Rayleigh Fading Multi-Antenna Channels

Alex Grant

Institute for Telecommunications Research, University of South Australia, Mawson Lakes Boulevard,
Mawson Lakes, SA 5095, Australia
Email: alex.grant@unisa.edu.au

Information theoretic properties of flat fading channels with multiple antennas are investigated. Perfect channel knowledge at the receiver is assumed. Expressions for maximum information rates and outage probabilities are derived. The advantages of transmitter channel knowledge are determined and a critical threshold is found beyond which such channel knowledge gains very little. Asymptotic expressions for the error exponent are found. For the case of transmit diversity closed form expressions for the error exponent and cutoff rate are given. The use of orthogonal modulating signals is shown to be asymptotically optimal in terms of information rate.

Keywords and phrases: space-time channels, transmit diversity, information theory, error exponents.

1. INTRODUCTION

Wireless access to data networks such as the Internet is expected to be an area of rapid growth for mobile communications. High user densities will require very high speed low delay links in order to support emerging Internet applications such as voice and video. Even in the low mobility indoor environment, the deleterious effects of fading and the need for very low transmit power combine to cause problems for radio transmissions. Regardless of advanced coding techniques such as turbo-codes [1], channel capacity remains an unmovable barrier (without yet considering inefficiencies introduced by higher layer network overheads). Without changing the channel itself not much can be done. Fortunately, increasing the number of antennas at both the base and mobile stations accomplishes exactly that, resulting in channels with higher capacity. Such systems can theoretically increase capacity by up to a factor equaling the number of transmit and receive antennas in the array [2, 3, 4, 5, 6].

Spatial diversity has been proposed for support of very high rate data users within third generation wideband Code-Division Multiple Access (CDMA) systems such as cdma2000 [7]. Using multiple antennas, these systems achieve gains in link quality and therefore capacity. Classically, diversity has been exploited through the use of either beam-steering (for antenna arrays with correlated elements), or through diversity combining (for independent antenna arrays) [8, 9]. Use of these array processing techniques can achieve any combination of the following: (a) reduction of multiple access interference through the "nulling" of strong interferers. Such techniques are complementary to (and share the mathematical formulations of) multiple-user

receivers such as the decorrelator and MMSE filter [10]; (b) mitigation of fading effects by averaging over the spatial properties of the fading process. This is a dual of interleaving techniques which average over the temporal properties of the fading process; (c) increased link margins by simply collecting more of the transmitted energy at the receiver.

Recently, it has been realized that coordinated use of diversity can be achieved through the use of space-time codes. Rather than relying solely on array processing of uncoded transmissions, forward error correction codes which add redundancy in both the temporal and spatial domains are designed specifically for channels with multiple transmit and receive antennas. There are currently two main approaches to realizing the capacity potential of these channels: *coordinated* space-time codes and *layered* space-time codes.

Coordinated space-time block codes [11, 12, 13] and trellis codes [14, 15, 16, 17, 18] are designed for coordinated use in space and time. The data is encoded using multidimensional codes that span the transmit array. Such codes are efficient for small arrays, and can achieve within 3 dB of the 90% outage capacity rate calculated in [3]. The other approach uses layered space-time codes [19, 20, 21], where the channel is decomposed into parallel single-input, single-output channels.

Increasing the number of antennas and using space-time codes involves both a physical and computational complexity/performance trade-off. Simple schemes such as multi-carrier transmit diversity or array processing techniques, such as maximum ratio combining may be easily implemented. How is their performance limited, when compared to powerful space-time codes? What is the advantage of channel knowledge at the transmitter? In this paper, we attempt

to answer some of these questions from an information theoretic point of view. We do not consider the design of space-time codes.

The paper is organized in the following way. The space-time channel model, along with the Rayleigh fading statistical model are introduced in Section 2.

Under the assumption of fast Rayleigh fading and perfect channel knowledge at the receiver, various channel capacity computations are carried out in Section 3.1. In particular, we find simple bounds on capacity for the case where the transmitter has no channel knowledge. These bounds are used to further investigate the rate of growth of capacity with the number of antennas. An asymptotic (in the number of users) expression for the multiple access space-time channel capacity is shown to be independent of the number of transmit antennas, indicating that for large numbers of users multiuser diversity dominates transmit diversity. We then provide bounds on the outage capacity for slow fading Rayleigh channels. These bounds provide additional support to a conjecture of Telatar concerning the optimal power distribution.

In Section 3.2, we consider perfect channel knowledge at the transmitter. Asymptotic expressions for capacity are given, and the advantage of using channel knowledge at the transmitter is quantified. We find a transmit power threshold beyond which transmitter channel knowledge yields only a marginal capacity advantage. We also consider a suboptimal approach, sometimes referred to as downlink beamforming, in which the transmitter uses only the most powerful eigenchannel. Expressions for capacity and outage capacity are given for this case.

In Section 4, we give asymptotic expressions for a random coding error exponent. In Section 5, we consider transmit diversity (using only a single receive antenna). In particular, we derive a closed form expression for the error exponent and show that the cut-off rate increases logarithmically with the number of antennas. We also show that asymptotically, the use of orthogonal carriers for each transmit antenna is optimal.

Throughout, we make extensive use of results from the theory of random matrices. Of particular interest are Wishart matrices, which model certain statistics of the Rayleigh channel. We make use of both the large systems approach, now familiar in the analysis of multiple-input multiple-output systems, as well as using results for finite dimension matrices. We collect in Appendix A several definitions and known results for reference. Several lemmas used throughout the paper are also proved in Appendix A. Finally, in Appendix B we give for reference definitions and some results concerning hypergeometric functions of matrix argument, which are useful in certain matrix variate distribution calculations.

We use the following notations. The vector $\mathbf{x} \in \mathbb{C}^n$ is a column vector with complex entries x_i, $i = 1, 2, \ldots, n$. Likewise $\mathbf{A} \in \mathbb{C}^{m \times n}$ is a matrix with complex entries A_{ij}, $i = 1, \ldots, m$, $j = 1, \ldots, n$. The superscripts $'$, $*$ and -1 denote, respectively, transposition, Hermitian adjoint, and matrix inverse. By $\mathbf{A} > 0$ we indicate that $\mathbf{A}$ is positive definite. Likewise $\mathbf{A} > \mathbf{X}$ means $\mathbf{A} - \mathbf{X} > 0$, $\det \mathbf{A}$ is the determinant, $\operatorname{tr} \mathbf{A}$ the trace and $\operatorname{etr} \mathbf{A} = \exp(\operatorname{tr} \mathbf{A})$, and $\mathbf{I}_n$ is the $n \times n$ identity matrix. For a random variable X, $\mathrm{E}[X]$ and $\operatorname{var} X$ are its expectation and variance.

2. THE SPACE-TIME CHANNEL

Consider a point-to-point communication link with t transmit antennas and r receive antennas. Throughout the paper we refer to $m = \min\{r, t\}$ and $n = \max\{r, t\}$. We let the match-filtered received signal y_k^j, at antenna $1 \leq j \leq r$ at time k be given by

$$y_k^j = \sum_{i=1}^{t} \sum_{l=0}^{L-1} x_{k-l}^i f_{k,l}^{i,j} + n_k^j, \qquad (1)$$

where x_k^i is the signal transmitted from antenna $1 \leq i \leq t$ at time k; $f_{k,l}^{i,j}$ is the response at antenna j at time l to an impulse applied to antenna i at time k; and n_k^j is a discrete-time white Gaussian noise process, independent over time and antennas, with $\mathrm{E}[n_k^j \bar{n}_k^j] = \sigma^2$ and $\mathrm{E}[n_k^j \bar{n}_{k'}^{j'}] = 0$ for $j \neq j'$ and/or $k \neq k'$. The integer $L > 0$ is the (maximum) delay spread of the fading channel.

The information theoretic analyses presented so far in the literature have concentrated on the flat fading case, we likewise restrict our present scope. For flat fading, the impulse response sequences $f_k^{i,j}$ become scalars. We simplify (1) obtaining a linear algebraic model [2]. At each symbol interval, the received vector $\mathbf{y} \in \mathbb{C}^r$ depends on the transmitted vector $\mathbf{x} \in \mathbb{C}^t$ according to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}. \qquad (2)$$

Element y_j is the matched-filter output from antenna j, while x_i is the signal transmitted from antenna i. The matrix $\mathbf{H} \in \mathbb{C}^{r \times t}$ has as elements H_{ji}, which is the complex gain between transmit antenna i and receive antenna j. The vector $\mathbf{n}$ contains i.i.d. circularly symmetric Gaussian noise samples [22, page 134], $\mathrm{E}[\mathbf{n}\mathbf{n}^*] = \sigma^2 \mathbf{I}_r$. We place the following power constraint on the transmitted signal (independent of t),

$$\mathrm{E}[\mathbf{x}^*\mathbf{x}] \leq P. \qquad (3)$$

We denote the signal-to-noise ratio (SNR) as $\gamma = P/\sigma^2$.

Definition 1 (Rayleigh channel). Randomly select the entries of $\mathbf{H}$, independently of $\mathbf{x}$ and $\mathbf{n}$ as follows. Let H_{ji}, $i = 1, 2, \ldots, t$, $j = 1, 2, \ldots, r$ be i.i.d., zero mean Gaussian with independent real and imaginary parts, each with variance $1/2$.

The complex entries of $\mathbf{H}$ are independent with uniformly distributed phase and Rayleigh distributed magnitudes, modelling the scenario in which the antenna separation is large enough to ensure independent fading. A matrix $\mathbf{H}$ selected according to Definition 1 is matrix normal distributed (i.e., $\mathbf{H} \sim \mathcal{N}(r, t)$ according to Definition A.1). The matrix product $\mathbf{H}\mathbf{H}^*$ will be of great interest and has

the Wishart distribution, $\mathbf{H}\mathbf{H}^* \sim \mathcal{W}(r, t)$ by Definition A.2. Appendix A describes some useful properties of Wishart matrices that we will use throughout the paper. Appendix B summarizes some results concerning hypergeometric functions of matrix argument that are required for certain calculations concerning Wishart matrices.

