Applied Mathematics & Information Sciences **5**(3) (2011), 460-483
– An International Journal

Channel capacity for MIMO systems with multiple frequencies and delay spread

Christopher S. Withers[1] and Saralees Nadarajah[2]
[1] Applied Mathematics Group, Industrial Research Limited, Lower Hutt, NEW ZEALAND
[2]School of Mathematics, University of Manchester, Manchester M13 9PL, UK
Email Address: mbbsssn2@manchester.ac.uk

Received 3 November 2010; Revised 4 April 2011; Accepted 21 June 2011

Consider N-transmit M-receive antenna systems with multiple frequencies and delay spread. Expansions are given for the distribution and quantiles of the channel capacity efficiency C in powers of $N^{-1/2}$ for fixed M. The first term gives normality. This gives a good approximation for M/N small. For $M < N$ the second or third term is generally sufficient for accuracy. An important duality principle is given: expansions for the distribution and quantiles of C in powers of $M^{-1/2}$ for fixed N follow. The first term gives a good approximation for N/M small. Both discrete and continuous time models are considered.

Keywords: Antenna systems, Channel capacity, Expansions, Quantiles.

1 Introduction

The aim of this paper is to provide asymptotic solutions of the capacity cumulant moments in the limit of many transmitter antennas M and fixed receiver antennas N and vice versa. The approach taken is different from other approaches in the literature, where the ratio M/N is kept fixed and both are assumed to be large (Foschini and Gans, 1998; Telatar, 1999; Hochwald *et al.*, 2004). We also mention Boche and Jorswieck (2002), where the full distribution is calculated for correlated antennas for the multi-transmitter single-receiver case; Martin and Ottersten (2002), where second order approximations are calculated for eigenvalue moments for MIMO channels; Moustakas *et al.* (2003), where the full distribution is calculated for any M, N for independent and identically distributed channels; Wang and Giannakis (2004), where the first three moments are calculated for correlated channels for large M, N. For more recent work, see Tulino and Verdu (2004) and Hachem *et al.* (2008).

Withers and Vaughan (2001) considered N-transmit M-receive antenna systems, where the power of the noise at the mth receiver is Q_m and the power of the nth transmitter is $P_n/(N\lambda)$ for $\lambda = 1$ if the mean total power is bounded and $N\lambda = 1$ if the mean total power is increasing. Adapting Foschini and Gans (1998), let $\mathbf{s}(t)$ denote the $N \times 1$ signal transmitted at time t, $\mathbf{e}(t)$ the $M \times 1$ noise at the receiver at time t, and $\mathbf{r}(t)$ the $M \times 1$ received signal. The vector equation describing the channel operating on the signal is

$$\mathbf{r}(t) = \mathbf{g}(t) \otimes \mathbf{s}(t) + \mathbf{e}(t), \tag{1.1}$$

where $\otimes$ denotes convolution and $\mathbf{g}(t)$ is the $M \times N$ matrix channel impulse response. Assuming that $\mathbf{s}(t) = \mathbf{0}$ for $t \leq 0$, its Fourier transform is

$$\mathbf{R}_w = \mathbf{G}_{0w}\mathbf{S}_w + \mathbf{E}_w, \tag{1.2}$$

say. The convolution and Fourier transforms are discrete for discrete time models and continuous for continuous time models. Assuming that a randomly selected channel is not changing during a burst, Foschini and Gans (1998) gave the capacity efficiency (capacity/bandwidth) as

$$C = \log_2 \left(\det \mathbf{A}_s \det \mathbf{A}_r / \det \mathbf{A}_u\right),$$

where $\mathbf{A}_s = E\mathbf{S}_w\mathbf{S}_w^+$, $\mathbf{A}_r = E\mathbf{R}_w\mathbf{R}_w^+$, $\mathbf{A}_u = E\mathbf{U}_w\mathbf{U}_w^+$, and $\mathbf{U}_w = \binom{\mathbf{S}_w}{\mathbf{R}_w}$, and $\mathbf{x}^T$, $\overline{\mathbf{x}}$ and $\mathbf{x}^+$ denote the transpose, conjugate and transpose conjugate of $\mathbf{x}$. Since $\mathbf{A}_s = \mathbf{D}_P/(N\lambda)$ and $\mathbf{A}_r = \mathbf{D}_Q + \mathbf{G}_{0w}\mathbf{D}_P\mathbf{G}_{0w}^+/(N\lambda)$, where

$$\mathbf{D}_P = diag\left(P_1, \ldots, P_N\right),$$
$$\mathbf{D}_Q = diag\left(Q_1, \ldots, Q_M\right),$$

this gives the capacity efficiency in bps/Hz as

$$C = \log_2 \det\left(\mathbf{I}_M + \mathbf{X}_w\mathbf{X}_w^+/(N\lambda)\right), \tag{1.3}$$

where $\mathbf{I}_M = \text{diag}\ (1, \ldots, 1)$ and

$$\mathbf{X}_w = \mathbf{D}_Q^{-1/2}\mathbf{G}_{0w}\mathbf{D}_P^{1/2}. \tag{1.4}$$

As in Foschini and Gans (1998) we replace $\{P_n, Q_m\}$ by their means $\{p_n, q_m\}$. For their case, $\mathbf{A}_s = PN^{-1}\mathbf{I}_N$ and $\mathbf{A}_r = Q\mathbf{I}_M + PN^{-1}\mathbf{G}_{0w}\mathbf{G}_{0w}^+$. Their formula goes back to equation (21) of Winters (1987). Also $\mathbf{G}_{0w} = (\mathbf{H}_{10w}, \ldots, \mathbf{H}_{N0w})$, where for independent transmitters, $\{\mathbf{H}_{n0w}\}$ are independent $\mathcal{CN}_M(\boldsymbol{\mu}_0, \mathbf{V}_0)$ and $\boldsymbol{\mu}_0$, $\mathbf{V}_0$ do not depend on $\{p_n\}$ or $\{q_m\}$. For Raleigh fading $\boldsymbol{\mu}_0 = \mathbf{0}$. For independent receivers $\mathbf{V}_0$ is diagonal. Typically $\mathbf{V}_0 \propto \mathbf{I}_M$ but for line-of-sight the elements of $\mathbf{V}_0$ all have the same value.

When multiple frequencies are used, (1.3) becomes

$$C = \int_0^\infty \log_2 \det \left(\mathbf{I}_M + \mathbf{X}_w \mathbf{X}_w^+ / (N\lambda) \right) d\nu(w), \tag{1.5}$$

where W is a random frequency independent of the random process $\mathbf{X}_w$ with distribution determined by the spectrum of frequencies used, say

$$P(W \leq w) = \nu(w). \tag{1.6}$$

The simplest example is $W \sim U(w_0, w_0 + B)$ uniform with bandwidth B and base frequency w_0, that is $d\nu(w) = B^{-1} I(w_0 < w < w_0 + B) dw$, where $I(A)$ is 1 or 0 for A true or false. Another example is W uniform over a number of non-overlapping intervals $I_1, \ldots, I_J$ of bandwidths $B_1, \ldots, B_J$ and total bandwidth $B = B_1 + \cdots + B_J$, that is

$$d\nu(w) = B^{-1} I \left(w \in I_1 \cup \cdots \cup I_J \right) dw. \tag{1.7}$$

The columns of $\mathbf{G}_{0w}$ of (1.4), $\{\mathbf{H}_{n0w}\}$, are again independent copies of $\mathbf{H}_{0w}$, the Fourier transform of a column of $\mathbf{g}(t)$, say $\mathbf{g}_{0t}$.

We consider two models for *delay spread*. The first assumes that each column of $\mathbf{g}_{0t}$ takes the form

$$\mathbf{g}_{0t} = \int \mathbf{Z}_{0\ell} \delta(t - \ell) dP(L \leq \ell) \text{ in } \mathcal{C}^M, \tag{1.8}$$

where $\{\mathbf{Z}_{0\ell}\}$ are independent $\mathcal{CN}_M(\boldsymbol{\mu}_0, \mathbf{V}_0)$. We assume that the transmitters are close enough together and the receivers close enough together so that the distribution $P(L \leq \ell)$ of the random delay L_{nm} from transmitter n to receiver m does not depend on n, m. To cover both continuous and discrete delay let f_ℓ be the density of the delay distribution with respect to a dominating measure ϵ_ℓ: $dP(L \leq \ell) = f_\ell d\epsilon_\ell$. Assuming *continuous time*,

$$\mathbf{H}_{0w} = \int \mathbf{Z}_{0\ell} \exp\left(-jw\ell\right) f_\ell d\epsilon_\ell \tag{1.9}$$

is finite with probability with probability 1. For *discrete delay*, ϵ_ℓ is counting measure so that

$$f_\ell = P(L = \ell), \ \mathbf{g}_{0t} = \left(f_t \mathbf{Z}_{0t}\right) \otimes_d \delta(t), \ \mathbf{H}_{0w} = \sum_\ell f_\ell \mathbf{Z}_{0\ell} \exp\left(-jw\ell\right), \tag{1.10}$$

where $\otimes_d$ denotes discrete convolution. The simplest example is *discrete rectangular delay*

$$L = \ell \text{ with probability } I^{-1} \text{ for } \ell = 0, 1, \ldots, I - 1 \tag{1.11}$$

for some integer $I \geq 1$ (labeled $f = f_1$ in Section 2), so that

$$\mathbf{g}_{0t} = I^{-1} \sum_{\ell=0}^{I-1} \mathbf{Z}_{0\ell} \delta(t - \ell),$$

$$\mathbf{H}_{0w} = I^{-1} \sum_{\ell=0}^{I-1} \mathbf{Z}_{0\ell} \exp\left(-j\ell w\right).$$

One could absorb the factor I^{-1} into $\mathbf{Z}_{0\ell}$ by replacing $\boldsymbol{\mu}_0$, $\mathbf{V}_0$ by $I^{-1}\boldsymbol{\mu}_0$, $I^{-2}\mathbf{V}_0$.

