Wireless Personal Communications **6:** 311–335, 1998.

On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas

G.J. FOSCHINI and M.J. GANS
Lucent Technologies, Bell Labs. Innovations, Crawford Hill Laboratory – R137, 791 Holmdel-Keyport Road, Holmdel, New Jersey 07733-0400, U.S.A.

Abstract. This paper is motivated by the need for fundamental understanding of ultimate limits of bandwidth efficient delivery of higher bit-rates in digital wireless communications and to also begin to look into how these limits might be approached. We examine exploitation of multi-element array (MEA) technology, that is processing the spatial dimension (not just the time dimension) to improve wireless capacities in certain applications. Specifically, we present some basic information theory results that promise great advantages of using MEAs in wireless LANs and building to building wireless communication links. We explore the important case when the channel characteristic is not available at the transmitter but the receiver knows (tracks) the characteristic which is subject to Rayleigh fading. Fixing the overall transmitted power, we express the capacity offered by MEA technology and we see how the capacity scales with increasing SNR for a large but practical number, n, of antenna elements at *both* transmitter and receiver.

We investigate the case of independent Rayleigh faded paths between antenna elements and find that with high probability extraordinary capacity is available. Compared to the baseline $n = 1$ case, which by Shannon's classical formula scales as one more bit/cycle for every 3 dB of signal-to-noise ratio (SNR) increase, remarkably with MEAs, the scaling is almost like n more bits/cycle for each 3 dB increase in SNR. To illustrate how great this capacity is, even for small n, take the cases $n = 2$, 4 and 16 at an average received SNR of 21 dB. For over 99% of the channels the capacity is about 7, 19 and 88 bits/cycle respectively, while if $n = 1$ there is only about 1.2 bit/cycle at the 99% level. For say a symbol rate equal to the channel bandwith, since it is the bits/symbol/dimension that is relevant for signal constellations, these higher capacities are not unreasonable. The 19 bits/cycle for $n = 4$ amounts to 4.75 bits/symbol/dimension while 88 bits/cycle for $n = 16$ amounts to 5.5 bits/symbol/dimension.

Standard approaches such as selection and optimum combining are seen to be deficient when compared to what will ultimately be possible. New codecs need to be invented to realize a hefty portion of the great capacity promised.

Key words: capacity, information theory, diversity, matrix channel, array.

1. Introduction

We report a theoretical and numerical estimation of the ultimate limits of bandwidth efficient delivery of high bit-rate digital signals in wireless communication systems when multi-element arrays (MEAs) are used. An information-theoretic approach is taken to investigate, for a fading environment, the value of using a substantial number of antenna elements at *both* transmitter and receiver. We constrain the channel bandwidth and total transmitted power and show that by forming a channel using increased spatial dimensions one can get extraordinarily large capacity.

The analysis is conducted in an idealized propagation context tailored to give insight into wireless LAN applications and other wireless applications where at least *nominally* there is extremely limited mobility, like when transmit and receive MEAs are affixed to buildings. None-the-less, we allow that changes in the propagation environment occur but on a very slow time scale compared to the burst rate. E.g., a user at a desk on a LAN can move within the workspace and cause the channel to change. Our idealized model allows the channel, fixed

during a burst, to randomly change from burst to burst. We conduct a "quasi-static" analysis, in which one calculates capacity as if a randomly selected channel is unchanged during a burst. The burst is assumed to be of long enough duration that the standard infinite time horizon view of information theory offers meaningful results. (Say, e.g., for several megasymbols per second burst rate with several thousand symbols in a burst with the channel changing on a scale of seconds.) The channel characteristic is not known at the transmitter but the receiver knows (tracks) the characteristic which is subject to Rayleigh fading. Lack of knowledge of the channel characteristic at the transmitter is deemed to be a practical assumption as otherwise a fast feedback link would be required. Significant processing can be involved in exploiting MEAs, so it is of interest to avoid the additional complication of accommodating yet another layer of processing to include a feedback link. Furthermore, if we had allowed feedback to inform the transmitter, the extra time involved in incorporating the feedback loop could erode the validity of assuming that the channel is virtually unchanged.