The channel matrix may either be fixed, or time varying. In the time varying case, we can further distinguish between fast variations (compared to the data transmission rate), which allow us to compute average information rates, and slowly varying channels, for which outage probability is more appropriate. We assume throughout that the channel matrix is known perfectly at the receiver. We consider both the case where the channel is unknown at the transmitter and the case where it is known perfectly. The theory for the case in which the channel matrix is unknown at both terminals is still in development. It is known that the capacity gains to be achieved depend upon the coherence time of the channel [23, 24, 25]. Special space-time codes have also been designed to operate in the absence of channel state information [26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36].

3. CAPACITIES

In this section, we try to gain some insight into the scaling of capacity with the number of transmit and receive antennas. We are particularly interested in quantifying the capacity increase available through the availability of channel knowledge at the transmitter. We assume throughout that the receiver has perfect channel knowledge.

Consider the channel (2), with average transmit power constraint (3). Restrict x_i, $i = 1, 2, \ldots, t$ to be i.i.d. zero mean Gaussian variables, each with variance P/t. Then the maximum transmission rate is given by [2] as

$$I(\gamma, r, t, \mathbf{H}) = \log_2 \det \left(\mathbf{I}_r + \frac{\gamma}{t}\mathbf{H}\mathbf{H}^* \right) \text{ bit/channel use.} \quad (4)$$

This is simply an achievable rate, rather than the capacity, since no optimization over the source distribution is performed. Such optimization would require knowledge of $\mathbf{H}$ at the transmitter. In the case of fast Rayleigh fading according to Definition 1, the channel capacity is given by $C(\gamma, r, t) = \mathrm{E}[I(\gamma, r, t, \mathbf{H})]$ and is achieved for the i.i.d. Gaussian input distribution just described.

Given perfect channel knowledge at the transmitter, channel capacity takes a different form, and is achieved by water filling on the parallel orthogonal channels defined by the singular vectors of $\mathbf{H}$, see, for example, [37, Theorem 8.5.1]. Let $\lambda_1, \ldots, \lambda_m$ be the m nonzero eigenvalues of $\mathbf{H}\mathbf{H}^*$. Then the capacity is given parametrically [2] by

$$C_{\mathrm{WF}} = \sum_{i:\lambda_i^{-1} \leq \mu} \log \mu \lambda_i, \quad (5)$$

$$P = \sum_{i:\lambda_i^{-1} \leq \mu} \mu - \lambda_i^{-1}. \quad (6)$$

Note that in the case of transmitter channel knowledge, the

concept of forming an average capacity is somewhat less straightforward, since the transmitter has the additional option of optimizing the power allocation over time as well as over the eigenvalues (while maintaining the required average power restriction). We do not consider average capacities for this case.

3.1. Channel unknown at transmitter

Now concentrate on the case in which only the receiver perfectly knows the time varying Rayleigh channel matrix $\mathbf{H}$. The capacity of the fast time varying Rayleigh channel was found by Telatar [2]. We rewrite this capacity using an alternative form in the following theorem.

Theorem 1. *At each symbol interval, let $\mathbf{H}$ be selected according to Definition 1, and assume that $\mathbf{H}$ is known to the receiver, but not at the transmitter. Let $m = \min\{r, t\}$ and $n = \max\{r, t\}$. Then the capacity of the channel (2) is given by*

$$C(\gamma, t, r) = \mathrm{E}\left[\log \det \left(\mathbf{I} + \frac{\gamma}{t}\mathbf{H}\mathbf{H}^* \right)\right]$$

$$= \frac{(m-1)!}{(n-1)!} \int_0^\infty \log \left(1 + \frac{\gamma}{t}x \right) e^{-x} x^{n-m} \quad (7)$$

$$\times \left[\left(L_{n-m+1}^{m-1}(x) \right)^2 - L_{n-m+1}^m(x) L_{n-m+1}^{m-2}(x) \right] dx,$$

where L_k^α is the generalized Laguerre polynomial of order k [38]. Capacity is achieved for $\mathbf{x}$ circularly symmetric zero mean complex Gaussian vector with $\mathrm{E}[\mathbf{x}\mathbf{x}^*] = \mathbf{I}_t P/t$.

Proof. The expression follows from [2, Theorem 2], using [39, equation (8.974.1)]. □

Although straightforward to numerically compute, (7) does not give much information about the scaling of capacity with r and t. Using limiting arguments, it is by now well known, and often quoted that in the case of $r = t = k$ that $C(\gamma, k, k)$ is linear in k. We now give convenient upper and lower bounds to (7).

Theorem 2. *The capacity (7) is upper and lower bounded according to*

$$C(\gamma, t, r) \leq m \log \frac{\gamma}{t} + \log m! + \log \left[L_m^{n-m}\left(-\frac{t}{\gamma} \right) \right], \quad (8)$$

$$C(\gamma, t, r) \geq m \log \frac{\gamma}{t} + \sum_{i=0}^{m-1} \psi(n - i), \quad (9)$$

where ψ is Euler's digamma function [39, equation (8.36)].

Proof. Inequality (8) follows directly from Jensen's inequality and Theorem A.4. The lower bound follows from $\det(\mathbf{I}+\mathbf{A}) \geq \det \mathbf{A}$ for Hermitian $\mathbf{A}$ and Lemma A.2. □

It is clear from the method of proof that the upper and lower bounds (8) and (9) are tight as $\gamma \to \infty$. In fact, expanding the digamma as $\psi(z) = \log z - (2z)^{-1} - \theta(12z)^{-2}$, for

some $0 < \theta < 1$, we may further approximate (9) by

$$m \log \frac{\gamma}{t} + \log m! + \log \binom{n}{m} - c, \qquad (10)$$

where $c \approx 0.577$ is the Euler-Mascheroni constant [40, Section 3.1.9.4]. Now

$$\lim_{\gamma \to \infty} L_m^{n-m} \left(-\frac{t}{\gamma} \right) = \binom{n}{m}, \qquad (11)$$

hence we expect polynomial tightness of the upper and lower bounds in γ.

The lower bound is however poor when simultaneously $\gamma \ll t$ and $t \approx r$. This problem may be somewhat offset by noting that C is monotonic in both n and m, and hence the lower bound near $n = m$ may be improved by using smaller values of either n or m.

It is therefore not unreasonable to use (8) as an approximation for the purposes of determining the rate of growth of capacity with t and r, since for sufficient SNR, the upper and lower bounds will display the same gross behavior. Three cases are of interest: (a) fix t and let $r \to \infty$, (b) fix r and let $t \to \infty$, and (c) fix t/r and let both $t, r \to \infty$.

Considering case (a), we hold t constant and send r large to see that

$$C(\gamma, t, r) \approx t \log \frac{\gamma}{t} + \log \frac{r!}{(r-t)!} \longrightarrow t \log \frac{\gamma}{t} + t \log r \quad (12)$$

since $r!/(r-t)! < r^t$ (the convergence indicated is in the sense of the ratio of the left- and right-hand sides approaching unity). Hence an asymptotically logarithmic increase in r results.

For case (b), we hold r constant and send t large, to obtain

$$C(\gamma, t, r) \longrightarrow r \log \frac{\gamma}{t} + r \log t = r \log \gamma, \qquad (13)$$

where once again, convergence is understood in terms of the ratio of the two quantities. As expected, this result indicates that increasing t serves only to eliminate the effect of fading, but provides no further gains.

Finally, consider case (c). It is already well known that if $t = r$, an asymptotically linear increase in capacity is observed. Now let $t/r \to \beta \leq 1$. Then

$$\frac{1}{t} C(\gamma, t, r) \longrightarrow \log \gamma + \log \beta^{-1}, \qquad (14)$$

which demonstrates that as long as t and r increase at the same rate, a linear capacity increase occurs. The proportion β serves only to scale the SNR and set the slope of the linear increase. Finally, $t/r \to \beta > 1$ results in $C/r \to \log \gamma$, which is independent of β. Asymptotically, there is no benefit from increasing t beyond r.

3.1.1 Multiple users

Now consider a K user multiple access channel, in which the channel from each user to the common receiver is an independent space-time channel. We suppose each user has t transmit antennas and that there are r receive antennas. At each symbol interval, user $k = 1, 2, \ldots, K$ transmits a vector $\mathbf{x}[k] \in \mathbb{C}^t$ by sending $\sqrt{P_k/t}\, x_i^{(k)}$ over antenna $i = 1, 2, \ldots, t$ such that $\mathrm{E}[\mathbf{x}[k]^* \mathbf{x}[k]] \leq 1$, corresponding to a transmit power constraint P_k. The received vector $\mathbf{y} \in \mathbb{C}^r$, observed at the output of the r receive antennas is $\mathbf{y} = \mathbf{H}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{n}$, where $\mathbf{x} = (\mathbf{x}[1], \mathbf{x}[2], \ldots, \mathbf{x}[K])$ and $\mathbf{H} \in \mathbb{C}^{r \times Kt}$ contains the fading coefficients. For a given integer $1 \leq k \leq K$, the element $H_{j,kt+i-1}$ is the complex gain between transmit antenna i of user k and receive antenna j. The $Kt \times Kt$ diagonal matrix $\mathbf{P} = \mathrm{diag}(P_1\mathbf{I}_t, P_2\mathbf{I}_t, \ldots, P_K\mathbf{I}_t)/t$. The vector $\mathbf{n} \in \mathbb{C}^r$ contains i.i.d. zero-mean Gaussian noise samples, $\mathrm{E}[\mathbf{n}\mathbf{n}^*] = \sigma^2\mathbf{I}$. We assume that $\mathbf{H}$ is known at the receiver, but not at the transmitter.