This includes the delay model of Pedersen *et al.* (2001)

$$\mathbf{g}(t) = \sum_{i=1}^{I} \mathbf{a}_i \delta\left(t - d_i\right)$$

so that

$$\mathbf{r}(t) = \sum_{i=1}^{I} \mathbf{a}_i \mathbf{s}\left(t - d_i\right) + \mathbf{e}(t). \tag{1.12}$$

Here, I is the number of paths, d_i is the delay of path i, and $\mathbf{a}_i$ is the gain of path i. Raleigh and Cioffi (1998) model $\mathbf{a}_i$ in terms of angles of departure and arrival. Note that (1.12) can be written as (1.1) with $\otimes = \otimes_c$ denoting continuous convolution and

$$\mathbf{g}(t) = f_t \widetilde{\mathbf{Z}}_{0t} = \mathbf{a}_i \tag{1.13}$$

for $t = d_i$, where $f_t = 0$ for $t \neq d_i$ and the columns of $\widetilde{\mathbf{Z}}_{0t}$ are independent copies of $\mathbf{Z}_{0t}$ as above, where the scalar f_t may now be complex and $\sum_t |f_t| < \infty$. That is, we assume that the columns of $\mathbf{a}_i$ are independently distributed as

$$g_{0t} = f_t \mathbf{Z}_{0t} \sim \mathcal{CN}_M\left(f_t \boldsymbol{\mu}_0, |f_t|^2 \mathbf{V}_0\right), \tag{1.14}$$

where $t = d_i$.

As well as (1.8), we also consider the *discrete time* delay model

$$\mathbf{g}_{0t} = f_t \mathbf{Z}_{0t} \tag{1.15}$$

for $\mathbf{Z}_{0t}$ as above, where the scalar f_t may be complex. This includes (1.12), the static delay model of Telatar and Tse (2000) - (1) and the equation before (18): (1.12) can be written as (1.1) with $\otimes = \otimes_d$ and $\mathbf{g}(t)$ of (1.13). So, (1.14) holds and (1.9) again holds - with ϵ_ℓ counting measure.

We assume that the delay distribution has finite Fisher information:

$$h_0 = \int |f_\ell|^2 \, d\epsilon_\ell < \infty.$$

We call h_0^{-1} *the delay factor* as it increases with mean delay. For example, the delay factor equals the mean delay if the delay is a scaled exponential random variable. We shall see that *the random delay L reduces the SNR by the delay factor*. It can be shown that $\mathbf{H}_{0w}$ is a Gaussian process in $\mathcal{C}^M$ with mean and covariance determined by $\boldsymbol{\mu}_0$, $\mathbf{V}_0$ and the (discrete or continuous) Fourier transforms of the delay density and its square:

$$E\mathbf{H}_{0w} = \boldsymbol{\mu}_0 F(w),\ cov\left(\mathbf{H}_{0w_1}, \mathbf{H}_{0w_2}\right) = \mathbf{V}_0 h\left(w_1 - w_2\right) \tag{1.16}$$

for

$$F(w) = \int \exp\left(-jw\ell\right) f_\ell d\epsilon_\ell = E \exp\left(-jwL\right), \tag{1.17}$$

$$h(w) = \int \exp(-jw\ell) \left|f_\ell\right|^2 d\epsilon_\ell = h_0 E \exp\left(-jw\widetilde{L}\right), \tag{1.18}$$

where $\widetilde{L}$ is a random variable with distribution

$$P\left(\widetilde{L} \leq t\right) = h_0^{-1} \int^t \left|f_\ell\right|^2 d\epsilon_\ell. \tag{1.19}$$

The use of E in (1.17) is for the case f_ℓ real and $\int f_\ell d\epsilon_\ell = 1$, as in (1.8). For the case (1.12), (1.17) and (1.18) take the form

$$F(w) = \sum_{i=1}^{I} \exp\left(-jwd_i\right) \widetilde{f}_i,$$
$$h(w) = \sum_{i=1}^{I} \exp\left(-jwd_i\right) \left|\widetilde{f}_i\right|^2,$$

where $\widetilde{f}_i = f_{t_{d_i}}$, and the analysis is conditional on the delay times $\{d_i\}$. We call $\widetilde{L}$ *the associated delay*.

For no delay $F(w) = h(w) = h_0 = 1$. For discrete delay (1.10), (1.16) assumes that

$$cov\left(\mathbf{Z}_{0\ell_1}, \mathbf{Z}_{0\ell_2}\right) = \mathbf{V}_0 \delta_{\ell_1 \ell_2}, \tag{1.20}$$

where $\delta_{\ell_1\ell_2} = 1$ or 0 for $\ell_1 = \ell_2$ or $\ell_1 \neq \ell_2$. For continuous delay, $d\epsilon_\ell = d\ell$ Lebesgue measure, (1.16) assumes that

$$cov\left(\mathbf{Z}_{0\ell_1}, \mathbf{Z}_{0\ell_2}\right) = \mathbf{V}_0 \delta\left(\ell_1 - \ell_2\right). \tag{1.21}$$

It is convenient to absorb the factor $\mathbf{D}_Q^{-1/2}$: set

$$\mathbf{H}_{nw} = \mathbf{D}_Q^{-1/2} \mathbf{H}_{n0w},$$
$$\mathbf{G}_w = \mathbf{D}_Q^{-1/2} \mathbf{G}_{0w} = \left(\mathbf{H}_{1w}, \ldots, \mathbf{H}_{Nw}\right).$$

So, $\{\mathbf{H}_{nw}\}$ are independent copies of $\mathbf{H}_w$, the Fourier transform of $\mathbf{g}_t = \mathbf{D}_Q^{-1/2} \mathbf{g}_{0t}$. For the continuous time delay model (1.8),

$$\mathbf{g}_t = \int \mathbf{Z}_\ell \delta(t - \ell) f_\ell d\epsilon_\ell$$

while for the discrete time delay model (1.15), $\mathbf{g}_t = f_t \mathbf{Z}_t$: in both cases $\{\mathbf{Z}_\ell = \mathbf{D}_Q^{-1/2} \mathbf{Z}_{0\ell}\}$ are independent $\mathcal{CN}_M(\boldsymbol{\mu}, \mathbf{V})$, where $\boldsymbol{\mu} = \mathbf{D}_Q^{-1/2} \boldsymbol{\mu}_0$ and $\mathbf{V} = \mathbf{D}_Q^{-1/2} \mathbf{V}_0 \mathbf{D}_Q^{-1/2}$.

It can be shown that for fixed M and $r \geq 1$, the rth cumulant of

$$\widehat{\theta} = C_0, \tag{1.22}$$

where $C_0 = M \ln \lambda + C \ln 2$, can be expanded as

$$\kappa_r\left(\widehat{\theta}\right) = \sum_{i=r-1}^{\infty} a_{ri} N^{-i}. \tag{1.23}$$

Note that (1.23) implies that as $N \to \infty$

$$Y_N = (N/a_{21})^{1/2}\left(\widehat{\theta} - a_{10}\right) \to \mathcal{N}(0,1) \tag{1.24}$$

for $\mathcal{N}(0,1)$ a unit real normal random variable, and that the distribution, density and quantiles of Y_N (and so of C_0 and C) can be expanded in powers of $N^{-1/2}$ about those of the unit normal. Let $\Phi(x)$ and $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ be the distribution and density of $\mathcal{N}(0,1)$. Set

$$P_N(x) = P\left(Y_N \leq x\right).$$

Then for $\widehat{\theta}$ non-lattice, the distribution, density and quantiles of Y_N (and so those of $\widehat{\theta}$) are given by the asymptotic expansions

$$\begin{aligned}
P_N(x) &\approx \Phi(x) - \phi(x)\sum_{r=1}^{\infty} N^{-r/2} h_r(x),\\
p_N(x) &\approx \phi(x)\left\{1 + \sum_{r=1}^{\infty} N^{-r/2} h_r'(x)\right\},\\
\Phi^{-1}\left(P_N(x)\right) &\approx x - \sum_{r=1}^{\infty} N^{-r/2} f_r(x),\\
P_N^{-1}\left(\Phi(x)\right) &\approx x + \sum_{r=1}^{\infty} N^{-r/2} g_r(x),
\end{aligned}$$

where $\{h_r(x), h_r'(x), f_r(x), g_r(x)\}$ are polynomials in both x and $A_{ri} = a_{ri}/a_{21}^{r/2}$, the standardized cumulant coefficients. These expansions are called the Edgeworth-Cornish-Fisher expansions. See Withers (1984) for these polynomials. For $R = 1, 2, \ldots$ *the Rth order approximations* truncate these to R terms giving remainder $O(N^{-R/2})$. For $R = 1$, the Central Limit Theorem (1.24), one needs a_{10}, a_{21}. For $R = 2$ one needs a_{11}, a_{32} and for $R = 3$ one needs a_{22}, a_{43}.

Now apply this to $\widehat{\theta} = C_0$ of (1.22). So, $a_{10} = M \ln \lambda + c \ln 2$ and $Y_N =$

$(\ln 2)(N/a_{21})^{1/2}(C-c)$, where c is the mean of C for large N. So, one obtains

$$\begin{aligned} c &= c_{Raleigh} + c_{Rice}, \\ c_{Rice} &= \int_0^\infty \log_2\left(1+2R_{\boldsymbol{\mu}}\left|F(w)\right|^2\right) d\nu(w) = E\log_2\left(1+2R_{\boldsymbol{\mu}}\left|F(W)\right|^2\right), \\ p_T &= \left(\sum_{n=1}^N p_n\right)/(N\lambda), \\ R_{\boldsymbol{\mu}} &= \boldsymbol{\mu}^+\left(p_T^{-1}\mathbf{I}_M + h_0\mathbf{V}\right)^{-1}\boldsymbol{\mu}/2, \end{aligned} \tag{1.25}$$

where

$$c_{Raleigh} = \log_2 \det\left(\mathbf{I}_M + h_0 p_T \mathbf{V}\right) \tag{1.26}$$

and p_T is *the average total power transmitted*. The delay reduces the power by the delay factor. For no delay $F(w) = h_0 = 1$ so that $c_{Rice} = \log_2(1+2R_{\boldsymbol{\mu}})$ does not depend on the spectrum used. Note $R_{\boldsymbol{\mu}}$ may be identified as a scaled Rice factor: it depends on the delay only through h_0. If the delay L is discrete so that (1.10) holds then $\min_\ell f_\ell \leq h_0 \leq \max_\ell f_\ell \leq 1$, so that the delay factor reduces the effective SNR. For rectangular delay (1.11), $h_0 = I^{-1}$.

