{\mu}_i(f)+\mu_i(f)-\bar{\mu}_i(f))\right) df \tag{59}$$

$$= \frac{1}{W}\int_{-W/2}^{W/2}\left(\sum_{i=1}^{N_R}\log\left[(1+\bar{\mu}_i(f))\left(1+\frac{\mu_i(f)-\bar{\mu}_i(f)}{1+\bar{\mu}_i(f)}\right)\right]\right) df \tag{60}$$

$$= \frac{1}{W}\int_{-W/2}^{W/2}\left(\sum_{i=1}^{N_R}\log(1+\bar{\mu}_i(f)) + \log\left(1+\frac{\mu_i(f)-\bar{\mu}_i(f)}{1+\bar{\mu}_i(f)}\right)\right) df \tag{61}$$

$$\approx \frac{1}{W}\int_{-W/2}^{W/2}\left(\sum_{i=1}^{N_R}\log(1+\bar{\mu}_i(f)) + \frac{\mu_i(f)-\bar{\mu}_i(f)}{1+\bar{\mu}_i(f)}\right) df \tag{62}$$

and, therefore, substituting (74) and (72) into (67)

$$\mathrm{var}[I_W] \approx \frac{N_R \sum_{i=1}^{N_T} \gamma_i^2}{\left(1+\sum_{i=1}^{N_T}\gamma_i\right)^2} \sum_{l=0}^{\infty} A_l^4. \tag{75}$$

If $\boldsymbol{Q}$ has eigenmodes along the channel eigenmodes (as $\boldsymbol{Q}_0$ (15) does), then we have

$$\mathrm{var}[I_W] \approx \frac{N_R \sum_{i=1}^{N_T} (\lambda_i p_i)^2}{\left(1+\sum_{i=1}^{N_T}\lambda_i p_i\right)^2} \sum_{l=0}^{\infty} A_l^4 \tag{76}$$

as desired. □

Remark: We can infer from Theorem 5 that the transmission strategy which optimizes the mean of the spectral efficiency also approximately maximizes the outage rate in a wideband system, as follows.

Note that the variance of the spectral efficiency (55) depends on the term $\sum_{l=0}^{\infty} A_l^4$, which multiplies a term dependent on the transmission strategy. It can easily be seen that the term dependent on the transmission strategy can be bounded as follows:

$$\frac{\sum_{i=1}^{N_T} \gamma_i^2}{\left(1+\sum_{i=1}^{N_T}\gamma_i\right)^2} \leq 1 \tag{77}$$

since $\gamma_i \geq 0 \ \forall \ i$. We show below that, under a mild technical condition on the power–delay profile (satisfied by all power–delay profiles that we have encountered in the literature), $\sum_{l=0}^{\infty} A_l^4 \sim \frac{1}{W}$, and thus, noting the bound in (77), it can be seen that, regardless of the transmission strategy, $\mathrm{var}[I_W]$ decreases with bandwidth as $\frac{1}{W}$. This means that for large enough bandwidths, changes in the *variance* of the spectral efficiency due to changes in the transmission strategy are small relative to changes in the *mean* of the spectral efficiency due to changes in the transmission strategy. (Note that the mean of the spectral efficiency is independent of bandwidth.) We now show that $\sum_{l=0}^{\infty} A_l^4 \sim \frac{1}{W}$ for most power delay profiles.

Proposition 1: For smooth power delay profiles $P_\tau(\tau')$ that are square integrable, i.e.,

$$\int_0^{\infty} P_\tau^2(\tau') d\tau' < \infty \tag{78}$$

it holds that, for large bandwidths

$$\sum_{l=0}^{\infty} A_l^4 \sim \frac{1}{W} \tag{79}$$

where the $A_l \propto \sqrt{P_\tau\left(\frac{l}{W}\right)}$, but are normalized such that $\sum_{l=0}^{\infty} A_l^2 = 1$.

Proof: We first note that the power delay profile is normalized to have unit energy

$$\int_0^{\infty} P_\tau(\tau') d\tau' = 1. \tag{80}$$

Converting from an integral to a Riemann sum, for large W, we have

$$\sum_{l=0}^{\infty} P_\tau\left(\frac{l}{W}\right)\frac{1}{W} \approx 1 \tag{81}$$

or

$$\sum_{l=0}^{\infty} P_\tau\left(\frac{l}{W}\right) \approx W. \tag{82}$$

Since $\sum_{l=0}^{\infty} A_l^2 = 1$, and $A_l^2 \propto P_\tau\left(\frac{l}{W}\right)$, this gives

$$A_l^2 \approx \frac{P\left(\frac{l}{W}\right)}{W}, \quad \forall\, l \tag{83}$$

and thus,

$$\sum_{l=0}^{\infty} A_l^4 \approx \sum_{l=0}^{\infty} \frac{P^2\left(\frac{l}{W}\right)}{W^2} \tag{84}$$

$$\approx \frac{1}{W}\int_0^{\infty} P_\tau^2(\tau') d\tau' \tag{85}$$

as long as $\int_0^{\infty} P_\tau^2(\tau') d\tau'$ is bounded, $\sum_{l=0}^{\infty} A_l^4$ decays as $\frac{1}{W}$. □