Passing to the band-limited case via the usual arguments, suppose that the total bandwidth available for transmission is KW, where W is some fixed quantity (KW should be less than the coherence bandwidth of the channel for the flat fading assumption to be realistic). Then the sum capacity of this channel is given by

$$C(\mathbf{P}, \mathbf{H}) = KW \log \det \left(\mathbf{I}_r + \frac{\mathbf{HPH}^*}{KW\sigma^2} \right) \text{ bit/s.} \qquad (15)$$

If $\mathbf{H}$ is independently selected at each symbol interval according to Definition 1, the capacity of the resulting channel is found by taking the expectation of (15) with respect to the random matrix $\mathbf{H}$.

Now as $K \to \infty$, $\mathbf{HPH}^* \to \mathbf{I}_r \sum_k P_k$ a.s., which leads directly to the following theorem[1] which shows that for sufficiently large numbers of users, the capacity gain from using multiple transmit antennas disappears.

Theorem 3. *For $k = 1, 2, \ldots$, let $\mathbf{H}_k \in \mathbb{C}^{r \times Kt}$ be chosen according to Definition 1, and let $0 \leq P_k < \infty$ define an arbitrary but nonrandom sequence of nonnegative real numbers. Then the sum capacity* (15), *normalized by the number of users tends to a fixed limit,*

$$\frac{C(\mathbf{H}_k)}{rK} \longrightarrow W \log \left(1 + \frac{\sum P_k}{KWN_0} \right) \quad \text{a.s.} \qquad (16)$$

which is independent of t.

Observe that a linear capacity increase with r is obtained, independent of the ratio of r and t. Also note that this theorem refers only to the sum capacity. Benefits may still be gained through the use of multiple transmit antennas for achieving other points in the capacity region. Note, however, that if it is the overall constraint that is tight, any increase in the capacity of a subset of users, obtained by increasing the

---

[1] The idea for this theorem arose out of discussions with S. Hanly at Melbourne University.

number of transmit antennas comes at the price of decreasing the capacity of the users in the complementary set.

3.1.2   Outage probability

So far, we have concentrated on the fast time varying Rayleigh channel. Now consider the slowly varying channel, for which outage probability has been proposed as the measure of interest [2, 3]. For a given information rate R and $t \times t$ diagonal matrix $\mathbf{Q} > 0$, define $P_{\text{out}}(R, \mathbf{Q}) = \Pr(\log \det(\mathbf{I}_r + \mathbf{HQH}^*) < R)$. For power constraint P, Telatar [2] gives

$$P_{\text{out}}(R, P) = \inf_{\mathbf{Q}:\mathbf{Q}>0,\text{tr}\,\mathbf{Q}\leq P} P_{\text{out}}(R, \mathbf{W}). \qquad (17)$$

We consider $\mathbf{H}$ chosen once for all time according to our Rayleigh channel definition. Now the distribution of $\log \det(\mathbf{I} + \mathbf{W})$ is unknown. Apart from the mean (Theorem A.4), the moments of this quantity do not have a simple expression (although in limiting cases they are known [41]). The following theorem demonstrates the type of difficulty one encounters.

Theorem 4. *The outage probability is upper bounded*

$$P_{\text{out}}(R, P) \leq \inf_{\substack{s>0 \\ \mathbf{Q}>0,\text{tr}\,\mathbf{Q}\leq P}} {}_2F_0\,(s,n;\mathbf{Q})\,e^{sR}, \qquad (18)$$

where $_mF_n\,(\cdot;\cdot;\cdot)$ is a hypergeometric function of matrix argument as defined in Appendix B.

Proof. Now $\mathbf{W} \triangleq \mathbf{HQH}^* \sim \mathcal{W}(m, n, \mathbf{Q})$ and

$$\begin{aligned} P_{\text{out}}(R, \mathbf{Q}) &= \Pr\,(\,\log \det(\mathbf{I} + \mathbf{W}) < R) \\ &\leq \mathrm{E}\,\big[\det(\mathbf{I} + \mathbf{W})^{-s}\big]\,e^{sR}, \end{aligned} \qquad (19)$$

for any $s > 0$ by Chernoff's bound. Now consider

$$\begin{aligned} \mathrm{E}\,\big[\det(\mathbf{I} + \mathbf{W})^{-s}\big] &= \int p(\mathbf{W})\det(\mathbf{I} + \mathbf{W})^{-s}d\mathbf{W} \\ &= \frac{\det \mathbf{Q}^{-n}}{\Gamma_m\,(n)} \int_{\mathbf{W}>0} {}_sF_0\,(-\mathbf{W};\text{etr})\,(\,-\mathbf{Q}^{-1}\mathbf{W}) \\ &\qquad\qquad\qquad \times \det \mathbf{W}^{n-m}d\mathbf{W}, \end{aligned} \qquad (20)$$

where the second line results from (B.7) and the Wishart density (A.3). The result is then obtained through application of the Laplace transform recursion for the hypergeometric functions (B.5). $\qquad\square$

The following is a more computable bound.

Theorem 5. *The outage probability is upper bounded*

$$P_{\text{out}}(R, P) \leq \Pr\,\big(X < e^{R-m\log(P/t)}\big), \qquad (21)$$

where $X \sim \prod_{i=0}^{m-1} X_i$, X_i chi-squared with $2(n - i)$ degrees of freedom.

Proof. The proof follows from $\det(\mathbf{I} + \mathbf{A}) > \det \mathbf{A}$ and the distribution of $\det \mathbf{W}$ for $\mathbf{W} \sim \mathcal{W}(m, n, \mathbf{Q})$, (A.3), which gives

$P_{\text{out}}(R, \mathbf{Q}) \leq \Pr(X < e^R \det \mathbf{Q}^{-1})$. For fixed m, the optimal $\mathbf{Q}$ is $\mathbf{Q} = \arg\min_{\tilde{\mathbf{Q}}>0,\text{tr}\,\tilde{\mathbf{Q}}=P} \det \tilde{\mathbf{Q}} = P\mathbf{I}/m$, which leaves only a minimization over m. $\qquad\square$

This upper bound lends additional support to the conjecture made in [2] that the outage probability is minimized by using a uniform power distribution over some subset of the antennas.

3.2.   Channel known at transmitter

Now consider the case when both the receiver and transmitter have perfect channel knowledge. We are interested in two cases, firstly we compute a large systems limit for the waterfilling capacity C_{WF} and establish the optimal power profile. Secondly, we consider a simple suboptimum case, in which transmitter selection diversity is used, that is, the transmitter simply transmits using all its power along the maximum eigenvector. In this second case, we compute the large systems capacity and the outage probability. Without loss of generality, we let $\sigma^2 = 1$, that is, $\gamma = P$. All of the results in this section will be obtained as large systems limits and rely on the convergence of the Wishart spectrum as $m, n \to \infty$ in such a way that $m/n \to \beta \leq 1$.

Given a particular eigenvalue distribution, the capacity may be found using (5). It is however a well-known fact that for a wide class of random matrices, the normalized eigenvalue distribution converges to a nonrandom limit as the matrices are taken large. We use this fact to determine the large systems capacity limit.

Theorem 6. *Let $m \leq n$ approach infinity such that $m/n \to \beta$. Then for a sequence of matrices $\mathbf{H}_{m,n}$ selected according to Definition 1*

$$\frac{C_{\text{WF}}}{m} \longrightarrow \int_{\max\{a(\beta),1/\mu\}}^{b(\beta)} \log \mu\lambda\,dF(\lambda), \qquad (22)$$

$$P \longrightarrow \int_{\max\{a(\beta),1/\mu\}}^{b(\beta)} (\mu - \lambda^{-1})\,dF(\lambda), \qquad (23)$$

where $F(\lambda)$, $a(\beta)$, and $b(\beta)$ are defined in Theorem A.7.

Proof. The proof follows from [37, Chapter 8] (using the eigenvectors of $\mathbf{H}$ as kernels, rather than the Fourier kernel), and the convergence in distribution of the empirical eigenvalue distribution to $F(\lambda)$ [42]. $\qquad\square$

According to this water-filling approach, if the transmit power P is such that $\mu^{-1} > a(\beta)$, the transmitter uses only a fraction $F(\mu^{-1})$ of the eigenchannels. If $\mu^{-1} \leq a(\beta)$ it transmits using all the parallel channels. In this latter case, we can find a closed form expression for the capacity.

Corollary 1. *Retaining the limiting regime from Theorem 6, let P be such that $\mu^{-1} \leq a(\beta)$. Then*

$$\frac{C_{\text{WF}}}{m} \longrightarrow \log \left(P + \frac{1}{1-\beta} \right) + \frac{1-\beta}{\beta} \log \left(\frac{1}{1-\beta} \right) - 1. \quad (24)$$
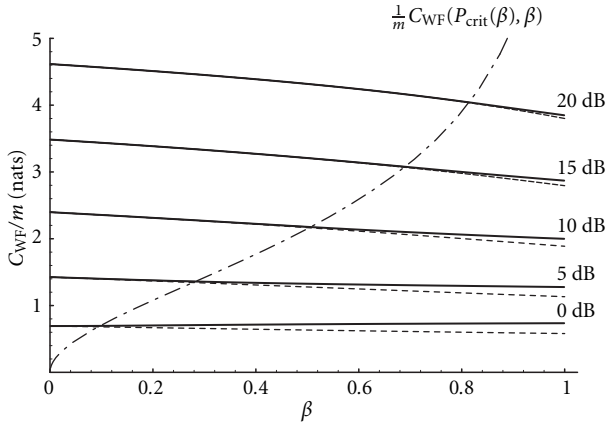
FIGURE 1: Asymptotic water-filling capacity.