So, in situations, where one can affect the distribution of the delay L, one should seek to maximize h_0.

However, if L is continuous, (for example, a scaled exponential), then h_0 can take on values greater than one if the scale factor is small enough! In fact, as the scale factor decreases to zero (corresponding to $L = 0$, that is no delay,) then $h_0 \to \infty$. This is counter-intuitive but it reflects the different assumptions (1.20) and (1.21).

Although (1.25) was calculated assuming transmitters (but not receivers) to be independent, by the Law of Large Numbers as $N \to \infty$, C converges to c even if transmitters are correlated.

The Raleigh component of the asymptotic capacity $c_{Raleigh}$, does not depend on $\boldsymbol{\mu}$ (so its name); nor on the frequency distribution $\nu(w)$! It depends on the delay only through h_0. The Rice component c_{Rice} depends on h_0, $F(w)$ of (1.17) and $\nu(w)$; it is zero if $\boldsymbol{\mu} = \mathbf{0}$.

Each cumulant coefficient $a_{ri} = a_{ri}(\boldsymbol{\mu})$ say, can also be written as the sum of a Raleigh term $a_{ri}(\mathbf{0})$ and a Ricean term. Each Raleigh term can be written in terms of

$$\gamma_r' = p_{rN}\Gamma_r \tag{1.27}$$

for

$$p_{rN} = N^{-1}\sum_{n=1}^N p_n^r, \tag{1.28}$$

$$\Gamma_r = \begin{cases} h_0^r \text{ trace } \left(\lambda^{-1}\mathbf{V}\left(\mathbf{I}_M + h_0 p_T \mathbf{V}\right)^{-1}\right)^r, & \\ p_{1N}^{-r} \text{ trace } \left(\mathbf{I}_M + h_0^{-1}p_T^{-1}\mathbf{V}^{-1}\right)^{-r}, & \text{if } \det(\mathbf{V}) \neq 0, \end{cases} \tag{1.29}$$

and real functions $[\cdot]'$ of the form $E\exp(-jK)=E\cos K$, where K is a real symmetric random variable determined by the distributions of L, W through those of $\widetilde{L}_1-\widetilde{L}_2$ and W_1-W_2, where $\widetilde{L}_1$, $\widetilde{L}_2$ independent copies of $\widetilde{L}$ of (1.19) and W_1, W_2 independent copies of W of (1.6). These *shrinkage functions* $[\cdot]'$ lie in $[-1,1]$ and act as shrinkage factors as they are less than 1 unless either W is constant (a single frequency) or if L is constant (a fixed delay), in which case $K\equiv 0$ so $[\cdot]'\equiv 1$. For example, by (1.22)-(1.23) the asymptotic variance of C is $a_{21}/N(\ln 2)^2$ for $a_{21}=a_{21}(\boldsymbol{\mu})$, where

$$a_{21}(\mathbf{0})=\gamma_2'[12]' \tag{1.30}$$

for

$$\begin{aligned}
[12]' &= h_0^{-2}E\left|h\left(W_1-W_2\right)\right|^2 && (1.31)\\
&= E\exp(-jK)=E\cos K && (1.32)\\
&= E\left|\widetilde{\nu}_{\widetilde{L}_1-\widetilde{L}_2}\right|^2, && (1.33)\\
K &= \left(W_1-W_2\right)\left(\widetilde{L}_1-\widetilde{L}_2\right), && (1.34)\\
\widetilde{\nu}_\ell &= E\exp\left(-j\ell W\right). && (1.35)
\end{aligned}$$

Note $\widetilde{\nu}_\ell$ is the Fourier transform of the frequency distribution.

It can be shown more generally for the Raleigh case $\boldsymbol{\mu}=\mathbf{0}$ that if only one frequency is used then *delay has no effect on capacity except for the delay factor in the effective SNR*, and if the delay is constant then *the choice of spectrum has no effect on capacity*.

If neither L nor W are constant and either are increased stochastically (typically by increasing a scale parameter, say $\epsilon\to\infty$) in such a way that $\widetilde{L}_1-\widetilde{L}_2$ or W_1-W_2 become stochastically unbounded, then each shrinkage function tends to zero so that each $a_{ri}(\mathbf{0})\to 0$ except for $(ri)=(10)$, so that for the Raleigh case $\boldsymbol{\mu}=\mathbf{0}$, C converges in probability to $c_{Raleigh}$ of (1.26).

Note that (1.31)-(1.33) remind one of Parseval's identity written in the form

$$(2\pi)^{-1}\int\left|E\exp\left(-jwX\right)\right|^2dw=(2\pi)^{-1}\int E\exp\left\{-jw\left(X_1-X_2\right)\right\}dw=\int f(x)^2dx,$$

where $X_1,\ X_2$ are independent random variables with density $f(x)$ with respect to Lebesgue measure.

For the asymptotic variance for the Ricean case (that is a_{21} when $\boldsymbol{\mu}\neq\mathbf{0}$).

Example 1.1. Suppose that the transmitters have the same average power $p_n=p$, and the receiver noises have the same average power $q_m=q$. Take $\mathbf{V}_0=v_0\mathbf{I}_M$. Set

$$\rho_3=h_0v_0p/q,\ \rho_1=\rho_3/\lambda,\ \rho_2=\rho_1/h_0. \tag{1.36}$$

These are scaled forms of the total signal to individual receiver noise ratio $\rho_2/v_0 = p/(\lambda q)$. Then

$$c_{Raleigh} = M \log_2 (1+\rho_1), \qquad (1.37)$$
$$\gamma'_r = M \left(1+\rho_1^{-1}\right)^{-r}.$$

So, ρ_1 is the effective SNR. Figure 1.1 plots $c_{Raleigh}/M$ in bits/sec/hz against the scaled SNR ρ_2 for $h_0 = 2^{-i}$, $0 \le i \le 4$.

How far can the assumption that M be bounded as $N \to \infty$ be relaxed? If $N \ge M \to \infty$ then $A_{11} \sim M^{3/2}$, $A_{32} \sim M^{-1/2}$, $A_{22} \sim M$, $A_{43} \sim M^{-1}$ so that $g_1(x)N^{-1/2} \sim (M^3/N)^{1/2} \to 0$ (implying the Central Limit Theorem) if $M = o(N^{1/3})$. (Here, $a_{M,N} \sim b_{M,N}$ means that $a_{M,N}/b_{M,N} \to 1$ in the limit.) This rules out the case when M, N have the same order of magnitude. The second order approximation to the percentiles requires the weaker requirement that $g_2(x)/N \sim M/N \to 0$. The third order approximation appears to require the intermediate condition $M = o(N^{3/5})$.

For the case of W or L constant Figures 1.2 plot the three approximations to the 1 percentile of capacity, while Figures 1.3 plot the mean and third order approximations to 8 percentiles.

Unlike the first order approximations, the third order approximations are not symmetric about the asymptotic mean c. As M increases to N, c crosses over the upper percentiles.

This example is continued in Example 2.1 with plots of c_{Rice}.

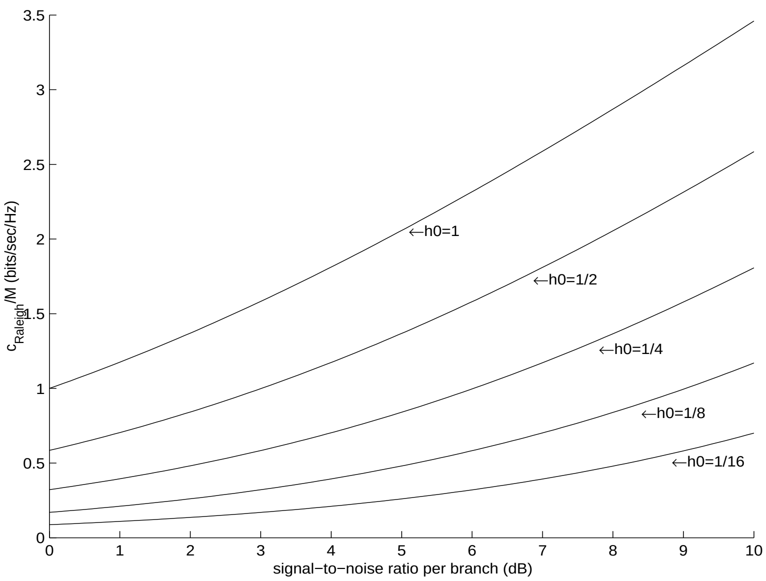

Figure 1.1 $c_{Raleigh}/M$ versus scaled SNR $10 \log_{10} \rho_2$.