Equation (85) in the preceding proof is consistent with the results in [3] for an exponential power-delay profile, for which

$$\sum_{l=0}^{\infty} A_l^4 \approx \frac{1}{2W\tau_{\mathrm{rms}}}$$

where τ_{rms} is the rms of the power delay profile. Thus, for large enough bandwidths, the variance of the spectral efficiency is small, and fairly insensitive to changes in the transmission strategy. In contrast, the mean of the spectral efficiency is unaffected by the bandwidth, and is therefore large compared to the variance for large W. It is therefore reasonable, when in the large bandwidth regime, to use the transmission strategy that maximizes the mean of the spectral efficiency, even when trying to keep outages at a minimum. We henceforth assume that the transmission covariance matrix used is that which maximizes ergodic capacity, namely, $\boldsymbol{Q}_0$.

We now give a numerical example illustrating that, for the wideband systems we consider, the antenna spacing that maximizes the ergodic capacity also approximates maximizes the outage rate $R(0.01)$. This is further confirmation of our hypothesis that, for large bandwidths, the outage rate is governed by the mean rather than the variance of the spectral efficiency.

Fig. 9 shows the 1% outage rate, $R(0.01)$, plotted as a function of antenna spacing for our running example where $N_T = 6$ and $N_R = 2$. As before, there are 1024 subcarriers spaced 25 kHz apart, the PAP is $L(0°, 5°)$, the PDP is exponential with $\tau_{\mathrm{rms}} = 0.5$ μs, and SNR $= 10$ dB. The solid line is the simulated outage rate while the dotted line is the estimated outage rate using (54) and (56). (The simulated value for $E[I_W]$ is used in (54)). The two curves are clearly well matched, indicating the validity of the Gaussianity assumption.

The changes in the variance of I_W as a function of d/λ are small compared to the changes in the ergodic capacity, hence the shape of the outage rate curve is very similar to that of ergodic capacity. This can be seen in Fig. 10. In fact, the optimal antenna spacing is the same for both curves. In general, for the wideband systems we consider, the antenna spacing which maximizes ergodic capacity also approximately maximizes $R(0.01)$. Intuitively, this is because there is enough frequency diversity

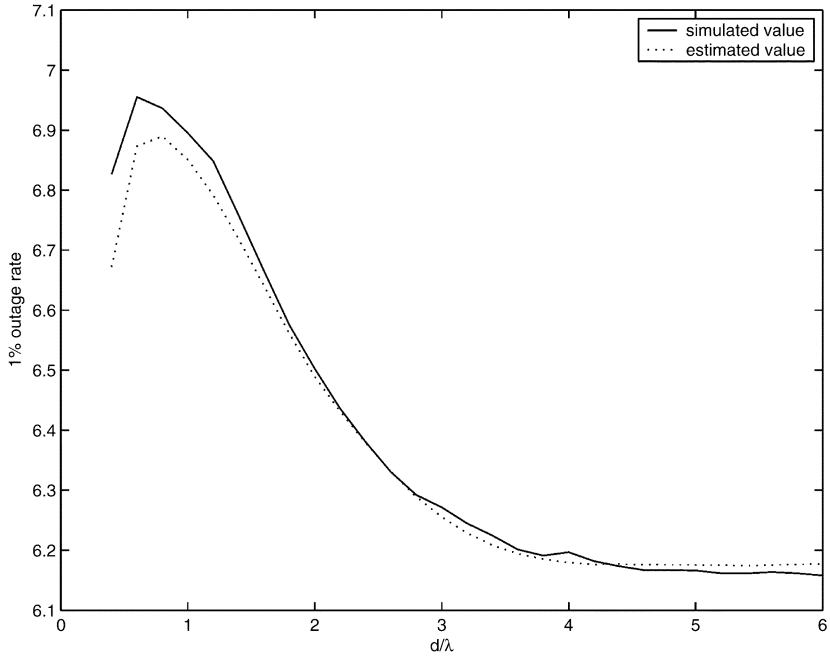


Fig. 9. 1% outage rate versus d/λ when $N_T = 6$, and $N_R = 2$. $\Omega \sim L(0^\circ, 5^\circ)$, and $\tau_{\rm rms} = 0.5\ \mu$s. The dotted line uses the Gaussian approximation for outage rates.