Proof. Under the assumption $\mu^{-1} \leq a(\beta)$, the limit (22) becomes

$$\frac{C_{\mathrm{WF}}}{m} \longrightarrow \log\mu + \int_{a(\beta)}^{b(\beta)} \log\lambda \, dF(\lambda)$$
$$= \log\mu + \frac{1-\beta}{\beta}\log\frac{1}{1-\beta} - 1, \qquad (25)$$

where the first line is due to integrating over the entire support of λ and the second line is from [42, Corollary 5.1]. It remains to find μ. Now from (23),

$$P \longrightarrow \mu - \frac{1}{2\pi\beta}\int_{a(\beta)}^{b(\beta)} \frac{\sqrt{(\lambda - a(\beta))(b(\beta) - \lambda)}}{\lambda^2} \, d\lambda. \qquad (26)$$

This integral is of the form [39, equation (2.267.2)] and after straightforward calculation is found to be $1/(1-\beta)$. □

The critical value of μ beyond which Corollary 1 holds is $\mu_{\mathrm{crit}} = (1 - \sqrt{\beta})^{-2}$ corresponding to a critical SNR given by $P_{\mathrm{crit}}(\beta) = (1 - \sqrt{\beta})^{-2} - (1 - \beta)^{-1}$.

Figure 1 shows the normalized asymptotic water-filling capacity C_{WF}/m plotted against β. The dot-dashed line shows $C_{\mathrm{WF}}(P_{\mathrm{crit}}(\beta), \beta)$. Capacity for values of β to the left of this line may be calculated using Corollary 1. Values to the right are calculated via numerical integration of (22). The solid lines are C_{WF}/m for fixed values of P. Shown for comparison as the dashed lines are the corresponding values of C, in the absence of transmitter channel knowledge.

Note that for β small enough such that $P > P_{\mathrm{crit}}$ (left of dot-dashed line), or for large SNR that there is only a small advantage in C_{WF}/m to be obtained from water-filling. This is investigated further in Figure 2, where the ratio of the water-filling capacity to the corresponding limiting value of C is shown for $P = P_{\mathrm{crit}}(\beta)$. This loss will be even less for higher power levels (to the left of the dot-dashed line on Figure 1). The relative difference, upper bounded by approximately 1.022 at $\beta \approx 0.1$ is preserved through the scaling by m. This result indicates that in certain scenarios that the absence
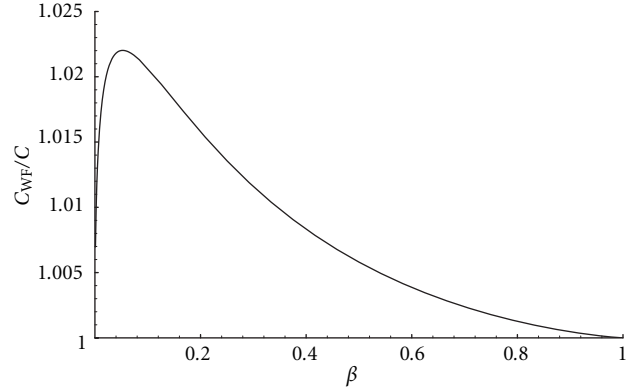
FIGURE 2: Relative capacity gain from water-filling along the locus of $P = P_{\mathrm{crit}}$.
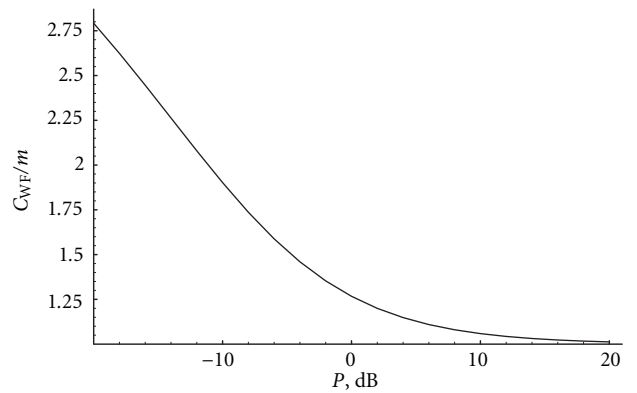
FIGURE 3: Relative capacity gain from water-filling at $\beta = 1$.

of transmitter channel knowledge may not cost too much. For $P > P_{\mathrm{crit}}$, the capacity advantage for using transmitter channel knowledge is at most about 2%.

For $P < P_{\mathrm{crit}}$, there is more to gain. Figure 3 shows C_{WF}/C versus P in dB for $\beta = 1$ (which maximizes the gain). We see that as P decreases the gain increases. For example, capacity is increased by approximately 25% at $P = 0$ dB and is approximately doubled for $P = -10$ dB.

We conclude our examination of transmitter channel knowledge by considering a suboptimal method, sometimes referred to as downlink beamforming, which has been proposed for 3G communications [7, 43, 44]. In this method, only the largest eigenvector is used for transmission. Theorem A.6 yields directly.

Theorem 7. *Let $m/n \to \beta \leq 1$, then*

$$C(\gamma) \longrightarrow \log\left(1 + \gamma m\left(1 + \sqrt{\beta}\right)^2\right) \quad \text{a.s.} \qquad (27)$$

This indicates that by transmitting along the maximum eigenvector, performance superior to the unfaded AWGN channel is obtained. Note however that the capacity increase is only logarithmic with $\min\{t, r\}$, as compared to the linear capacity increase available through the use of all transmit antennas, even in the absence of channel knowledge at the

transmitter. For finite systems, the following theorem may be used to compute outage probabilities for the selection diversity approach.

Theorem 8. *Let $\alpha = (e^R - 1)/P$, then*

$$P_{\text{out}}(R, P) = \frac{\Gamma_m(m)}{\Gamma_m(n+m)} \alpha^{nm} {}_1F_1(n; n+m; -\alpha\mathbf{I}). \quad (28)$$

Proof. From Theorem A.5, $P_{\text{out}}(R, P) = \Lambda(\alpha)$.   □

4. ERROR EXPONENTS

In addition to maximum information transmission rates, it is interesting to consider the error exponent [37] for space-time channels. The error exponent gives some indication of how difficult it may be to achieve a certain bit error rate. In this section, we derive a limiting expression for the error exponent. Telatar [2] gives the following lower bound to the error exponent.

Lemma 1. *The probability of error averaged over all randomly selected (N, NR) block codes is bounded*

$$P_e \le \exp\left[-N \max_{0 \le \rho \le 1} (E_0(\rho) - \rho R)\right], \quad (29)$$

where

$$E_0(\rho) = -\log \mathrm{E}\left[\det\left(\mathbf{I}_r + \frac{\gamma}{t(1+\rho)}\mathbf{HH}^*\right)^{-\rho}\right]. \quad (30)$$

Reliability functions for fading channels with correlated multiple antennas have also been considered in [45] for $t = 1$, see also [46].

The following theorem uses the theory of random determinants to give a limiting expression for the error exponent.

Theorem 9. *Let t/r be fixed. The following limiting relation holds.*

$$\sup_{t \to \infty} E_0(\rho) = \mu(\rho)\rho - \frac{\sigma^2(\rho)\rho^2}{2}, \quad (31)$$

where, for convenience $\mathbf{A}(\rho) = \mathbf{I} + \gamma/(t(1+\rho))\mathbf{HH}^*$ and

$$\mu(\rho) = \lim_{t \to \infty} \mathrm{E}[\log \det \mathbf{A}] = \lim_{t \to \infty} C\left(\frac{\gamma}{1+\rho}, t, r\right), \quad (32)$$

$$\sigma^2(\rho) = \lim_{t \to \infty} \mathrm{var} \log \det \mathbf{A}.$$

Proof. We can rewrite $E_0(\rho)$ as

$$E_0(\rho) = -\log \mathrm{E}\left[\exp(-\rho x)\right], \quad (33)$$

where the random variable $x = \log \det \mathbf{A}$. From [47], we have that x is asymptotically normal with mean $\mu(\rho)$ and variance $\sigma^2(\rho)$. The right-hand side of (33) is simply the negative logarithm of the moment generating function $M(s) = \mathrm{E}[e^{sx}]$,

which is well known for the normal distribution to be given by

$$M(s) = \exp\left(\mu s + \frac{s^2\sigma^2}{2}\right). \quad (34)$$

Taking the supremum on both sides of (33) and using (34) gives the result.   □

$\mu(\rho)$ may be calculated using [2, equation (13)], or may be lower bounded using [48, Theorem 3].

5. TRANSMIT DIVERSITY

In this section, we consider transmit diversity, $r = 1, t > 1$. From (4), it is easy to see that

$$C = \mathrm{E}\left[\log\left(1 + \frac{\gamma}{2t}X\right)\right], \quad (35)$$

where X is a chi-squared random variable with $2t$ degrees of freedom.[2] As t increases, $C \to \log(1+\gamma)$ (since $X \to 2t$ by the law of large numbers), which is as if all the power were transmitted over a single, nonfaded link. By increasing the number of antennas, the effects of the fading may be completely removed. In practice, only a small number of transmit antennas (typically two or three for a wide range of γ) are required (see [2, Figure 3]).

Increasing the number of transmit antennas however has the desirable effect of increasing the error exponent, as we now show.

Theorem 10. *Let $r = 1$, then*

$$E_0(\rho) = -\log\left[\left(\frac{t(1+\rho)}{\gamma}\right)^t \Psi\left(t, t+1-\rho; \frac{t(1+\rho)}{\gamma}\right)\right], \quad (36)$$
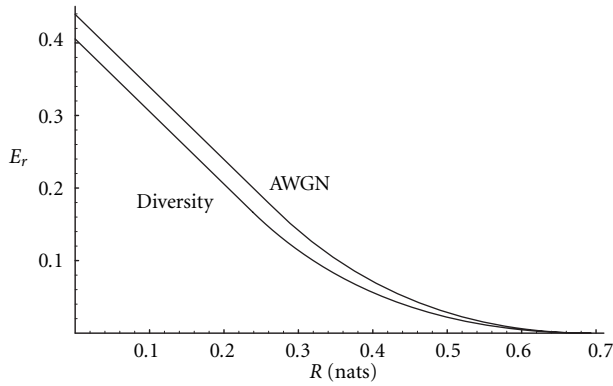
where Ψ is the confluent hypergeometric function [39, Section 9.2]. Furthermore, letting $C_{\mathrm{G}}(\gamma) = \log(1+\gamma)$ be the Gaussian channel capacity with SNR γ we have the following limiting expression for the error exponent:

$$\lim_{t \to \infty} E_r = \begin{cases} C_{\mathrm{G}}\left(\dfrac{\gamma}{2}\right) - R, & \text{if } 0 \le R \le C_{\mathrm{G}}\left(\dfrac{\gamma}{2}\right) - \dfrac{\gamma}{2(\gamma+2)}, \\[2mm] \rho C_{\mathrm{G}}\left(\dfrac{\gamma}{1+\rho}\right) - \rho R(\rho), & \text{otherwise,} \end{cases} \quad (37)$$

where for the second case, $C_{\mathrm{G}}(\gamma/2) - (\gamma/2(\gamma+2)) < R \le C_{\mathrm{G}}(\gamma)$ we have a parametric representation via $0 \le \rho \le 1$ and

$$R(\rho) = C_{\mathrm{G}}\left(\frac{\gamma}{1+\rho}\right) - \frac{\gamma\rho}{(1+\rho)(1+\rho+\gamma)}. \quad (38)$$

---

[2]The factor 2 in the denominator accounts for the fact that the real and imaginary parts of each independent Gaussian variate have variance 1/2.