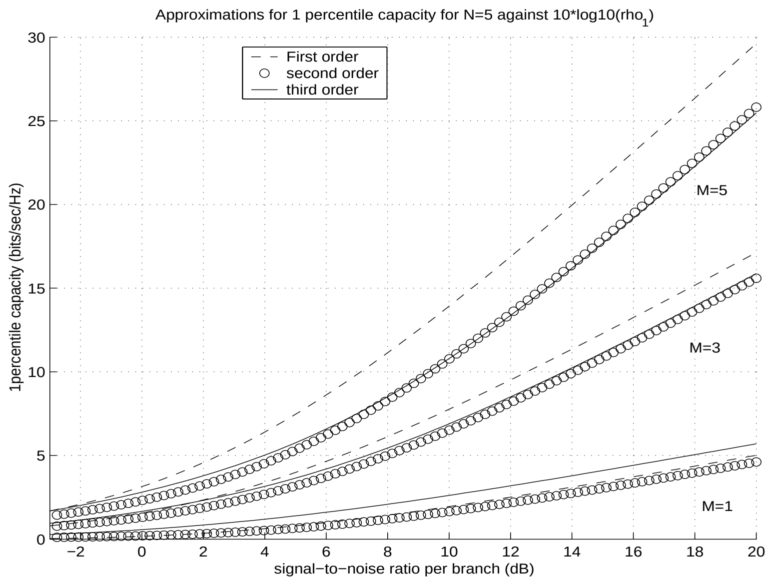

Figure 1.2a First three approximations to 1 percentile against ρ_1 for $N = 5$.

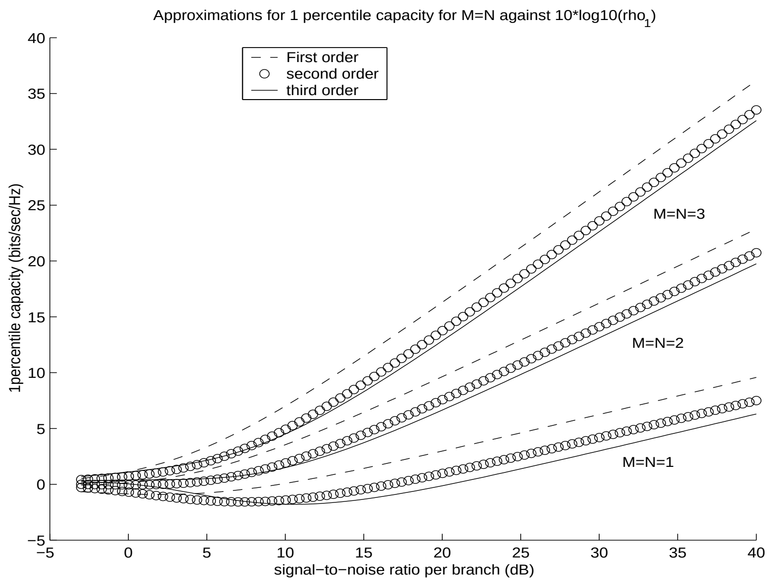

Figure 1.2b First three approximations to 1 percentile against ρ_1 for $M = N \leq 3$ become negative for low db for $M \leq 2$ and fail to show increasing capacity below 5db for $M = 2$ and below 10b for $M = 1$.

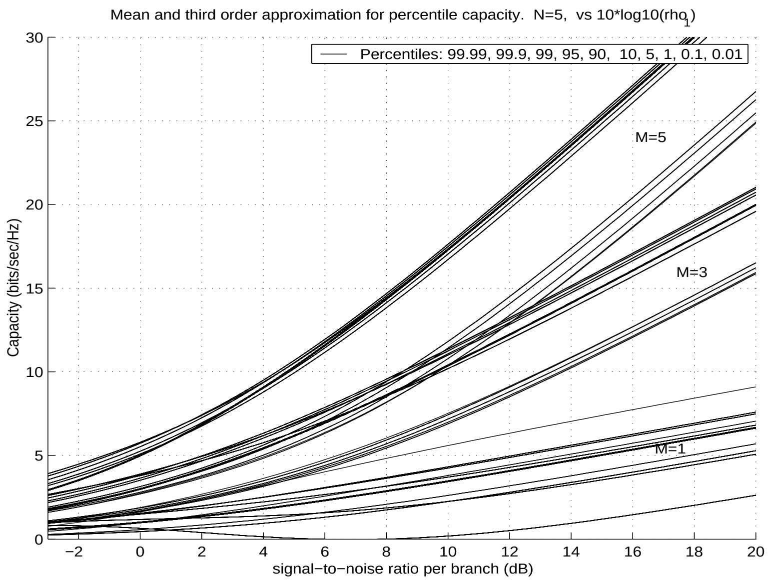

Figure 1.3a Mean and third approximations to percentiles against ρ_1 for $N = 5$.

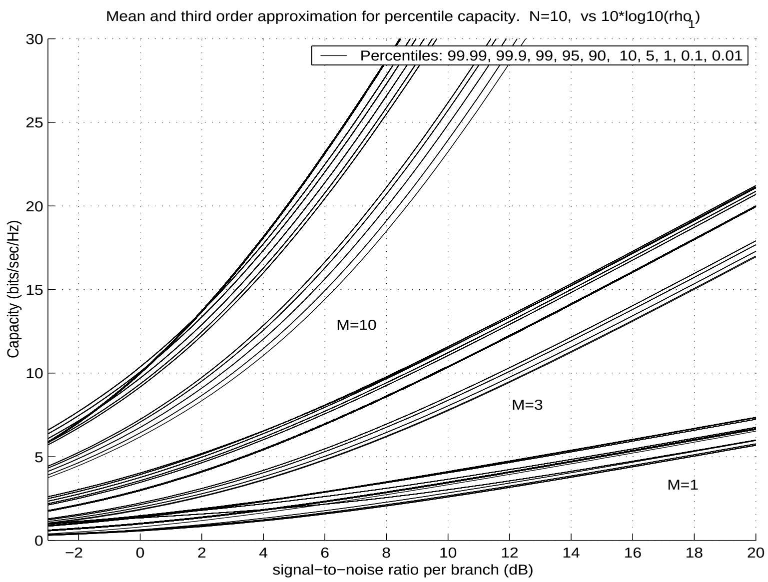

Figure 1.3b Mean and third approximations to percentiles against ρ_1 for $N = 10$.

The 1 percentile of the distribution of the capacity efficiency C are plotted for some examples in Section 2, showing as in Withers and Vaughan (2001) that even for N as small as three (for the case of no delay) the third order approximations hardly differ from the second order approximations unless $M \gg N$.

Section 3 gives an important duality result between $\{p_n\}$ and $\{q_m^{-1}\}$. This allows all these expansions in powers of N^{-1} or $N^{-1/2}$ for fixed M to be re-interpreted as powers of M^{-1} or $M^{-1/2}$ for fixed N:

For example, for Raleigh fading with $p_n \equiv p$, $q_m \equiv q$ then for fixed N, as $M \to \infty$, $C \to c^*_{Raleigh} = N \log_2(1+\rho_1^*)$ for $\rho_1^* = \rho_1 M/N$. If M, N are both large and of the same magnitude then asymptotic normality no longer holds: see, for example, Smith and Shafi (2001) and Johnstone (2001).

Throughout, we shall write $C_n \approx \sum_{r=0}^{\infty} c_{rn}$ to mean that that for $i \geq 1$ under suitable regularity conditions $C_n - \sum_{r=0}^{i-1} c_{rn}$ converges to zero as $n \to \infty$. We shall also write $\dot{\omega}(\cdot)$ to denote the first derivative of $\omega(\cdot)$.

2 Examples

Here, we assume that the transmitters have the same average power and receivers have the same average power, say $p_n = p$, $q_m = q$ so that $p_{rN} = p^r$. Define ρ_i as in (1.36).

We consider (1.8) for both discrete and continuous delay L. The discrete delay density is *the rectangular* (1.11), $f = f_1$ say. For $f = f_1$, the mean delay is $(I-1)/2$, $h_0 = I^{-1}$ and $F(w)$, $h(w)$ of (1.17), (1.18) are

$$I\ F(w) = h(w) = \left\{1 - \exp\left(-jwI\right)\right\} / \left\{1 - \exp\left(-jw\right)\right\}.$$

So,

$$I^2 \left|F(w)\right|^2 = \left|h(w)\right|^2 = \left(1 - \cos Iw\right) / \left(1 - \cos w\right).$$

Also for $f = f_1$, $\widetilde{L}$ has the same distribution as L.

The continuous delay density, f_2 say, is that of $L = \sigma G_\alpha$, where σ is a scale parameter and G_γ is a gamma random variable with mean γ and density

$$dP\left(G_\gamma \leq x\right)/dx = x^{\gamma-1} \exp(-x)/\Gamma(\gamma) \tag{2.1}$$

for $x, \gamma > 0$. To obtain F and h for f_2, note that for $L = \sigma L_0$, if f_0, F_0, h_0 denote f, F, h for L_0 with $d\epsilon_\ell = d\ell$, then

$$\begin{aligned} f_\ell &= \sigma^{-1} f_{0\ell/\sigma}, \\ F(w) &= F_0(w\sigma), \\ h(w) &= \sigma^{-1} h_0(w\sigma), \\ \widetilde{L} &= \sigma \widetilde{L}_0. \end{aligned}$$

Also for $L_0 = G_\alpha$, $F_0 = F_\alpha$ and $h_0(w) = h_\alpha(w)$, where $F_\alpha(w) = (1+jw)^{-\alpha}$, $h_\alpha(w) = h_\alpha(0) b_\alpha (1+jw/2)^{-\alpha_1}$ for $\alpha_1 = 2\alpha - 1 > 0$, $h_\alpha(0) = b_\alpha 2^{-\alpha_1}$, $b_\alpha = \Gamma(\alpha_1)\Gamma(\alpha)^{-2}$. For $\alpha \leq 1/2$, $b_\alpha = \infty$. So, for $f = f_2$, the mean delay is $\sigma\alpha$, $h_0 = \sigma^{-1} h_\alpha(0)$, $\widetilde{L} = \sigma 2^{-1} G_{\alpha_1}$.