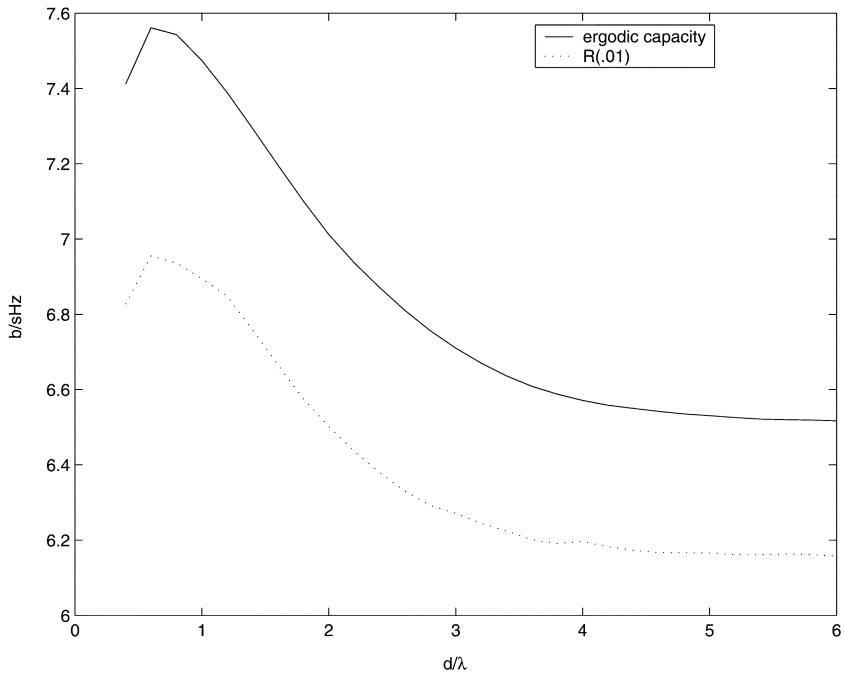


Fig. 10. Ergodic capacity and 1% outage rate versus d/λ when $N_T = 6$, and $N_R = 2$. $\Omega \sim L(0^\circ, 5^\circ)$, and $\tau_{rms} = 0.5\ \mu$s.

such that reducing the antenna spacing, and hence the spatial diversity, does not have the negative effect on outage rates that it does in a narrowband system.

VI. Channels With Multiple Clusters

Covariance feedback can be obtained and used by the transmitter regardless of the number of clusters in the channel, since the space–time channels for different frequency bins are still identically distributed. Although the theorems relating to the hypothetical system in Section IV still hold when there are multiple clusters, it becomes more difficult to control the actual eigenvalue distribution of the channel. Reducing the antenna spacing tends to shift the energy toward the stronger eigenmodes, but the energy in the dominant eigenmode is no longer strictly increasing with decreasing spacing. Also, the presence of multiple clusters may make it impossible to obtain the desired number of dominant eigenmodes. Despite these difficulties, appropriately chosen antenna spacing, guided by the results