FIGURE 4: Error exponent, $\gamma = 0$ dB.

FIGURE 5: Cut-off rate versus number of antennas.

Proof. From (30), we have

$$E_0(\rho) = -\log E\left[\left(1 + \frac{\gamma}{2t(1+\rho)}X\right)^{-\rho}\right], \qquad (39)$$

where X is a chi-squared random variable with $2t$ degrees of freedom. Using an integral representation for Ψ [39, equation (9.211.4)], we get (36) after some algebraic manipulations.

As t increases, $X \to 2t$ in probability we have

$$\lim_{t\to\infty} E_0(\rho) = \rho C_G\left(\frac{\gamma}{1+\rho}\right), \qquad (40)$$

which, after some manipulations (along the lines of [37, Section 7.4]) proves (37). □

Figure 4 compares the error exponents for the additive white Gaussian noise channel [37, equation (7.4.33)] and the limiting $t \to \infty$ transmit diversity channel with Rayleigh fading (37). The signal-to-noise ratio is zero dB. Further numerical investigations have shown that the gap in the error exponents only disappears as $\gamma \to 0$. It is interesting to note that although $t \to \infty$ results in $C \to C_G(\gamma)$ there is an inherent complexity penalty to be paid (at least in terms of the error exponent).

Closed form determination of E_r for finite t using (36) seems difficult, although numerical methods can be used. We therefore concentrate on the cut-off rate, $R_0 = E_0(1)$, which gives the error exponent up the critical rate, and a lower bound beyond that. Substituting $\rho = 1$ into the expressions of Theorem 10, we obtain the following corollary.

Corollary 2. *The cut-off rate is given by*

$$R_0(t, \gamma) = t \log \frac{\gamma}{2t} - \frac{2t}{\gamma} - \log \Gamma\left(1 - t, \frac{2t}{\gamma}\right), \qquad (41)$$

where $\Gamma(\alpha, z) = \int_z^\infty x^{\alpha-1} e^{-t}\, dt$ is the incomplete gamma function. Furthermore,

$$\lim_{t\to\infty} R_0(t, \gamma) = \log\left(1 + \frac{\gamma}{2}\right). \qquad (42)$$

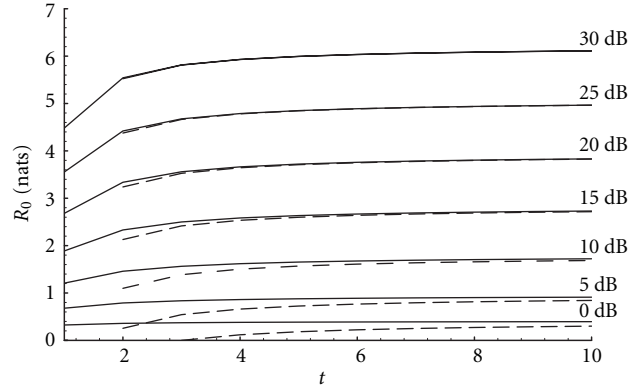*Proof.* Equation (41) is obtained from (39) via the identity $\Psi(\alpha, \alpha; z) = e^z \Gamma(1 - \alpha, z)$. □

For $t > 1$ the cut-off rate may be lower bounded as follows:

$$R_0 \geq \log\left[\left(1 + \frac{\gamma}{2}\right)\left(1 - \frac{1}{t}\right)\right]. \qquad (43)$$

This bound is tight as $t \to \infty$ or $\gamma \to \infty$. Hence R_0 increases at least as the logarithm of $t/(t-1)$. Figure 5 shows R_0 (solid line) and the lower bound (43) (dashed) plotted versus t for various SNR.

5.1. *Orthogonal transmit diversity*

It is interesting to consider the use of mutually orthogonal transmit waveforms for each transmit antenna,[3] as proposed for cdma2000 [7, Section 3.2.1.1.5]. We continue to consider transmit diversity only ($r = 1$), and assume codebooks with i.i.d. circularly symmetric complex Gaussian letters. This means that for a given channel, transmission rates are upper bounded by I as given by (4). If the channel is random, according to Definition 1, the transmission rate is upper bounded by C as given by (7).

Examples of orthogonal waveforms include (a) time division, in which the scalar input is successively transmitted from each antenna at disjoint time intervals; (b) frequency division (multi-carrier transmit diversity [7, Section 3.2.1.1.5.1]), in which the signal for each antenna is modulated by an orthogonal frequency carrier; and (c) code division (direct-sequence transmit diversity [7, Section 3.2.1.1.5.2]), where each antenna is modulated by (for example) an orthogonal Walsh code.

It is easy to see (e.g., through application of the data processing theorem [49]) that restriction to orthogonal carriers can only decrease the maximum possible mutual information, when compared to the use of unconstrained space-time codes. In this section, we show that use of a sufficiently large

---

[3]What counts is that the waveforms are orthogonal at the receiver, for example, flat fading and no Doppler spread.

number of transmit antennas makes this penalty insignificant. This may be interesting since use of orthogonal carriers should result in simple receiver structures [11, 12]. Narula et al. [6] have found the average mutual information (assuming a Gaussian codebook) in the case of orthogonal signaling.

Theorem 11. *Let* **H** *be a fixed matrix, known at the receiver. Then the average mutual information under the assumption of orthogonal transmit waveforms is given by*

$$I_\perp = \frac{1}{t} \sum_{i=1}^{t} \log_2 \left(1 + \frac{|H_i|^2 P}{\sigma^2} \right). \qquad (44)$$

Furthermore, for this channel $I_\perp \leq I$ with equality if and only if $|H_i| = |H_j|$ for all $i, j = 1, 2, \ldots, t$.

Hence for any given channel, the use of either multicarrier transmit diversity, or orthogonal direct-sequence transmit diversity, while convenient from an implementation point of view, incurs a loss, when compared to space-time coding. Since this is true of any given channel, the averages over random channel selection (e.g., Definition 1) also obey $E[I_\perp] \leq E[C]$. The following result gives an expression for $E[I_\perp]$ and shows its limiting behavior for large signal-to-noise ratios.

Theorem 12. *Let* **H** *be randomly selected every symbol interval according to Definition 1, and let* **H** *be known at the receiver. Then the maximum transmission rate, under the assumption of orthogonal transmit diversity is given by $E[I_\perp] = -e^{1/\gamma} \operatorname{Ei}(-1/\gamma)$ nats/channel use, where $\operatorname{Ei}(x) = \int_{-x}^{\infty} e^{-x}/x \, dx$ is the exponential integral function. Furthermore,*

$$\lim_{\gamma \to \infty} \log\left(1 + \gamma\right) - E\left[I_\perp\right] = c, \qquad (45)$$

where $c \approx 0.577$ is the Euler-Mascheroni constant. $E[I_\perp] \to \log(1 + \gamma)$ in the sense that their ratio tends to unity.

It is easy to verify that for $r = 1$ and increasing t, $C \to \log(1 + \gamma)$. Hence even though a rate penalty is incurred for using orthogonal transmit diversity, this penalty disappears as the signal-to-noise ratio and the number of antennas is increased.

Proof. From (44), Definition 1, and the linearity of the expectation operator, we have that $E[I_\perp] = E[\log(1 + \gamma X)]$, where X is an exponentially distributed[4] random variable [22]. Hence

$$E\left[I_\perp\right] = \int_0^\infty \log(1 + \gamma x) e^{-x} \, dx$$
$$= -e^{1/\gamma} \operatorname{Ei}\left(-\frac{1}{\gamma}\right), \qquad (46)$$

according to [39, equation (4.337(2))], which is the first part of the theorem. Continuing, we note from [39, equation

---

[4]Or equivalently, a χ_2^2 variable.

(8.214.1)] that

$$\operatorname{Ei}(x) = c + \log(-x) + \sum_{k=1}^{\infty} \frac{x^k}{k \cdot k!}, \quad x < 0. \qquad (47)$$

Using this identity, and performing some simple rearrangements, we can write

$$E\left[I_\perp\right] - e^{1/\gamma} \log(\gamma) = -e^{1/\gamma} \left[c + \sum_{k=1}^{\infty} \frac{(-1/\gamma)^k}{k \cdot k!} \right]. \qquad (48)$$

Taking limits as $\gamma \to \infty$ on both sides yields the desired result, since $1/\gamma \to 0$, forcing the exponential term to unity, and each element inside the sum to zero. ☐

6. CONCLUSION

In this paper, we have given several calculations regarding the capacity and error exponents of multiple antenna Gaussian channels. We used simple bounds to determine the scaling of capacity with the number of antennas. In particular, we found that if the number of transmit and receive antennas are increased proportionately, a linear increase is obtained, and the rate of this increase was determined. We further discovered that asymptotically, there is no capacity advantage for the number of transmit antennas to exceed the number of receive antennas. In the case that the number of transmit antennas is held constant, we found that capacity increases logarithmically with the number of receive antennas. In the reverse situation, holding the number of receive antennas fixed, increasing the number of transmit antennas results in convergence to a fixed capacity. A simple calculation showed that for multiple-access space-time channels that capacity is dominated by multiuser diversity, rather than space diversity. For large numbers of users, there is no total capacity advantage in using multiple antennas. We also gave bounds on the outage probability for nonergodic channels and discovered that a constant power allocation over some subset of antennas minimized one of these bounds.