Where f_2 is used below we take exponential delay ($\alpha = 1$) with mean delay σ, $b_1 = 1$, $h_0 = (2\sigma)^{-1}$,

$$f_{2\ell} = \sigma^{-1} \exp(-\ell/\sigma), \ F(w) = (1 + jw\sigma)^{-1}, \ h(w) = h_0 (1 + jw\sigma/2)^{-1}, \ \widetilde{L} = \sigma G_1/2.$$

We consider five frequency distributions (1.6):

$$\begin{aligned}
&\text{for } \nu_1, W = w_0 + i\delta \text{ with probability } J^{-1} \text{ for } i = 0, 1, \ldots, J-1;\\
&\text{for } \nu_2, W \sim \text{Uniform}(w_0, w_0 + 2\delta);\\
&\text{for } \nu_3, W = w_0 G_\beta, \text{ where } w_0, \beta > 0;\\
&\text{for } \nu_4, W = \begin{cases} w_0, & \text{with probability } p,\\ w_0 + \delta, & \text{with probability } q = 1 - p; \end{cases}\\
&\text{for } \nu_5, W \sim \text{Uniform}(I_1 \cup I_2)
\end{aligned}$$

for G_γ of (2.1) and I_1, I_2 the non-overlapping intervals $[w_0, w_0+B_1]$, $[w_0+\delta, w_0+\delta+B_2]$. Note ν_2, ν_5 are of standard type (1.7). Note ν_1 approximates a spectrum of J equally spaced narrow bands each of bandwidth B say with total bandwidth JB. Note ν_2 is for one broad band of bandwidth 2δ. Note ν_5 is for two broad bands of bandwidths B_1, B_2 and total bandwidth $B = B_1 + B_2$. Note that for ν_1, ν_2, ν_4 and ν_5, when $\boldsymbol{\mu} = \mathbf{0}$ capacity does not depend on the base frequency w_0 since the shrinkage functions $[\cdot]'$ do not. Their Fourier transforms (1.35), are for $\ell \neq 0$

$$\begin{aligned}
\widetilde{\nu}_{1\ell} &= J^{-1} \exp(-jw_0\ell) \{1 - \exp(-jJ\delta\ell)\} / \{1 - \exp(-j\delta\ell)\},\\
\widetilde{\nu}_{2\ell} &= \exp(-jw_0\ell) \{1 - \exp(-j2\delta\ell)\} / (2\delta\ell),\\
\widetilde{\nu}_{3\ell} &= (1 + jw_0\ell)^{-\beta},\\
\widetilde{\nu}_{4\ell} &= \exp(-jw_0\ell) \{p + q \exp(-j\delta)\},\\
\widetilde{\nu}_{5\ell} &= (Bj\ell)^{-1} \exp(-jw_0\ell) \{1 - \exp(-jB_1\ell) + [1 - \exp(-jB_2\ell)] \exp(-j\delta\ell)\}.
\end{aligned}$$

Note that $\widetilde{\nu}_{1\ell}$ has period 2π in δ, so the same is true for $[12]'$ with $\nu = \nu_1$. As $J \to \infty$, $\widetilde{\nu}_{1\ell} \to 0$, so that $[12]' \to 0$ when $\nu = \nu_1$.

In every case except for ν_4, there is a parameter ϵ say, such that when $\boldsymbol{\mu} = \mathbf{0}$ (Raleigh fading), $C \to c$ as $\epsilon \to \infty$. For, take $\epsilon = J$, δ, w_0 or β, B for $i = 1, 2, 3, 5$. Then as $\epsilon \to \infty$, $\widetilde{\nu}_\ell \to 0$ for $\ell \neq 0$ so that $[12]' \to 0$.

Where ν_3 is used below we take W exponential ($\beta = 1$).

Example 2.1. This continues Example 1.1. Consider Raleigh fading, that is, $\boldsymbol{\mu}_0 = \mathbf{0}$, $\mathbf{V}_0 = v_0 \mathbf{I}_M$. By (1.26), the asymptotic (large N) capacity is given by (1.37).

Example 2.2. Consider Raleigh fading (that is $\boldsymbol{\mu}_0 = \mathbf{0}$) with $\mathbf{V} = v_0 \boldsymbol{\tau}\boldsymbol{\tau}^+$, where $|\boldsymbol{\tau}| = 1$. This includes the regular line of sight case $\boldsymbol{\tau} = M^{-1/2} \mathbf{1}_M$. For example, for $V_{rs} \equiv 1$, $v_0 = M$. Then $\{a_{ri}\}$ except for a_{10} are the same as for Example 2.1 with $M = 1$. Also

$$c_N = c_{Raleigh} = \log_2(1 + \rho_1), \ a_{10} = M \ln \lambda + \ln(1 + \rho_1), \ \gamma_r' = \left(1 + \rho_1^{-1}\right)^{-r}$$

for ρ_1 of (1.36). So, for Raleigh fading the $N-M$ MIMO system behaves like an $N-1$ MIMO system. But for regular line of sight v_0 is amplified by a factor M over its value in Example 2.1, so that the effective SNR is also amplified by a factor M. So, the figures given in Examples 1.1, 2.1 for the case $M=1$ apply with ρ_i interpreted as $M\rho_i$.

More generally, we have

Example 2.3. Let $\{v_i, \tau_i\}$ be the eigenvalues and eigenvectors of $\mathbf{V}_0$. So,

$$\mathbf{V}_0 = \sum_{i=1}^{M} v_i \boldsymbol{\tau}_i \boldsymbol{\tau}_i^+,$$

where $\boldsymbol{\tau}_{i_1}^+ \boldsymbol{\tau}_{i_2} = \delta_{i_1 i_2}$. If $\boldsymbol{\mu}_0 = \mathbf{0}$ then

$$c_{Raleigh} = \sum_{i=1}^{M} \log_2 \left(1 + v_i \widetilde{\rho}\right),$$
$$\gamma_r' = \sum_{i=1}^{M} \left\{ \left(1 + v_i^{-1} \widetilde{\rho}^{-1}\right)^{-r} : 1 \le i \le M, v_i \neq 0 \right\}$$

for $\widetilde{\rho} = h_0 p/(\lambda q)$. A case of interest is that intermediate between $\mathbf{V}_0 = v_0 \mathbf{I}_M$ of Example 2.1 and $\mathbf{V}_0 = v_0 \boldsymbol{\tau}\boldsymbol{\tau}^+$ of Example 2.2, where $|\boldsymbol{\tau}| = 1$. That is $\mathbf{V}_0/v_0 = \eta \mathbf{I}_M + (1-\eta)\boldsymbol{\tau}\boldsymbol{\tau}^+$ for $0 \le \eta \le 1$, with eigenvalues $1, \eta, \ldots, \eta$. So,

$$c_{Raleigh} = \log_2 \left(1 + \rho_1\right) + (M-1) \log_2 \left(1 + \eta \rho_1\right),$$
$$\gamma_r' = \left(1 + \rho_1^{-1}\right)^{-r} + (M-1) \left(1 + \eta^{-1} \rho_1^{-1}\right)^{-r}$$

for $\rho_1 = v_0 \widetilde{\rho} = h_0 v_0 p/(\lambda q)$.

Example 2.4. Consider again Example 2.1, Raleigh fading. The approximations used above can be improved by stabilizing the variance, $a_{21} = M[12]'(1+\rho_1^{-1})^{-2}$ for ρ_1 of (1.36). Set $\widehat{w} = C_0/M$ and $w = a_{10}/M = \ln(\lambda(1+\rho_1))$. Note (1.23) holds for $\widehat{\theta} = \widehat{w}$ with cumulant coefficients $a_{ri0} = M^{-r} a_{ri}$. So, by Withers (1982) (1.23) holds for $\widehat{\theta} = t(\widehat{w})$ for any smooth function $t(\widehat{w})$ with cumulant coefficients given as follows in terms of $t_r = t^{(r)}(w)$:

$$a_{10}' = t(w),$$
$$a_{21}' = t_1^2 a_{210},$$
$$a_{11}' = t_1 a_{110} + t_2 a_{210}/2,$$
$$a_{32}' = t_1^3 a_{320} + 3 t_1^2 t_2 a_{210}^2,$$
$$a_{22}' = t_1^2 a_{220} + t_1 t_2 a_{320} + \left(t_2^2/2 + t_1 t_3\right) a_{210}^2 + 2 t_1 t_2 a_{110} a_{210},$$
$$a_{43}' = t_1^4 a_{430} + 12 t_1^3 t_2 a_{210} a_{320} + 4 \left(3 t_1^2 t_2^2 + t_1^3 t_3\right) a_{210}^3.$$

Applying this to $t(\widehat{w}) = \ln\{\exp(\widehat{w}) - \lambda\} - \ln \rho_3$ gives

$$t_1 = 1 + \rho_1^{-1},$$
$$t_2 = -t_1/\rho_1,$$
$$t_3 = (1+\rho_1)(2+\rho_1)/\rho_1^3,$$

so that $a'_{10} = 0$ and $a'_{21} = M^{-1}[12]'$, that is, the variance has been "stabilized", no longer depending on ρ_1. We illustrate this for the case of no delay. One obtains

$$\begin{aligned}
a'_{21} &= M^{-1}, \\
a'_{11} &= -\left(M\rho_1 + M^{-1}\right)(1+\rho_1)^{-1}/2, \\
a'_{32} &= -M^{-2}\left(3 + 7\rho_1 + \rho_1^2\right)\rho_1^{-1}(1+\rho_1)^{-1}, \\
a'_{22} &= \{\rho_1^3 - 2\rho_1^2 + M^{-2}\left(4\rho_1^2 + 13\rho_1 + 6\right)\}\rho_1^{-1}(1+\rho_1)^{-2}/2, \\
a'_{43} &= 2M^{-3}\left(18 + 37\rho_1 - 10\rho_1^2 - 11\rho_1^3\right)\rho_1^{-1}(1+\rho_1)^{-2}.
\end{aligned}$$

Figures 2.1a, b plot these "stabilized" approximations for the 1 percentiles of capacity for the cases $M \le N = 3$ and $M = N \le 3$. For $M = N = 1$ there is no Law of Large Numbers - so the wild third order curve.