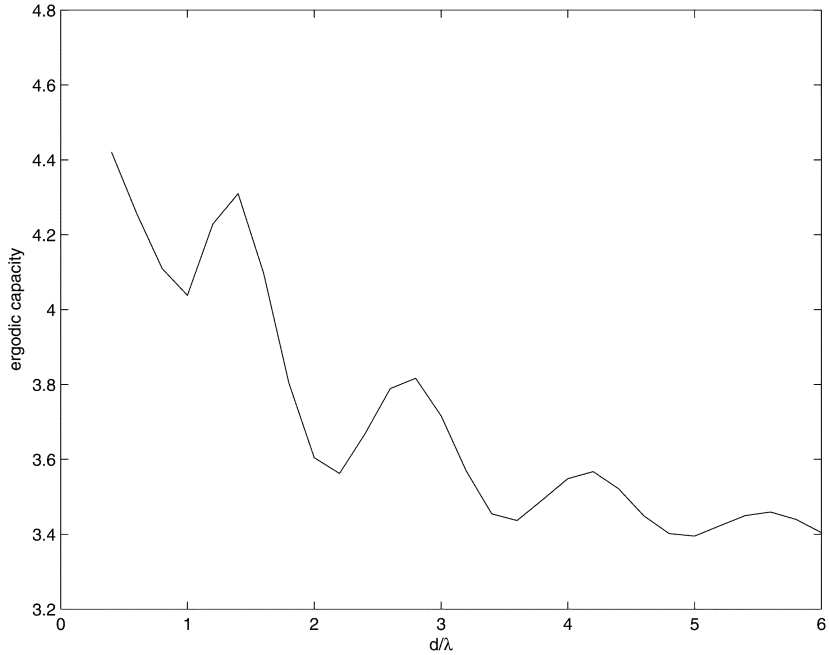


Fig. 11. Ergodic capacity versus d/λ when $N_T = 6$, and $N_R = 1$. $\Omega_1 \sim L(0^\circ, 5^\circ)$, $\Omega_2 \sim L(45^\circ, 5^\circ)$, and $\tau_{\rm rms} = 0.5\ \mu$s.

in Section IV, still leads to substantial gains in both ergodic capacity and outage rates.

For illustrative purposes, we consider a two cluster channel. Let $\{P_{\Omega_j}(\Omega), P_{\tau_j}(\tau) : j = 1, 2\}$ be the PAPs/PDPs of the clusters and $\{\boldsymbol{C}_j : j = 1, 2\}$ be the corresponding covariance matrices. The lth tap of cluster j is proportional to the jth PDP: $A_{jl} \propto \sqrt{P_{\tau_j}\left(\frac{l}{W}\right)}$, where we normalize such that

$$\sum_{j=1}^{2}\sum_{l=0}^{\infty} A_{jl}^2 = 1.$$

Let $A_l^2 \equiv A_{1l}^2 + A_{2l}^2$. In [3], we showed that the TDL channel response for a two-cluster system can be written as

$$\boldsymbol{h}_W(t) = \sum_{l=0}^{\infty} A_l \boldsymbol{v}_l \delta\left(t - \frac{l}{W}\right) \tag{86}$$

where

$$\boldsymbol{v}_l \sim CN(0, \boldsymbol{V}_l) \tag{87}$$

$$\boldsymbol{V}_l \equiv \alpha_{1l}\boldsymbol{C}_1 + \alpha_{2l}\boldsymbol{C}_2 \tag{88}$$

and

$$\alpha_{jl} = \frac{A_{jl}^2}{A_{1l}^2 + A_{2l}^2}. \tag{89}$$

Fig. 11 shows the ergodic capacity as a function of antenna spacing for a system where $\Omega_1 \sim L(0^\circ, 5^\circ)$, $\Omega_2 \sim L(45^\circ, 5^\circ)$, and the power delay profiles for both clusters are exponential with $\tau_{\rm rms} = 0.5\ \mu$s. The other parameters are $N_T = 6$, $N_R = 2$, and SNR $= 10$ dB. Because the power–delay profiles are the same for both clusters, all the taps in the tap delay line model (86) have the same variance. In other words, the covariance matrices $\boldsymbol{V}_l$ are the same for all l (this is not true for multicluster channels in which different clusters have different PDPs). Thus, we can approximate the variance of the spectral efficiency using the formula in Theorem 5, and then calculate the outage rates using (54). Fig. 12 shows the simulated 1% outage rate plotted with the 1% outage rate estimated using the Gaussian approximation (54), and (56). As before, the ergodic capacity and the outage rate have similar shapes and are both maximized for the same value of d/λ, namely, 0.4. (Antenna spacings smaller than $d/\lambda = 0.4$ are not considered, since their practical feasibility is limited by complications due to electromagnetic coupling.) As can be seen, the capacity at $d/\lambda = 0.4$ is much better than the capacity achieved at $d/\lambda = 6$, which corresponds to the best capacity attainable with a standard space–time code. This is an interesting example of a setting in which the physical system cannot exactly realize the optimal eigenvalue distribution for the hypothetical system, but the latter still provides valuable guidance. Theorem 4 prescribes one dominant channel eigenmode for the hypothetical system, since the receiver has one antenna element. However, the physical channel has two equi-powered clusters spaced sufficiently apart, so that there cannot be a single dominant eigenmode. In this case, the capacity-maximizing antenna spacing in the physical system, $d/\lambda = 0.4$, corresponds to two (the smallest possible number of) dominant eigenmodes.