Large systems limits were calculated for the capacity in the case of perfect channel knowledge at the transmitter. For power levels above a certain critical threshold, this limit can be easily calculated in closed form. We found numerically that for power levels greater than this same critical threshold that there is very little to be gained by the water-filling approach, although for smaller powers we found that capacity may be increased significantly. We also considered briefly the use of selection diversity at the transmitter and found that this simple strategy outperforms a single antenna link with no fading.

The remainder of the paper was devoted to calculation of error exponents. In particular, we found a limiting expression for $E_0(\rho)$, and a closed form expression for the case of transmit diversity ($r = 1$). We further showed that for $r = 1$ the cut-off rate increases logarithmically with the number of transmit antennas. Although capacity is near to maximum for a small number of antennas (typically 2 or 3), use of additional antennas improves performance through increase of

the error exponent. We have also shown that use of a sufficiently large number of transmit antennas offsets the rate penalty suffered due to orthogonal transmissions from each antenna, such as those described for cdma2000.

APPENDICES

A. WISHART MATRICES

Wishart matrices are of interest in statistics and signal processing. We now review some facts concerning Wishart matrices, and derive some results concerning expectations of certain determinantal forms and further investigate the distribution and asymptotics of the largest eigenvalue. Books providing an introduction to matrix variate distributions are [50, 51, 52, 53]. These texts concentrate on real variables. Although the corresponding complex cases that we require are easily developed, they are noticeably absent from these references.

Definition A.1 (complex normal matrix). Let the real and imaginary parts of each element of $\mathbf{X} \in \mathbb{C}^{m \times n}$ be a selected i.i.d. Gaussian with zero mean and variance 1/2. Then $\mathbf{X}$ is *complex normal*, denoted $\mathbf{X} \sim \mathcal{N}(m, n)$.

Definition A.2 (complex Wishart). Let $\mathbf{W} = \mathbf{X}\mathbf{X}^*$, where $\mathbf{X} \sim \mathcal{N}(m, n)$. Then $\mathbf{W} \in \mathbb{C}^{m \times m}$ is *complex Wishart* [54, 55], denoted $\mathbf{W} \sim \mathcal{W}(m, n)$.

Real normal and Wishart matrices may also be defined, but we are more interested in the complex case. Henceforth we refer to a complex Wishart matrix as simply Wishart.

Theorem A.1 (Wishart density). *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$, $m \le n$. *Then*

$$p(\mathbf{W}) = \frac{1}{\Gamma_m(n)} \operatorname{etr}(-\mathbf{W}) \det \mathbf{W}^{n-m}, \quad \text{(A.1)}$$

where

$$\Gamma_m(a) = \pi^{m(m-1)/2} \prod_{i=1}^{m} \Gamma(a - i + 1) \quad \text{(A.2)}$$

is the complex multivariate gamma function.

In fact Theorem A.1 may be used as the definition of the Wishart matrix, from which the statement of Definition A.2 may be proved as a theorem. The ordering of definition and theorem however become less important in the light of [53, Theorem 3.3.3], where it is proved that a matrix $\mathbf{W}$ with the Wishart distribution may always be factored $\mathbf{W} = \mathbf{X}\mathbf{X}^*$, where $\mathbf{X} \sim \mathcal{N}(m, n)$. It is useful however to remember that there are other matrix forms besides Definition A.2 that lead to the Wishart distribution.

We can also consider Wishart matrices with arbitrary covariance structure. For $\mathbf{A} \in \mathbb{C}^{m \times m}$ nonsingular and $\mathbf{\Sigma} = \mathbf{A}\mathbf{A}^*$, the product $\mathbf{W} = \mathbf{A}\mathbf{X}\mathbf{X}^*\mathbf{A}^*$ is Wishart $\mathcal{W}(m, n, \mathbf{\Sigma})$ with density

$$p(\mathbf{W}) = \left(\det \mathbf{\Sigma}^n \Gamma_m(n) \right)^{-1} \operatorname{etr}(-\mathbf{\Sigma}^{-1}\mathbf{W}) \det \mathbf{W}^{n-m}. \quad \text{(A.3)}$$

The following theorem gives the distribution of a complex Wishart matrix. The proof of this theorem follows [53, Theorem 3.2.3], using however the complex hypergeometric functions, as outlined in Appendix B.

Theorem A.2 (Wishart distribution). *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$. *Then*

$$P(\mathbf{W} < \mathbf{X}) = \frac{\Gamma_m(m)}{\Gamma_m(n+m)} \det \mathbf{X}^n {}_1F_1(n; n+m; -\mathbf{X}). \quad \text{(A.4)}$$

Theorem A.3 (Ordered eigenvalue distribution). *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$ *with ordered eigenvalues* $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m$. *Then*

$$p(\lambda_1, \lambda_2, \ldots, \lambda_m) = \frac{\pi^{m(m-1)}}{\Gamma_m(n)\Gamma_m(m)}$$
$$\times \exp\left[\sum_{i=1}^{m} \lambda_i \right] \prod_{i=1}^{m} \lambda_i^{n-m} \prod_{i<j} (\lambda_i - \lambda_j)^2. \quad \text{(A.5)}$$

It is well known that $2^m \det \mathbf{W}$ is distributed as a product of independent random variables $\prod_{i=0}^{m-1} X_i$, where X_i is chi-squared with $2(n-i)$ degrees of freedom (see [51, pages 100–101] and [53, Theorem 3.3.22]). This property leads easily to the following lemma.

Lemma A.1 (generalized variance). *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$. *Then the moments of* $\det \mathbf{W}$ *are*

$$\mathrm{E}\left[\det \mathbf{W}^h\right] = \prod_{i=0}^{m-1} \frac{\Gamma(n - i + h)}{\Gamma(n - i)}. \quad \text{(A.6)}$$

In particular,

$$\mathrm{E}[\det \mathbf{W}] = \frac{n!}{(n-m)!},$$
$$\operatorname{var} \det \mathbf{W} = \left[\frac{n!}{(n-m)!} \right]^2 \frac{m}{n-m+1}. \quad \text{(A.7)}$$

Likewise, it is easy to determine moments of $\log \det \mathbf{W}$.

Lemma A.2. *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$. *Then the following identities hold.*

$$\mathrm{E}[\log \det \mathbf{W}] = \sum_{i=0}^{m-1} \psi(n - i),$$
$$\operatorname{var} \log \det \mathbf{W} = \sum_{i=0}^{m-1} \psi'(n - i), \quad \text{(A.8)}$$

where $\psi(x) = \Gamma'(x)/\Gamma(x)$ *is Euler's digamma function [39, equation (8.36)].*

The following lemma is from [56, Section 37].

Lemma A.3. *For* $\alpha \in \{1, 2, \ldots, n\}$, *let* $\mathbf{A}(\alpha)$ *be the principal submatrix [57, page 17] formed from* $\mathbf{A} \in \mathbb{C}^{n \times n}$ *by deleting rows and columns not indexed by the elements of* α. *Then for*

any $n \times n$ matrix $\mathbf{A}$

$$\det(\lambda\mathbf{I}_n + \mathbf{A}) = \sum_{i=0}^{n} \lambda^{n-i} \sum_{\substack{\alpha_i \subseteq \{1,2,\dots,n\} \\ |\alpha_i|=i}} \det\mathbf{A}(\alpha_i). \qquad (A.9)$$

Note that Lemma A.3 yields an explicit representation for the characteristic polynomial of a matrix. We will make extensive use of the following theorem, for which we believe the proof is new.

Theorem A.4. *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$ *and let* ω *be a given scalar. Then* $E[\det(\mathbf{I} + \omega\mathbf{W})] = m!\omega^m L_m^{n-m}(-1/\omega)$.

Proof. Using Lemma A.3 and the linearity of the expectation operator, we can write

$$E[\det(\mathbf{I} + \omega\mathbf{W})]$$

$$= \sum_{i=0}^{m} \omega^i \sum_{\substack{\alpha_i \subseteq \{1,2,\dots,m\} \\ |\alpha_i|=i}} E[\det\mathbf{W}(\alpha_i)] \qquad (A.10)$$

$$= \sum_{i=0}^{m} \binom{m}{i} \frac{n!}{(n-i)!} \omega^i \qquad (A.11)$$

$$= m!\omega^m \sum_{j=0}^{m} \binom{n}{m-j} \frac{(-1)^j}{j!} \left(-\frac{1}{\omega}\right)^j, \qquad (A.12)$$

where (A.11) follows from Lemma A.2 and the fact that $E[\det\mathbf{W}(\alpha_i)] = E[\det\mathbf{W}_i]$, where $\mathbf{W}_i \sim \mathcal{W}(n, i)$ (by marginalization of the Wishart distribution [53, Theorem 3.3.9]). The result follows from (A.12) which is the explicit representation of $L_m^{n-m}(-1/\omega)$ [38]. $\square$

We remark that Theorem A.4 yields (under suitable change of variables).

Corollary 3. *The expected characteristic polynomial of a Wishart matrix is*

$$E[\det(t\mathbf{I} - \mathbf{W})] = (-1)^m m! L_m^{n-m}(t). \qquad (A.13)$$

This result was apparently first derived by Edelman [58, Theorem 9.1], although our proof, based on the well-known Lemma A.3 is somewhat more direct. Unfortunately, we cannot generalize Theorem A.4 to give the variance, since the terms $\mathbf{W}(\alpha_i)$ are not independent (if they were, var would distribute over addition resulting in a simple expression).