Figures 2.2a, b compare the two methods for $M = N \le 3$. The solid lines are the stabilized approximations and the dashed lines are the original approximations. The a and b figures are for the first and third approximations. Agreement is fairly close: the stabilized version removes the dip and negative values for the case $M = N = 1$. An alternative when $M = 1$ is to use the exact result $C = \log_2(1 + \rho_1 G_N/N)$ for G_N gamma with mean N.

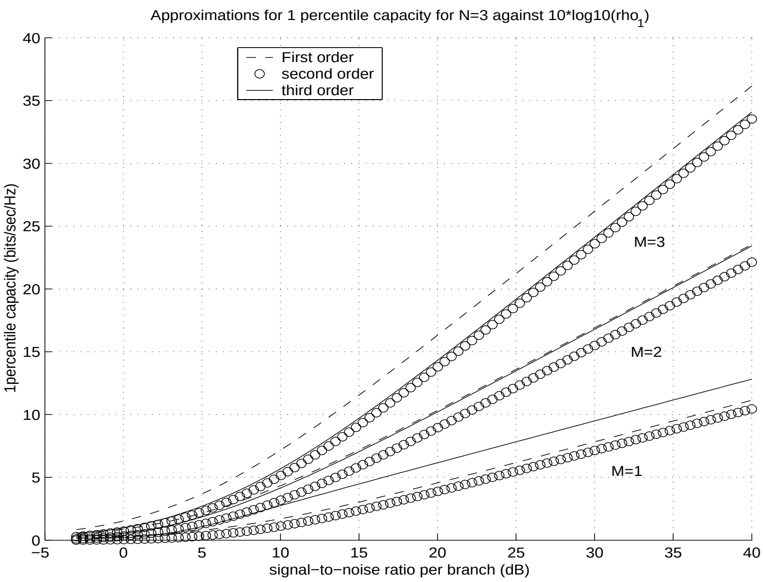

Figure 2.1a First three stabilized approximations to 1 percentile against ρ_1 for $M \leq N = 3$.

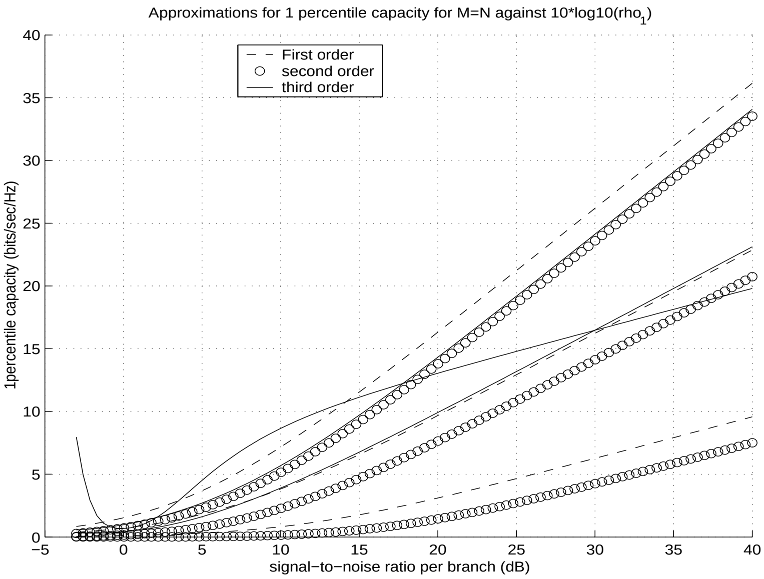

Figure 2.1b First three stabilized approximations to 1 percentile against ρ_1 for $M = N \leq 3$.

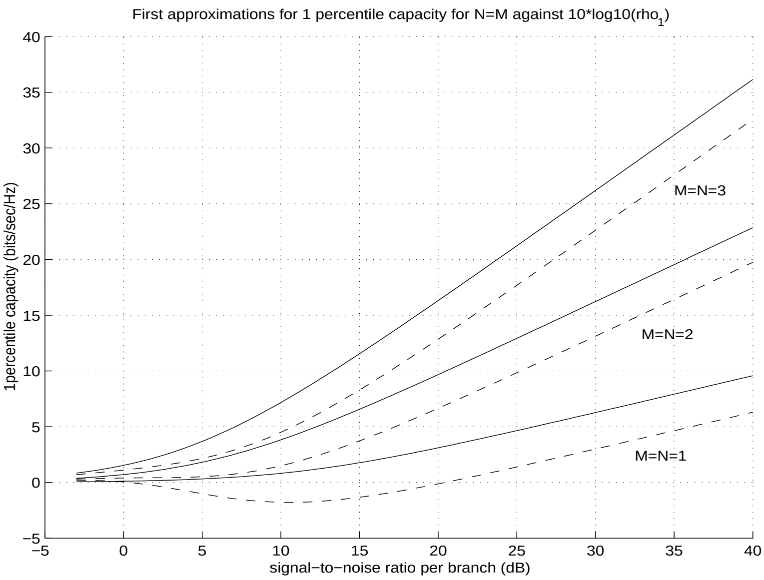

Figure 2.2a First approximations to 1 percentile against ρ_1 for $M = N \leq 3$ (solid for stabilized, dashed for original).

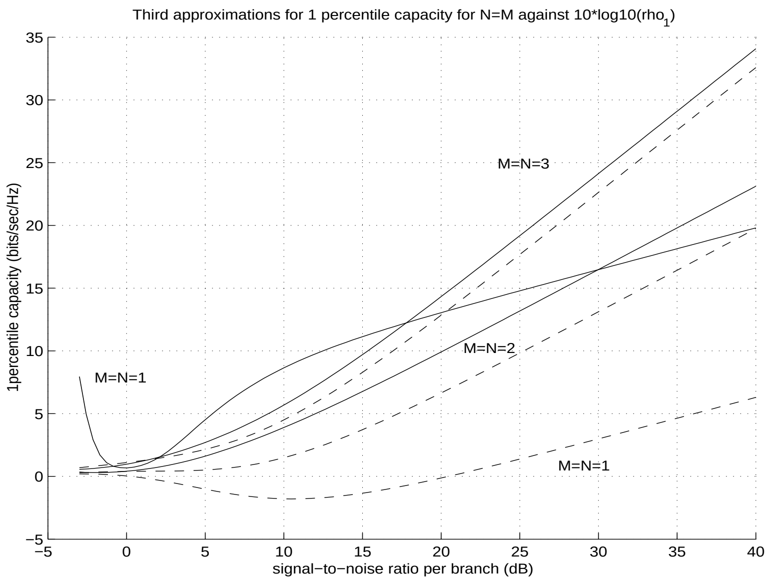

Figure 2.2b Third approximations to 1 percentile against ρ_1 for $M = N \leq 3$ (solid for stabilized, dashed for original).

For a given delay distribution we might seek to choose the spectrum ν to minimize the asymptotic variance. Consider the case of Raleigh fading so that $a_{21} = \gamma_2'[12]'$ for $[12]'$ of (1.34). We now give two examples showing that for delay distribution ν_4, the best we can do is to reduce $[12]'$ to 1/2. For ν_4, $[12]' = 1 - 2pq(1 - E\cos w_1 \Delta_L)$, where Δ_L is the second factor in (1.34); so $[12]'$ has minimum $1/2 + h_0^{-2} \inf_w |h(w)|^2$ achieved at $p = 1/2$ and at $w = \delta$ minimizing $|h(w)|^2$.

Example 2.5. For $f = f_1$ and $I > 1$, $|h(w)|^2$ has a minimum of zero at $w = 2L\pi/I$ for $L = 1, 2, \ldots$, giving $[12]' = 1/2$.

Example 2.6. For $f = f_2$, $|h(w)|^2$ has a minimum of zero at $w_1 = \infty$. So, $[12]'$ can be arbitrarily close to 1/2.

3 Duality

In (1.5), we gave the capacity of an $N - M$ MIMO system with delay speed and frequency spread in terms of $\mathbf{X}_w$ of (1.4). But

$$\det\left(\mathbf{I}_M + \mathbf{X}_w \mathbf{X}_w^+ / (N\lambda)\right) = \det\left(\mathbf{I}_N + \mathbf{X}_w^* {\mathbf{X}_w^*}^+ / \left(M\lambda^*\right)\right),$$

where

$$M\lambda^* = N\lambda,\ \mathbf{X}_w^* = \mathbf{X}_w^+ = \mathbf{D}_p^{1/2} \mathbf{G}_{0w}^* \mathbf{D}_Q^{-1/2},\ \mathbf{G}_{0w}^* = \mathbf{G}_{0w}^+ = \left(\mathbf{H}_{10w}^*, \ldots, \mathbf{H}_{M0w}^*\right), \quad (3.1)$$

say. We use $*$ to denote a dual quantity. Assume that the transmitters and the receivers are independent of each other and behave identically.

Then $\{\mathbf{H}^*_{m0w}\}$ are independent copies of $\mathbf{H}^*_{0w}$, the Fourier transform of a row of $\mathbf{g}(t)$, say

$$\mathbf{g}^*_{0t} = E\mathbf{Z}^*_{0L}\delta(t-L)|\mathbf{Z}_0 = \int \mathbf{Z}^*_{0\ell}\delta(t-\ell)dP(L \leq \ell) = \int \mathbf{Z}^*_{0\ell}\delta(t-\ell)f_\ell d\varepsilon_\ell,$$

where $\mathbf{Z}^*_{0\ell}$ are independent $\mathcal{CN}_N(\boldsymbol{\mu}^*_0, \mathbf{V}^*_0)$. If we write $\mathbf{Z}_{0\ell}$ for the nth column as $\mathbf{Z}_{n0\ell}$ and $\mathbf{Z}^*_{0\ell}$ for the mth row as $\mathbf{Z}^*_{m0\ell}$, then

$$(\mathbf{Z}^*_{10\ell} \cdots \mathbf{Z}^*_{m0\ell}) = (\mathbf{Z}_{10\ell} \cdots \mathbf{Z}_{N0\ell})^+ .$$

This allows transmitters to be correlated and non-stationary with receivers independent, just as Section 1 allowed receivers to be correlated and non-stationary with transmitters independent.