VII. Discussion

For wideband wireless channels, statistical reciprocity provides a powerful means of generating robust channel feedback with no overhead. Using such feedback on a cellular downlink, we can simultaneously improve performance, reduce encoder/decoder complexity, and reduce the required antenna spacing at the BS. Our results indicate that the BS antenna spacing should be optimized so as to create approximately K dominant eigenmodes, where K is a function of N_T, N_R, and SNR. For systems with small N_R, and moderate SNR, $K \approx N_R$. Beamforming along these eigenmodes creates an $N_R \times N_R$ virtual MIMO system, allowing the use of any of

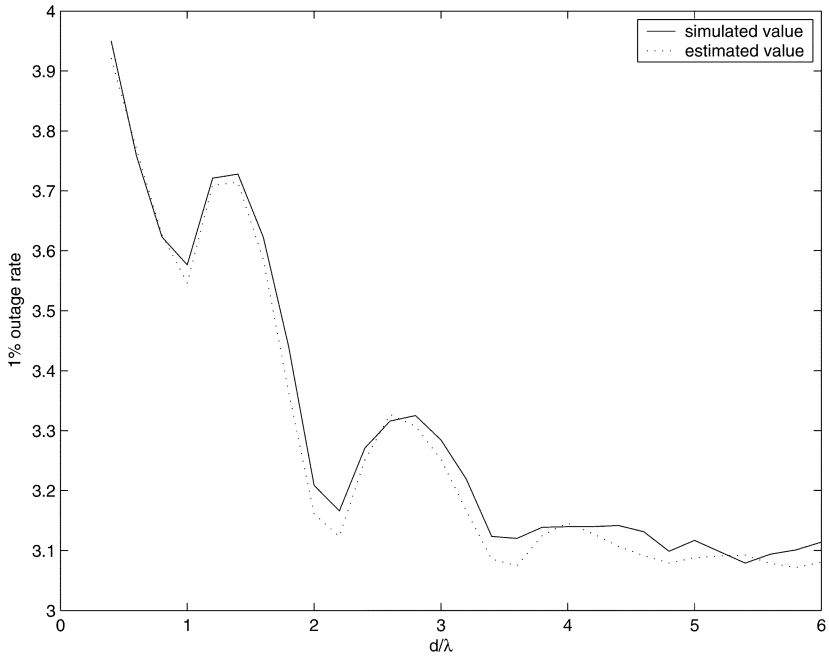


Fig. 12. 1% outage rate versus d/λ when $N_T = 6$, and $N_R = 1$. $\Omega_1 \sim L(0^\circ, 5^\circ)$, $\Omega_2 \sim L(45^\circ, 5^\circ)$, and $\tau_{\rm rms} = 0.5\,\mu$s.

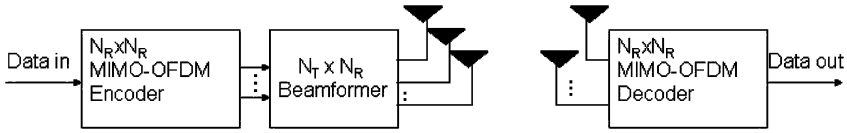


Fig. 13. MIMO–OFDM system with beamforming.

a number of strategies to achieve different points along the diversity–multiplexing tradeoff curve [18].

A block diagram of the MIMO OFDM system with beamforming is shown in Fig. 13. As can be seen, creation of the virtual MIMO system using N_R beamformers can be done in the time domain, after OFDM encoding for an $N_R \times N_R$ MIMO system. This is because the N_R sets of beamforming weights are the same across all subcarriers. The "virtual" MIMO system created is really just a standard MIMO system where the channel matrix is modified to account for the beamforming block. To see this, note that the received symbol vector for a particular frequency tone (10) is now

$$\hat{\boldsymbol{s}}_i = \boldsymbol{H}_i \boldsymbol{U} \boldsymbol{s}_i + \boldsymbol{n}_i \tag{90}$$

where $\boldsymbol{U}$ is the $N_T \times N_R$ beamforming matrix, $\boldsymbol{H}_i$ is the $N_R \times N_T$ channel matrix, $\boldsymbol{s}_i$ is the $N_{\hat{R}} \times 1$ transmitted symbol vector, and $\boldsymbol{n}_i$ is added white Gaussian noise. Since the receiver knows both $\boldsymbol{U}$ and $\boldsymbol{H}_i$, it can consider an effective channel $\tilde{\boldsymbol{H}}_i \equiv \boldsymbol{H}_i \boldsymbol{U}$, and perform standard MIMO processing with the modified $N_R \times N_R$ channel $\tilde{\boldsymbol{H}}_i$. (Note that it is possible to send different powers along the eigenmodes.)