The largest eigenvalue λ of a Wishart matrix is of interest in statistical analysis and in signal processing. In principal the distribution of the largest (indeed any) eigenvalue may be obtained from (A.5) by $(m-1)$-fold marginalization. The required integrals however quickly become unwieldily. The distribution of the largest eigenvalue of a Wishart matrix was first found by Sugiyama [59, 60]. This distribution may be easily derived from the c.d.f. of the matrix itself, as described in [51, page 421].

Theorem A.5. *Let* $\mathbf{W} \sim \mathcal{W}(m, n)$ *and let the random variable* $\lambda > 0$ *be the largest eigenvalue of* $\mathbf{W}$. *Then the distribution function* $\Lambda(x) = \Pr(\lambda < x)$ *is given by*

$$\Lambda(x) = \frac{\Gamma_m(m)}{\Gamma_m(n+m)} x^{nm} \,_1F_1(n; n+m; -x\mathbf{I}). \qquad (A.14)$$

Proof. A direct consequence of $\lambda < x$ if and only if $\mathbf{W} < x\mathbf{I}$ and Theorem A.2. $\square$

Computation of the hypergeometric function is difficult, and the series expansion into zonal polynomials (Appendix B) converges slowly. Sugiyama also gave an approximation to $\Lambda(x)$ in terms of a product of chi-squared distributions [61] (see also [51, Theorem 9.7.5]), but this approximation is only accurate for $n \gg m$. Alternatively, the largest root of (A.13) may be used as an approximation to $E[\lambda]$. Krishnaiah and Chang [62] have developed an alternate expression for $\Lambda(x)$ which avoids the use of zonal polynomials. Their expression involves linear combinations of certain double integrals.

As the matrix dimensions grow large proportionally, the extreme eigenvalues converge to nonrandom quantities.

Theorem A.6. *Let the elements of* $\mathbf{X} \in \mathbb{C}^{m \times n}$ *be selected i.i.d. with zero mean, unit variance and finite fourth moment. Let m and n increase without bound such that $n/m \to \beta \leq 1$ a.s. Let $\lambda_{\max}$ and $\lambda_{\min}$ be the largest and smallest eigenvalues of $\mathbf{XX}^*/m$. Then*

$$\begin{aligned} \lambda_{\max} &\longrightarrow b(\beta) \triangleq \left(1 + \sqrt{\beta}\right)^2 \quad \text{a.s.,} \\ \lambda_{\min} &\longrightarrow a(\beta) \triangleq \left(1 - \sqrt{\beta}\right)^2 \quad \text{a.s.} \end{aligned} \qquad (A.15)$$

The proof for this theorem can be found in [63] (smallest eigenvalue) and [64] for the largest eigenvalue. In fact, the entire empirical distribution of a randomly selected eigenvalue converges, see, for example, [42].

Theorem A.7. *Let* $\mathbf{X}$ *be selected according to Theorem A.6 and let m and n increase without bound such that $m/n \to \beta \leq 1$ a.s. Then the empirical distribution $F(x) = \Pr(\lambda < x)$ of $\mathbf{XX}^*/m$ converges to the nonrandom limit defined by*

$$dF(x)$$

$$= \begin{cases} (2\pi\beta x)^{-1}\sqrt{(x - a(\beta))(b(\beta) - x)}, & a(\beta) \leq x \leq b(\beta), \\ 0, & \text{otherwise.} \end{cases}$$

$$(A.16)$$

B. HYPERGEOMETRIC FUNCTIONS OF MATRIX ARGUMENT

Many calculations involving matrix variate distributions can be written in terms of hypergeometric functions of matrix argument. We now give for reference the definitions and results that we require. As was the case for the matrix variate distributions in Appendix A, we develop the complex variate case, which although straightforward is missing from the obvious texts.

Definition B.1 (partition). Let $k > 0$ then $\kappa = (k_1, k_2, \ldots, k_p)$ such that $k = \sum_{j=1}^{p} k_j$, $k_1 \geq k_2 \geq \cdots \geq k_p \geq 0$ is a *partition* of k into p parts.

Partitions may be ordered lexicographically as follows. If $\kappa = (k_1, k_2, \ldots)$ and $\lambda = (l_1, l_2, \ldots)$, then $\kappa > \lambda$ if $k_i > l_i$ for the first index i where the partitions differ. Now let $y_1, \ldots, y_p$ be p variables. Then we say that the monomial $y_1^{k_1} \cdots y_p^{k_p}$ is of order κ and that $y_1^{k_1} \cdots y_p^{k_p}$ is of higher order than $y_1^{l_1} \cdots y_p^{l_p}$ if $\kappa > \lambda$. The degree of a monomial in p variables is the sum of degrees of the individual variables. The degree of a polynomial is the maximum degree of the monomials making up the polynomial.

Define V_k to be the vector space of symmetric homogenous polynomials of degree k in p variables.[5] Further let V_κ be the subspace of V_k defined by polynomials of order κ. Then V_k is the direct sum of the irreducible invariant subspaces V_κ.

Zonal polynomials were introduced by James [65, 66, 67] and Constantine [68]. A concise definition of these polynomials for real symmetric matrix argument is given in [53]. We define the polynomials on Hermitian matrices.

Definition B.2 (zonal polynomial). Let V_k be defined on the eigenvalues of a $p \times p$ Hermitian matrix $\mathbf{X}$. Then the polynomial $(\operatorname{tr}\mathbf{X})^k \in V_k$ has a unique decomposition into polynomials $C_\kappa(\mathbf{X}) \in V_\kappa$ according to

$$(\operatorname{tr}\mathbf{X})^k = \sum_\kappa C_\kappa(\mathbf{X}). \tag{B.1}$$

The *zonal polynomial* $C_\kappa(\mathbf{X})$ is the component of $(\operatorname{tr}\mathbf{X})^k$ in V_κ.

A general formula for the coefficients of zonal polynomials has not been found. A recurrence for the coefficients may be found in [51]. An easily proved useful identity is $C_\kappa(\alpha\mathbf{X}) = \alpha^k C_\kappa(\mathbf{X})$.

The following constant occurs frequently in integrals concerning zonal polynomials. Note that this differs from the usual definition (e.g., [51, pages 247–248]), since we are considering zonal polynomials on the space of Hermitian matrices.

Definition B.3 (generalized hypergeometric coefficient). Let κ be a partition of k. Then

$$(a)_\kappa = \prod_{j=1}^{p} (a - j + 1)_{k_j}, \tag{B.2}$$

where $(a)_n = a(a+1)\cdots(a+n-1)$ is the Pochhammer symbol.

The zonal polynomials of the identity matrix are known [68, 69]. Suppose κ has exactly r nonzero parts. Then

$$C_\kappa(\mathbf{I}) = 2^{2k}k! \left(\frac{p}{2}\right)_\kappa \frac{\prod_{i<j}^{r}(2k_i - 2k_j - i + j)}{\prod_{i=1}^{r}(2k_i + r - i)!}. \tag{B.3}$$

Hypergeometric functions of matrix argument were introduced by Herz [70], defining the functions iteratively using Laplace transforms [70, equation (2.1)], starting from

$$_0F_0\mathbf{X} = \operatorname{etr}\mathbf{X}. \tag{B.4}$$

Note that we have modified Herz's definition to be suitable for Hermitian matrices (rather than symmetric).

Definition B.4 (hypergeometric function of matrix argument). Let $\mathbf{Z}$ be a Hermitian $p \times p$ matrix. The hypergeometric function of matrix argument is defined iteratively by

$$_{m+1}F_n(a_1, \ldots, a_m, c; b_1, \ldots, b_n; -\mathbf{Z}^{-1})$$
$$= \frac{\det\mathbf{Z}^c}{\Gamma_m(c)} \int_{\mathbf{X}>0} \operatorname{etr}(\mathbf{XZ})_m F_n(a_1, \ldots, a_m; b_1, \ldots, b_n; -\mathbf{X})$$
$$\times \det(\mathbf{X})^{c-m} d\mathbf{X}, \tag{B.5}$$

$$_m F_{n+1}(a_1, \ldots, a_m; b_1, \ldots, b_n, c; -\mathbf{X}^{-1})$$
$$= \frac{\det\mathbf{X}^{m-c}\Gamma_m(c)}{(2\pi i)^n}$$
$$\times \int_{\Re\mathbf{Z}>0} \operatorname{etr}(\mathbf{XZ})_m F_n(a_1, \ldots, a_m; b_1, \ldots, b_n; -\mathbf{Z}^{-1})$$
$$\times \det(\mathbf{Z})^{-c} d\mathbf{Z}. \tag{B.6}$$

Another particularly simple and useful form is

$$_1F_0(a; \mathbf{X}) = \det(\mathbf{I} - \mathbf{X})^{-a}. \tag{B.7}$$

The power series representation given below is due to Constantine [68].

Theorem B.1. *Let $\mathbf{X}$ be a Hermitian $p \times p$ matrix. The hypergeometric function of matrix argument can be written as*

$$_m F_n(a_1, \ldots, a_m; b_1, \ldots, b_n; \mathbf{X})$$
$$= \sum_{k=0}^{\infty} \sum_\kappa \frac{(a_1)_\kappa \cdots (a_m)_\kappa}{(b_1)_\kappa \cdots (b_n)_\kappa} \frac{C_\kappa(\mathbf{X})}{j!}. \tag{B.8}$$

REFERENCES

[1] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, 1996.

[2] I. Emre Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Telecomm.*, vol. 10, no. 6, pp. 585–595, 1999.

[3] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1998.

[4] J. H. Winters, J. Salz, and G. D. Gitlin, "The capacity of wireless communication systems can be substantially increased

---

[5]A polynomial that is unchanged by permutation of variable subscripts and such that every term has degree k.

by the use of antenna diversity," in *Conf. Inform. Sci. Syst.*, Princeton, NJ, USA, 1992.

[5] J. H. Winters, J. Salz, and R. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, no. 2-4, pp. 1740–1751, 1994.