So, we have the same expression for capacity with M, N, p_n, q_m, $\boldsymbol{\mu}_0$, $\mathbf{V}_0$, λ replaced by their dual quantities $M^* = N$, $N^* = M$, q_m^{-1}, p_n^{-1}, $\boldsymbol{\mu}^*_0$, $\mathbf{V}^*_0$, λ^*. So, the results of Section 1 hold with this switch. So,

$$\begin{aligned}
&C^*_0 = N\ln\lambda^* + C\ln 2,\\
&\kappa_r\left(C^*_0\right) = \sum_{i=r-1}^{\infty} a^*_{ri}M^{-i},\\
&a^*_{10} = N\ln\lambda^* + c^*\ln 2,\\
&Y^*_M = \ln 2(M/a^*_{21})^{1/2}\left(C - c^*\right) = (M/a^*_{21})^{1/2}\left(C^*_0 - a^*_{10}\right) \rightarrow \mathcal{N}(0,1)
\end{aligned}$$

as $M \rightarrow \infty$, where

$$\begin{aligned}
&c^* = c^*_{Raleigh} + c^*_{Rice},\\
&c^*_{Raleigh} = \log_2\det\left(\mathbf{I}_N + h_0 p^*_T\mathbf{V}^*\right),\\
&c^*_{Rice} = \int_0^\infty \log_2\left(1 + 2R^*_{\boldsymbol{\mu}}\left|F(w)\right|^2\right)d\nu(w),\\
&p^*_T = \left(\sum_{m=1}^{M} q_m^{-1}\right)/\left(M\lambda^*\right),\\
&R^*_{\boldsymbol{\mu}} = \boldsymbol{\mu}^{*+}\left({p^*_T}^{-1}\mathbf{I}_N + h_0\mathbf{V}^*\right)^{-1}\boldsymbol{\mu}^*/2,\\
&\boldsymbol{\mu}^* = \mathbf{D}_p^{1/2}\boldsymbol{\mu}^*_0,\ \mathbf{V}^* = \mathbf{D}_p^{1/2}\mathbf{V}^*_0\mathbf{D}_p^{1/2}.
\end{aligned}$$

As $M \rightarrow \infty$, C converges to c^* for N fixed even if receivers are correlated.

The distribution, density and quantiles of Y^*_M (and so of C) can be expanded in powers of $M^{-1/2}$ about those of the standard real normal $\mathcal{N}(0,1)$. The Rth order approximations give remainder $O(M^{-R/2})$. The truncated forms of these expansions increase in accuracy as M increases, just as those of Section 1 do as N increases.

For total transmitted power bounded, $\lambda = 1$ so that $\lambda^* = N/M$. For total transmitted power increasing, $\lambda = N^{-1}$ so that $\lambda^* = M^{-1}$.

For Raleigh fading $\boldsymbol{\mu}^* = \mathbf{0}$ so that by (1.30)

$$(\ln 2)^2 v^* = a^*_{21}(\mathbf{0}) = {\gamma'}^*_2[12]'$$

for ${\gamma'}^*_r$ the dual of (1.27):

$$\begin{aligned} {\gamma'}^*_r &= {p^*}_{rM}\Gamma^*_r, \\ {p^*}_{rM} &= M^{-1}\sum_{m=1}^{M} q_m^{-r}, \\ \Gamma^*_r &= {p^*}_{1M}^{-r} \text{ trace}\left(\mathbf{I}_M + h_0 {p^*}_T^{-1}\mathbf{V}^{*-1}\right)^{-r} \end{aligned}$$

if $det(\mathbf{V}^*) \neq 0$. The shrinkage functions $[\cdot]'$ do not change for the dual results. Now suppose that both receivers and transmitters are independent. Then for some scalar τ,

$$\begin{aligned} &\boldsymbol{\mu}_0 = \tau\mathbf{1}_M,\ \boldsymbol{\mu}^*_0 = \overline{\tau}\mathbf{1}_N, \\ &\mathbf{V}_0 = v_0\mathbf{I}_M,\ \mathbf{V}^*_0 = v_0\mathbf{I}_N, \\ &c_{Raleigh} = \sum_{m=1}^{M}\log_2\left\{1 + \left(\sum_{n=1}^{N} p_n\right) q_m^{-1} h_0 v_0 \eta^{-1}\right\}, \\ &c^*_{Raleigh} = \sum_{n=1}^{N}\log_2\left\{1 + p_n\left(\sum_{m=1}^{M} q_m^{-1}\right) h_0 v_0 \eta^{-1}\right\}, \end{aligned}$$

where $\mathbf{1}_M$ is the vector of M ones, $\eta = N\lambda = M\lambda^*$. These agree to a first approximation if $h_0\eta^{-1}v_0$ or $\{p_n q_m^{-1}\}$ are small then $c_{Raleigh}$, $c^*_{Raleigh}$ both equal

$$(\ln 2)^{-1}\left(\sum_{n=1}^{N} p_n\right)\left(\sum_{m=1}^{M} q_m^{-1}\right) h_0 v_0 \eta^{-1} + O\left(h_0^2\eta^{-2}\right).$$

To make a numerical comparison with the examples of Section 2, we also assume that the transmitters are independent of each and behave identically, then $\mathbf{V}_0 = v_0\mathbf{I}_M$, $\mathbf{V}^*_0 = v_0\mathbf{I}_N$, $\boldsymbol{\mu}_0 = \tau\mathbf{1}_M$, $\boldsymbol{\mu}^*_0 = \overline{\tau}\mathbf{1}_N$. Here, $\tau = \mu_{01}$ is zero for the Raleigh case and one for the Ricean case.

Example 3.1. Suppose that $p_n \equiv p$, $q_m \equiv q$, $\mathbf{V}^*_0 = v_0\mathbf{I}_N$. Then for ρ_i of (1.36),

$$\rho^*_3 = \rho_3,\ \rho^*_1 = \rho_1 M/N,\ \rho^*_2 = \rho^*_1/h_0,\ c^*_{Raleigh} = N\log_2\left(1 + \rho^*_1\right).$$

So, for fixed N the effective SNR is not ρ_1 but ρ^*_1. Also $1/\lambda^*$ is proportional to M for fixed N regardless of whether $\lambda = 1$ (bounded total power) or $\lambda = N^{-1}$ (increasing total power). So,

$$
\begin{aligned}
(\ln 2)EC &= M\ln(1+\rho_3) + O\left(N^{-1}\right) \text{ if } \lambda = 1, \\
&= M\ln(\rho_3 N) + O\left(N^{-1}\right) \text{ if } \lambda = N^{-1}, \\
&= N\ln(\rho_3 M) + O\left(M^{-1}\right),
\end{aligned}
$$

where the first two lines hold for fixed M, and the last line holds for fixed N regardless of whether total power is bounded or increasing.

The asymptotic variances of $C\ln 2$ by the two methods are

$$
\begin{aligned}
N^{-1}a_{21}(\mathbf{0}) &= N^{-1}M\left(1+\rho_1^{-1}\right)^{-2}[12]', \\
M^{-1}a_{21}(\mathbf{0})^* &= M^{-1}N\left(1+\rho_1^{*-1}\right)^{-2}[12]'.
\end{aligned}
$$

To recap, for $M \ll N$

$$
\begin{aligned}
&C \approx c_{Raleigh} = M\log_2(1+\rho_1), \\
&\text{var } C \approx (\ln 2)^{-2} M N^{-1} [12]' \left(1+\rho_1^{-1}\right)^2
\end{aligned}
$$

while for $N \ll M$

$$
\begin{aligned}
&C \approx c^*_{Raleigh} = N\ln(1+\rho_1^*), \\
&\text{var } C \approx (\ln 2)^{-2} N M^{-1} [12]' \left(1+\rho_1^{*-1}\right)^2,
\end{aligned}
$$

where $\rho_1{}^* = \rho_1 N M^{-1}$. A different theory applies if M, N are both large and comparable.

All the figures from Figure 1.1 can be re-interpreted by replacing M, $\rho_1, \ldots$ by their duals $N, \rho_1^*, \ldots$ If $M = N$ then dual quantities coincide; for example, $c^*_{Raleigh} = c_{Raleigh}$ and $v^* = v$.

Note that expansions like

$$
(\ln 2)EC = -M\ln\lambda + \sum_{i=0}^{\infty} a_{1i}N^{-i} = -N\ln\lambda^* + \sum_{i=0}^{\infty} a^*_{1i}M^{-i}
$$

are not comparable as the first is likely to diverge for M/N large and the second for M/N small, like the expansions for $\ln(1+x)$ valid for $|x| < 1$ and for $|x| > 1$ when $x = M/N$.

Note that the use of (3.1) has remedied the loss of reciprocity noted in Remark 2 of Telatar (1999).

4 The effect of variable transmitter power

In (1.5) we followed Foschini and Gans (1998) by replacing the random transmitter powers $\{P_n/(N\lambda)\}$ by their means. The values of a_{10} and c_N are not changed but the other $\{a_{ri}\}$ do change. See there for details. For example, a_{21} changes to

$$a_{21} = p_{2N}\Gamma_2 + k_p(2)\Gamma_1^2$$

for Γ_r of (1.29), p_{rN} of (1.28), and

$$k_p(2) = N^{-1}\sum_{n=1}^{N} var\,(P_n)\,.$$

So, if $q_m \equiv q$ and $\mathbf{V}_0 = v_0\mathbf{I}_M$, then

$$a_{21} = M\left(1+\rho_1^{-1}\right)^{-2}\{p_{2N} + Mk_p(2)\}\,p_{1N}^{-2}$$

for $\rho_1 = h_0 v_0 p_T/q$ as in Example 1.1. If also each P_n behaves like a random variable P of mean p then

$$a_{21} = M\left(1+\rho_1^{-1}\right)^{-2}\left\{1 + (M+1)var(P)/p^2\right\}.$$

The effect of allowing variable receiver power is more profound as normality breaks down.