For the mobile, the decoding complexity equals that for an $N_R \times N_R$ MIMO system, regardless of the number N_T of BS antenna elements. Thus, N_T can be scaled up (leading to larger beamforming gains) with no impact on the complexity at the mobile receiver.

An important topic for future work is to devise constructive space–time strategies based on our information-theoretic prescriptions. In this context, an interesting issue is how to choose the antenna spacing when the BS must transmit to mobiles for whom the optimal antenna spacing differs significantly. One possible approach is to equip the BS with a large number of closely spaced antenna elements, and send using a subset of these in a manner optimized for each mobile. The specifics of such a scheme are currently under investigation.

APPENDIX
PROOF OF THEOREM 4

This proof uses a mix of analytical and numerical techniques. Since the proof for each N_R is identical, with the exception of parameter values, we only present the proof for $N_R = 1$. In this case, from Theorem 2, $K^* = 1$. We substitute numerical values only when needed in order to keep the exposition as general as possible.

Step 1 We first fix $N_T = 50$, and show that $K_{\rm opt}(P) \leq K^*$ for $.01 \leq P \leq 100$. In order to do this, we use the following algorithm:

1a: Initialize P. (Let $P = .01$)

1b: Evaluate $C(K, P)$ for $1 \leq K \leq N_T$ where $C(K, P)$ is the capacity (25) with the dependence on P made explicit. This can be done numerically using the formula

$$C(K,P) = N_R E\left[\log\left(1 + P N_T \frac{\nu_K}{K^2}\right)\right] \tag{91}$$

and the distribution of ν_K given in (28). Then check to see if $K_{\rm opt}(P) \leq K^*$

1c: Pick the value $P' = P + \Delta$ for the next iteration such that $K_{\rm opt}(P)$ must remain optimal within the interval $[P, P+\Delta]$. To pick the appropriate value, we first note that

$$\frac{\partial C(K,P)}{\partial P} = \frac{N_R N_T \frac{\nu_K}{K^2}}{1 + P N_T \frac{\nu_K}{K^2}} < \frac{N_R}{P} \qquad (92)$$

and that $C(K,P)$ is an increasing function of P for all K. Thus, by increasing P by Δ, $C(K,P)$ can increase at most $\frac{N_R \Delta}{P}$. Denoting $K_{nextbest}$ as the value of K which gives the next highest value of $C(K,P)$ after $K_{\rm opt}$, we can then say that if

$$\frac{N_R \Delta}{P} < C(K_{\rm opt}, P) - C(K_{\text{next best}}, P) \qquad (93)$$

then $K_{\rm opt}(P)$ is the optimal K between $[P, P+\Delta]$. P' can then be chosen as $P + \Delta$ where Δ satisfies condition (93). If $P' < 100$ go back to Step 1b, substituting P' for P. If $P' \geq 100$, terminate the iterations.

Step 2 We now show that $K_{\rm opt}(P) \leq K^*$ for all $P > 100$ (keeping $N_T = 50$). Using the inequality $\log(1+x) - \log(x) < \frac{1}{x}$, we can write

$$\begin{aligned} 0 &\leq N_R E\left[\log\left(1 + P N_T \frac{\nu_K}{K^2}\right)\right] \\ &\quad - N_R E\left[\log\left(P N_T \frac{\nu_K}{K^2}\right)\right] \\ &\leq N_R E\left[\frac{K^2}{P N_T \nu_K}\right] \leq \frac{N_R}{P} E\left[\frac{K}{\nu_K}\right] \end{aligned} \qquad (94)$$

where the last inequality comes about since $K \leq N_T$. The first term on the left-hand side (LHS) is simply $C(K,P)$ and the second term approximates $C(K,P)$ at large P. For convenience, we denote this term by $\tilde{C}(K,P)$. Let $K^*_{\text{next best}}$ be the particular K which gives the next highest value of $\tilde{C}(K,P)$ after K^*, where $K^*_{nextbest} > K^*$. If

$$\begin{aligned} \frac{1}{N_R}\left[\tilde{C}(K^*,P) - \tilde{C}(K^*_{\text{next best}}, P)\right] \\ > \frac{1}{P} E\left[\frac{K}{\nu_K}\right] \end{aligned} \qquad (95)$$

for $K > K^*$, then $K_{\rm opt}(P)$ (which maximizes the first term on the LHS of (94)) cannot be greater than K^*. Since

$$\begin{aligned} \frac{1}{N_R}\left[\tilde{C}(K^*,P) - \tilde{C}(K^*_{\text{next best}}, P)\right] &= E\left[\frac{\nu_{K^*}}{K^{*2}}\right] \\ - E\left[\frac{\nu_{K^*_{\text{next best}}}}{K^{*2}_{\text{next best}}}\right] \end{aligned} \qquad (96)$$

this quantity is independent of P and N_T. For $N_R = 1$, the LHS of (95) is approximately 0.4, and $E\left[\frac{K}{\nu_K}\right] \leq 2$ for $K > K^*$. Thus, the LHS of (95) is greater than the RHS of (95) for $P \geq 100$, and we have the desired result.