[6] A. Narula, M. D. Trott, and G. W. Wornell, "Performance limits of coded diversity methods for transmitter antenna arrays," *IEEE Transactions on Information Theory*, vol. 45, no. 7, pp. 2418–2433, 1999.

[7] S. Dennett, "The cdma2000 ITU-R RTT candidate submission (0.18)," Technical report, TIA, July 1998.

[8] W. C. Jakes, *Microwave Mobile Communications*, John Wiley and Sons, New York, NY, USA, 1974.

[9] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice Hall, NJ, USA, 1993.

[10] S. Verdú, *Multiuser Detection*, Cambridge University Press, Cambridge, UK, 1998.

[11] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.

[12] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.

[13] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 3, pp. 451–460, 1999.

[14] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 744–765, 1998.

[15] A. F. Naguib, V. Tarokh, N. Seshadri, and A. R. Calderbank, "A space-time coding modem for high-data-rate wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1459–1478, 1998.

[16] D. Agrawal, V. Tarokh, A. Naguib, and N. Seshadri, "Space-time coded OFDM for high data-rate wireless communication over wideband channels," in *Proc. 48th IEEE Vehicular Technology Conference*, vol. 3, pp. 2232–2236, Ottawa, Canada, May 1998.

[17] V. Tarokh, A. Naguib, N. Seshadri, and A. R. Calderbank, "Combined array processing and space-time coding," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1121–1128, 1999.

[18] S. Baro, G. Bauch, and A. Hansmann, "Improved codes for space-time trellis-coded modulation," *IEEE Communications Letters*, vol. 4, no. 1, pp. 20–22, 2000.

[19] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, no. 2, pp. 41–159, 1996.

[20] P. W. Wonianski, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over rich scattering wireless channels," in *ISSSE*, pp. 295–300, 1998.

[21] H. El Gamal and A. R. Hammons, "The layered space-time architecture: a new perspective," submitted to *IEEE Transactions on Information Theory*, 2000.

[22] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, NY, USA, 2nd edition, 1984.

[23] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, 1999.

[24] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," Technical memorandum, Bell Laboratories, Lucent Technologies, April 2000, http://mars.bell-labs.com.

[25] L. Zheng and D. Tse, "Sphere packing in the Grassman manifold: a geometric approach to the noncoherent multi-antenna channel," in *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, June 2000, also submitted to *IEEE Transactions on Information Theory*. Preprint, http://degas.eecs.berkeley.edu/dtse/pub.html.

[26] B. Hochwald and W. Sweldens, "Differential unitary space-time modulation," *IEEE Trans. Communications*, vol. 48, no. 12, pp. 2041–2052, 2000.

[27] T. L. Marzetta, B. Hochwald, and B. Hassibi, "Space-time autocoding: arbitrarily reliable communication in a single fading interval," in *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, June 2000, also submitted to *IEEE Transactions on Information Theory*. Preprint, http://mars.bell-labs.com.

[28] D. Agrawal, T. J. Richardson, and R. Urbanke, "Multiple-antenna signal constellations for fading channels," in *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, June 2000, also submitted to *IEEE Transactions on Information Theory*. Preprint, http://mars.bell-labs.com.

[29] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 543–564, 2000.

[30] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, "Systematic design of unitary space-time constellations," *IEEE Transactions on Information Theory*, vol. 46, no. 6, pp. 1962–1973, 2000.

[31] B. L. Hughes, "Extensions to the theory of differential space-time modulation," in *Proc. IEEE Int. Symp. Inform. Theory*, Sorrento, Italy, June 2000, also submitted to *IEEE Transactions on Information Theory*.

[32] B. L. Hughes, "Differential space-time modulation," *IEEE Transactions on Information Theory*, vol. 46, no. 7, pp. 2567–2578, 2000.

[33] B. L. Hughes, "Optimal space-time constellations from groups," submitted to *IEEE Transactions on Information Theory*, 2000.

[34] B. L. Hughes, "Space-time coding without channel estimation," in *Conf. Inform. Sci. Sys.*, Johns Hopkins University, Baltimore, Md, USA, March 1999.

[35] B. L. Hughes, "Differential space-time modulation," in *IEEE Wireless Communications and Networking Conference*, vol. 1, pp. 145–149, New Orleans, La, USA, September 1999.

[36] B. L. Hughes, "Further results on differential space-time modulation," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, Mass, USA, March 2000.

[37] R. G. Gallager, *Information Theory and Reliable Communication*, Jonh Wiley and Sons, New York, NY, USA, 1968.

[38] G. Szegö, *Orthogonal Polynomials*, American Mathematical Society, Providence, Rhode Island, 1939.

[39] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, London, 5th edition, 1994.

[40] I. N. Bronshtein and K. A. Semendyayev, *Handbook of Mathematics*, Springer-Verlag, Berlin, Germany, 3rd edition, 1997.

[41] V. L. Girko, *Theory of Random Determinants*, vol. 45 of *Mathematics and Its Applications (Soviet Series)*, Kluwer Academic, Dordrecht, The Netherlands, 1990.

[42] D. Jonsson, "Some limit theorems for the eigenvalues of a sample covariance matrix," *J. Multivariate Anal.*, vol. 12, no. 1, pp. 1–38, 1982.

[43] 3GPP, "Multiplexing and channel coding (FDD) TS 25.212," Technical report, 3GPP, April 1999, http://www.3gpp.org.

[44] D. Gerlach and A. Paulraj, "Adaptive transmitting antenna arrays with feedback," *IEEE Signal Processing Letters*, vol. 1,

no. 10, pp. 150–152, 1994.

[45] W. K. M. Ahmed and P. J. McLane, "Error exponents for two-dimensional time correlated flat fading channels with space diversity and channel estimation," in *Proc. IEEE Int. Symp. Inform. Theory*, p. 240, Ulm, Germany, 29 June–4 July 1997.

[46] W. K. M. Ahmed, *Information theoretic reliability function for flat fading channels*, Ph.D. thesis, Queen's University, Kingston, Ontario, Canada, September 1997.

[47] V. L. Girko, "The central limit theorem for random determinants," *Theory of Probability and Its Applications*, vol. 24, no. 4, pp. 729–740, 1979.

[48] A. J. Grant and P. D. Alexander, "Random sequence multisets for synchronous code-division multiple-access channels," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2832–2836, 1998.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, NY, USA, 1991.

[50] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley and Sons, New York, NY, USA, 1958.

[51] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley and Sons, New York, NY, USA, 1982.

[52] M. Mehta, *Random Matrices*, Academic Press, San Diego, Calif, USA, 2nd edition, 1991.

[53] A. K. Gupta and D. K. Nagar, *Matrix Variate Distributions*, vol. 104 of *Monographs and Surveys in Pure and Applied Mathematics*, Chapman and Hall/CRC, Boca Raton, 2000.

[54] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, 1963.

[55] A. Edelman, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, M. I. T. Press, Cambridge, Mass, USA, May 1989.

[56] A. C. Aitken, *Determinants and Matrices*, Oliver and Boyd, Edinburgh, 9th edition, 1956.

[57] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Oxford University Press, Cambridge, New York, NY, USA, 1985.

[58] A. Edelman, "Eigenvalues and condition numbers of random matrices," *SIAM J. Matrix Anal. Appl.*, vol. 9, no. 4, pp. 543–560, 1988.

[59] T. Sugiyama, "On the distribution of the largest latent root and corresponding latent vector for principal component analysis," *Ann. Math. Statist.*, vol. 37, pp. 995–1001, 1966.

[60] T. Sugiyama, "On the distribution of the largest latent root of the covariance matrix," *Ann. Math. Statist.*, vol. 38, pp. 1148–1151, 1967.

[61] T. Sugiyama, "Approximation for the distribution of the largest latent root of a Wishart matrix," *Austral. J. Statist.*, vol. 14, no. 1, pp. 17–24, 1972.

[62] P. R. Krishnaiah and T. C. Chang, "On the exact distributions of the extreme roots of the Wishart and MANOVA matrices," *J. Multivariate Anal.*, vol. 1, pp. 108–117, 1971.

[63] Z. D. Bai and Y. Q. Yin, "Limit of the smallest eigenvalue of a large dimensional sample covariance matrix," *Ann. Probab.*, vol. 21, no. 3, pp. 1275–1294, 1993.

[64] Z. D. Bai, J. W. Silverstein, and Y. Q. Yin, "A note on the largest eigenvalue of a large dimensional sample covariance matrix," *J. Multivariate Anal.*, vol. 26, no. 2, pp. 166–168, 1988.

[65] A. T. James, "The distribution of the latent roots of the covariance matrix," *Ann. Math. Statist.*, vol. 31, pp. 151–158, 1960.

[66] A. T. James, "The distribution of noncentral means with known covariance matrix," *Ann. Math. Statist.*, vol. 32, pp. 874–882, 1961.

[67] A. T. James, "Zonal polynomials of the real positive definite symmetric matrices," *Ann. Math.*, vol. 74, no. 3, pp. 456–469, 1961.

[68] A. G. Constantine, "Some noncentral distribution problems in multivariate analysis," *Ann. Math. Statist.*, vol. 34, pp. 1270–1285, 1963.

[69] A. T. James, "Distribution of matrix variate and latent roots derived from normal samples," *Ann. Math. Statist.*, vol. 35, pp. 475–501, 1964.

[70] C. Herz, "Bessel functions of matrix argument," *Ann. Math.*, vol. 61, no. 3, pp. 474–523, 1955.

Alex Grant received the B.E. and Ph.D. degrees from the University of South Australia in 1993 and 1996, respectively. In 1997, he was a post-doctoral research fellow at the Laboratory for Signal and Information Processing, Swiss Federal Institute of Technology (ETH), Zürich. Since March 1998 he has been with the Institute for Telecommunication Research, where he is now a Senior Lecturer. Alex's research interests include information theory and its applications to multiple user and other communications problems. Alex is Leader of the Coding and Information Theory Research Group at ITR and serves as Chairman for the Australian Chapter of the IEEE Information Theory Society.