5 The effect of normalization

Consider again Example 1.1 with $\boldsymbol{\mu} = \mathbf{0}$. As $N \to \infty$ capacity approaches $c = M\log_2(1 + h_0 v_0\rho)$, where $\rho = p/(\lambda q)$ is the ratio of the average total power transmitted to the average noise of a single transmitter, as in Foschini and Gans (1998). In fact, this reduces to their result $M\log_2(1+\rho)$ if we normalize by replacing $\mathbf{G}_{0w}$ in (1.2) by

$$\mathbf{G}'_{0w} = NM||\mathbf{G}_{0w}||^{-1}\mathbf{G}_{0w} \tag{5.1}$$

for

$$||\mathbf{G}_{0w}||^2 = \sum_{i=1}^{M}\sum_{k=1}^{N} trace\ covar\ \mathbf{G}_{0wik} = N\ trace\ covar\ \mathbf{H}_{0w},$$

$$trace\ covar\ \mathbf{Z} = \ trace\ covar\ \mathbf{Z} = E|\mathbf{Z}|^2 - |E\mathbf{Z}|^2 \tag{5.2}$$

for $\mathbf{Z}$ a random complex vector. For, this is the same as replacing $\mathbf{Z}_{0t}$ in Section 1 by $NM||\mathbf{G}_{0w}||^{-1}\mathbf{Z}_{0t}$, that is $\boldsymbol{\mu}$ by $NM||\mathbf{G}_{0w}||^{-1}\boldsymbol{\mu}$ and $\mathbf{V}$ by $(NM)^2||\mathbf{G}_{0w}||^{-2}\mathbf{V}$. Let C' denote C when we normalize in this way.

In this example, $||\mathbf{G}_{0w}||^2 = NMh_0v_0$ and if $\boldsymbol{\mu} = \mathbf{0}$ then

$$C'\ln 2 \approx \mathcal{N}\left(M\ln(1+\rho), [12]'\left(1+\rho^{-1}\right)^{-2}M/N\right)$$

as $N \to \infty$ and if there is no delay or for narrowband (1 frequency) C' behaves exactly as if there is no delay and narrowband.

For given $\mathbf{S}_w$ and $\mathbf{E}_w$, (1.2) gives

$$trace\ covar\ \mathbf{R}_w = N^{-1}||\mathbf{G}_{0w}||^2\,|\mathbf{S}_w|^2\,.$$

If $\boldsymbol{\mu} = \mathbf{0}$, the left hand side is equal to $E\{|\mathbf{G}_{0w}\mathbf{S}_w|^2|\mathbf{S}_w\}$. So, if instead of (5.1) we normalize by replacing $\mathbf{G}_{0w}$ in (1.2) by $\mathbf{G}''_{0w} = M^{-1}\mathbf{G}'_{0w}$ then

$$trace\ covar\ \mathbf{R}''_w = |\mathbf{S}_w|^2\,,$$

where $\mathbf{R}''_w$ is $\mathbf{R}_w$ with $\mathbf{G}_{0w}$ replaced by $\mathbf{G}''_{0w}$. Some authors, for example, Chiurtu *et al.* (2001), argue that one should use this second normalization to ensure that the total receive power equals the total transmit power when averaged over the random channel matrix. This has the effect of replacing ρ by ρ/M so that the large N capacity becomes $M\log_2(1+\rho/M) \to \rho$ as $M \to \infty$. However, here we do not need to normalize keep capacity finite when both $N, M \to \infty$.

For the general case by (1.16)

$$||\mathbf{G}_{0w}||^2 = Nh_0\ trace\ \mathbf{V},$$

so for $\boldsymbol{\mu} = \mathbf{0}$ this normalization gives

$$C'\ln 2 \approx \mathcal{N}\left(\ln\det\left(\mathbf{I}_M + p_T M\mathbf{V}/\ trace\ \mathbf{V}\right), [12]'p_{2N}\Gamma'_2/N\right),$$

where $\Gamma'_r = p_{1N}^{-r}\ trace\ (\mathbf{I}_M + M^{-1}\mathbf{V}^{-1}\ trace\ \mathbf{V})^{-r}$ for $\det\mathbf{V} \neq 0$, and again if there is no delay or narrowband C' behaves exactly as if there is no delay and narrowband.

One cannot drop the term $E\mathbf{Z}$ in (5.2) if $\boldsymbol{\mu} \neq \mathbf{0}$ as otherwise $||\mathbf{G}_{0w}||^2$ will depend on the frequency w.

6 Conclusions

The formula of Foschini and Gans (1998) for capacity has been extended to allow for

- transmitters of different powers;
- receivers of different powers;
- correlated transmitters (for the dual expansion);
- correlated receivers;
- Ricean as well as Rayleigh fading;
- line of sight;
- multiple frequencies;

- delay spread;
- normalization.

In each case we have shown how to obtain approximate normality, and how to calculate the distribution and outage probabilities as power series in $N^{-1/2}$ (or in $M^{-1/2}$ in the dual case).

We have shown that one or two terms in these expansions is usually sufficient unless N is extremely small or N and M are close in magnitude.

We have shown that by spreading the frequencies used one can reduce the variance of the capacity, and so achieve a given outage probability with less total power.

References

[1] H. Boche and E. Jorswieck, *Analysis of diversity and multiplexing tradeoff for multi-antenna systems with covariance feedback*, in: Proceedings of the IEEE 56th Vehicular Technology Conference, Vancouver, BC, Canada, volume 2, 2002, pp. 864868.

[2] N. Chiurtu, B. Rimoldi and E. Telatar, *Dense multiple antenna systems*, in: Proceedings of the 2001 IEEE Information Theory Workshop, Cairns, Australia, 2001, pp. 108-109.

[3] G. J. Foschini and M. J. Gans, On limits of wireless communications in a fading environment when using multiple antennas, *Wireless Personal Communications* **6** (1998), 311-335.

[4] W. Hachem, P. Loubaton and J. Najim, A CLT for information theoretic statistics of gram random matrices with a given variance profile, *Annals of Applied Probability* **18** (2008), 2071-2130.

[5] B. M. Hochwald, T. L. Marzetta and V. Tarokh, Multiple-antenna channel hardening and its implications for rate feedback and scheduling, *IEEE Transactions on Information Theory* **50** (2004), 1893-1909.

[6] I. M. Johnstone, On the distribution of the largest eigenvalue in principal components analysis, *Annals of Statistics* **29** (2001), 295-327.

[7] C. Martin and B. Ottersten, *Analytic approximations of eigenvalue moments and mean channel capacity for MIMO channels*, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, volume 3, 2002, pp. 2389-2392.

[8] A. L. Moustakas, S. H. Simon and A. M. Sengupta, MIMO capacity through correlated channels in the presence of correlated interferers and noise: a (not so) large N analysis, *IEEE Transactions on Information Theory* **49** (2003), 2545-2561.

[9] K. I. Pedersen, J. B. Andersen, J. P. Kermoal and P. Mogensen, *A stochastic multiple-input-multiple-output radio channel model for evaluation of space-time coding algorithms*, in: Proceedings of the 2000 IEEE Vehicular Technology Conference, Boston, MA, volume 2, 2000, pp. 893-897.

[10] G. G. Raleigh and J. M. Cioffi, Spatio-temporal coding for wireless communication, *IEEE Transactions on Communications* **46** (1998), 357-366.
[11] P. Smith and M. Shafi, *On the capacity of MIMO Systems*, Technical Report, Department of Electrical Engineering, Canterbury University, 2001.
[12] I. E. Telatar, Capacity of multi-antenna Gaussian channels, *European Transactions on Telecommunications* **10** (1999), 585-595.
[13] I. E. Telatar and D. N. C. Tse, Capacity and mutual information of wideband multipath fading channels, *IEEE Transactions on Information Theory* **46** (2000), 1384-1400.
[14] A. Tulino and S. Verdu, Random matrix theory and wireless communications, *Foundations and Trends of Communications and Information Theory*, volume 1 (2004).
[15] Z. Wang and G. B. Giannakis, Outage mutual information of space-time MIMO channels, *IEEE Transactions on Information Theory* **50** (2004), 657-662.
[16] J. H. Winters, On the capacity of radio communication systems with diversity in a Rayleigh fading environment, *IEEE Journal on Selected Areas in Communications* **SAC-5** (1987), 871-878.
[17] C. S. Withers, The distribution and quantiles of a function of parameter estimates, *Annals of the Institute of Statistical Mathematics* A **34** (1982), 55-68.
[18] C. S. Withers, Asymptotic expansions for distributions and quantiles with power series cumulants, *Journal of the Royal Statistical Society* B **46** (1984), 389-396.
[19] C. S. Withers and R. Vaughan, *The distribution and percentiles of channel capacity for multiple arrays*, in: Proceedings of the 11th Symposium on Wireless Communications, Virginia Tech, Blacksburg, Virginia, 2001, pp. 141-152.

Christopher Withers is Senior Scientist at the Induatrial Research Limited, Lower Hutt, New Zealand. His research areas include signal processing, extreme value theory and Cumsum statistics.

Saralees Nadarajah is a Senior Lecturer in the School of Mathematics, University of Manchester, UK. His research interests include climate modeling, extreme value theory, distribution theory, information theory, sampling and experimental designs, and reliability. He is an author/co-author of four books, and has over 300 papers published or accepted. He has held positions in Florida, California, and Nebraska.