Step 3 Invoking Theorem 1 for $P < .01$, and using the results from Steps 1 and 2, we have that for $N_R = 1$ and $N_T = 50$

$$K_{\rm opt}(P) \leq K^*, \quad \forall\, P. \qquad (97)$$

Since decreasing N_T only decreases the necessary search range for $K_{\rm opt}$, and changes the effective power term PN_T in (91), (97) implies that $K_{\rm opt}(P) \leq K^* \,\forall\, P$ for all $N_T < 50$.

Step 4 It remains to be shown what happens to $K_{\rm opt}(P)$ when $N_T > 50$. In this case, $K_{\rm opt}(P)$ could potentially be in a region that has not previously been searched (i.e., $K_{\rm opt}$ could be greater than 50). We would thus like to bound the search to $1 < K < K_{\max}$, where in this particular case, $K_{\max} = 50$. We start by rewriting (91) as follows:

$$\begin{aligned} C(K,P) &= \log(PN_T) + E\left[\log\left(\frac{1}{PN_T} + \frac{\nu_K}{K^2}\right)\right] \\ &\leq \log(PN_T) + \log\left(\frac{1}{PN_T} + \frac{1}{K}\right) \end{aligned} \qquad (98)$$

where the inequality follows from Jensen's bound. If, for some P, N_T, and $\tilde{K}$

$$\begin{aligned} \log(PN_T) + \log\left(\frac{1}{PN_T} + \frac{1}{\tilde{K}}\right) &< \log(PN_T) \\ &\quad + E\left[\log\left(\frac{1}{PN_T} + \frac{\nu_{K^*}}{K^{*2}}\right)\right] \end{aligned} \qquad (99)$$

then the search can be terminated at $K_{\max} = \tilde{K}$. The inequality in (99) can be simplified to

$$\log\left(\frac{1}{PN_T} + \frac{1}{\tilde{K}}\right) < E\left[\log\left(\frac{\nu_{K^*}}{K^{*2}}\right)\right] \qquad (100)$$

since

$$E\left[\log\left(\frac{\nu_{K^*}}{K^{*2}}\right)\right] < E\left[\log\left(\frac{1}{PN_T} + \frac{\nu_{K^*}}{K^{*2}}\right)\right]. \qquad (101)$$

For $N_R = 1$, $K^* = 1$, and the RHS of (100) ≈ -0.57. Substituting $\tilde{K}$ with 50, (100) holds for $\frac{1}{PN_T} < 0.54$. When $N_T = 50$, this means (100) holds for $P > .03$, and when $N_T > 50$, P must be greater than some value smaller than 0.03. We have thus shown that $K_{\rm opt}(P) \leq K^*$ for $.03 < P < \infty$, and for all N_T. The lower bound for P can be made smaller by using the RHS of (99) instead of the RHS of (100), and by setting N_T to something larger than 50 in Step 1. Also note that for $N_T \leq 50$, $K_{\rm opt} \leq K*$ for *all* P. □

REFERENCES

[1] N. Jacobsen, G. Barriac, and U. Madhow, "Noncoherent eigenbeamforming for a wideband cellular uplink," in *Proc. Int. Symp. Information Theory*, Chicago, IL, June 2004, p. 381.

[2] G. Barriac and U. Madhow, "Characterizing outage capacity for space-time communication over wideband wireless channels," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, vol. 2, 2002, pp. 1513–1517.

[3] ——, "Characterizing the outage capacity for space-time communication over wideband channels," *IEEE Trans. Commun.*, to be published.

[4] D. Morgan, "Downlink adaptive array algorithms for cellular mobile communications," *IEEE Trans. Commun.*, vol. 51, pp. 476–488, Mar. 2003.

[5] B. Hochwald and T. Marzetta, "Adapting a downlink array from uplink measurements," *IEEE Trans. Signal Processing*, vol. 49, pp. 642–653, Mar. 2001.

[6] Y. Liang and F. Chin, "Downlink channel covariance matrix estimation and its applications in wireless DS-CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 19, pp. 222–232, Feb. 2001.

[7] G. Raleigh and V. Jones, "Adaptive antenna transmission for frequency duplex digital wireless communication," in *Proc. Int. Conf. Communications(ICC)*, vol. 6, June 1997, pp. 641–646.
[8] E. Telatar, "Capacity of multi-antenna Gaussian channels," AT&T Bell Labs Internal Tech. Memo BL0112170-950615-07TM, 1995.
[9] ——, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, pp. 585–595, Dec. 1999.
[10] E. Visotsky and U. Madhow, "Space-time transmit precoding with imperfect channel feedback," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2632–2639, Sept. 2001.
[11] S. Jafar and A. Goldsmith, "On optimality of beamforming for multiple antenna systems with imperfect feedback," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, June 2001, p. 321.
[12] S. Jafar, S. Vishwanath, and A. Goldsmith, "Channel capacity and beamforming for multiple transmit and receive antennas with covariance feedback," in *Proc. IEEE Conf. Communication*, vol. 7, 2001, pp. 2266–2270.
[13] H. Boche and E. Jorswieck, "Optimum power allocation, and complete characterization of the impact of correlation on the capacity of MISO systems with different CSI at the transmitter," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, June/July 2003, p. 353.
[14] ——, "Optimum power allocation for MISO systems, and complete characterization of the impact of correlation on the capacity," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, 2003, pp. 373–376.
[15] ——, "Optimal transmission strategies and impact of correlation in multi-antenna systems with different types of channel state information," *IEEE Trans. Signal Processing*, vol. 52, pp. 3440–3453, Dec. 2004.
[16] S. Simon and A. Moustakas, *Optimizing MIMO Antenna Systems With Channel Covariance Feedback*: Bell Labs Tech. Memo., 2002.
[17] H. Bolcskei, D. Gesbsert, and A. Paulraj, "On the capacity of OFDM-based spatial multiplexing systems," *IEEE Trans. Commun.*, vol. 50, pp. 225–234, Feb. 2002.
[18] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental trade-off in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.
[19] S. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1451–1458, Oct. 1998.
[20] V. Tarokh, H. Jafarkhani, and A. Calderbank, "Space-time block coding for wireless communications: Performance results," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 451–460, Mar. 1999.
[21] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi element antennas," *Bell Labs Tech. J.*, vol. 1, pp. 41–59, Mar. 1996.
[22] M. S. Haykin, "Turbo-blast for wireless communications: Theory and experiments," *IEEE Trans. Signal Processing*, vol. 50, pp. 2538–2546, Oct. 2002.
[23] S. Zhou and G. Giannakis, "Optimal transmitter eigen-beamforming and space-time block coding based on channel mean feedback," *IEEE Trans. Signal Processing*, vol. 50, pp. 2599–2613, Oct. 2002.
[24] P. Xia, S. Zhou, and G. Giannakis, "Adaptive MIMO OFDM based on partial channel state information," *IEEE Trans. Signal Processing*, vol. 52, pp. 202–213, Jan. 2004.
[25] G. Jongren, M. Skoglund, and B. Ottersten, "Combining beamforming and orthogonal space-time block coding," *IEEE Trans. Inform. Theory*, vol. 48, pp. 611–627, Mar. 2002.
[26] F. W. Vook, T. Thomas, and X. Zhuang, "Transmit diversity and transmit adaptive arrays for broadband mobile OFDM systems," *Wireless Commun. Networking*, vol. 1, pp. 44–49, Mar. 2003.
[27] A. Saleh and R. Valenzuela, "A statistical model for indoor multi-path propagation," *IEEE J. Select. Areas Commun.*, vol. 5, pp. 128–137, Feb. 1987.
[28] K. I. Pedersen, P. Mogensen, and B. H. Fleury, "Dual-polarized model of outdoor propagation environments for adaptive antennas," in *Proc. IEEE 49th Vehicular Technology Conf.*, vol. 2, May 1999, pp. 990–5.
[29] ——, "Power-azimuth spectrum in outdoor environments," in *IEE Electron. Lett.*, vol. 33, Aug. 1997, pp. 1583–84.
[30] U. Martin, "A directional radio channel model for densely built up urban areas," in *Proc. 2nd EMPCC*, Oct. 2002, pp. 237–244.
[31] M. Nicoli, O. Simeone, and U. Spagnolini, "Multislot estimation of fast-varying space-time channels in TD-CDMA systems," *IEEE Commun. Lett.*, vol. 6, pp. 376–378, Sept. 2002.
[32] O. Oyman, R. Nabar, H. Bolcskei, and A. Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Processing*, vol. 51, pp. 2784–2795, Nov. 2003.