This paper is restricted to analyzing the narrow band case where the bandwidth is taken to be narrow enough that the channel can be treated as flat over frequency. Then we express capacity in units of bps/Hz, or, equivalently, bits/cycle. For illustrative purposes, we concentrate on the case of an equal number of transmit and receive antennas but our numerical results will also include comparisons with more standard diversity methods like when there is only one transmitting antenna but many receiving antennas and vice-versa. We will often assume an environment of a large number of scatterers so that the Rayleigh fading model is appropriate. The assumption of *independent* Rayleigh paths that we will also often make, is to be thought of as an idealized version of the result that for antenna elements placed on a rectangular lattice with half wavelength ($\lambda/2$) spacing, the path losses tend to roughly decorrelate [1]. Note that, for example, with a 5 GHz carrier frequency $\lambda/2$ is only about 3 cm. So at sufficiently high carrier frequencies there can be great opportunity for accommodating numerous antennas in the regions of space occupied by the communicating stations.

Mostly, capacity is treated here as a random variable and a key goal is to find the complementary cumulative distribution functions (ccdfs). Such functions show how enormous the capacity can be. For the baseline case of a single transmit and a single receive antenna, it is well known that Shannon's classical capacity formula indicates that in the high signal-to-noise (SNR) ratio realm a 3 dB increase in SNR is roughly worth about one more bit/cycle of capacity. Analyzing the case of independent Rayleigh faded paths between n antenna elements at both transmitter and receiver, we will find that, remarkably, for large n the scaling is n more bits/cycle for every 3 dB SNR improvement. We will set a threshold percentage, say, e.g., 99%, and then read from the ccdf graphs the capacity that we can provide with 99% probability. This 99% level amounts to 1% probability of outage (write $P_{\text{out}} = 1\%$). For a symbol rate equal to the channel bandwidth the achievable capacity may, in many cases, at first seem inordinately high, some examples will involve tens, even hundreds, of bits/cycle. However, viewed in terms of bits/cycle/dimension the capacity will be seen to be much more reasonable.

We will see the capacity deficiency of standard MEA architectures relative to when n antenna elements are used at both transmitter and receiver. Looking toward implementations, it is of value for an architecture to be based on the highly developed technology of one dimensional (1-D) codecs. For the simplest interesting case, $n = 2$, we report the relative merit of some efficient architectures that employ two antennas at both communication sites and that are also based on 1-D codecs.

The exceedingly large capacities that we uncover opens up the path to achieve the ambitious bit-rates at very low error rates that will be needed in the future. Indeed, the work reported here of the effect on capacity of deploying MEAs was originally motivated by the growing efforts on evolving indoor wireless LAN technology [2–4]. There is very extensive work on the subject of MEAs. E.g., see [1] and [5–20] and their references for leads into this literature. It will be evident from the results that we will present here, that those capacities pale in comparison with capacities that can be projected when MEAs are deployed at both transmitter *and* receiver.

We will not deal here with the subject of multiple access (or interferers) but we will study a form of self-interference. In the references joint detection [20] and interference cancellation [17] schemes are employed to improve the performance of multiple user systems. Regarding other related work we also mention [21] where measurement and ray tracing calculations were used to examine indoor path loss variations versus both transmitter and receiver local displacements.

2. Mathematical Model for Wireless Channels

Our focus is on a single point-to-point channel. The perspective is a complex baseband view involving a fixed linear matrix channel with additive white gaussian noise (AWGN). Although fixed, the channel will often be taken to be random. Time is taken to be discrete. We need to list more notation and some basic assumptions:

- Number of antennas: Transmit: n_T and Receive: n_R. We will use the descriptor (n_T, n_R).
- Transmitted signal $s(t)$: The total power is constrained to $\hat{P}$ regardless of the value of n_T (the dimension of $s(t)$). The bandwidth is narrow enough that we can treat the channel frequency characteristic as flat over frequency.
- Noise at receiver $\nu(t)$: complex $n_R - D$ AWGN with statistically independent components of identical power N at each of the n_R receiver branches.
- Received signal $r(t)$: $n_R - D$ received signal so that at each point in time there is one complex vector component per receive antenna. When there is only one transmit antenna, it radiates power $\hat{P}$ and we denote the average power at the output of each of the receiving antennas by P.
- Average SNR at each receiver branch: $\rho = P/N$ independent of n_T.
- Matrix channel impulse response: $g(t)$ has n_T columns and n_R rows. We use $G(f)$ for the Fourier transform of $g(t)$. Consistent with the narrowband assumption, we treat this (matrix) transform as constant over the band of interest, writing G, suppressing the frequency dependence. So, except for $g(0)$, $g(t)$ is the zero matrix. Often, it will be convenient to represent the matrix channel response in normalized form, $h(t)$. Specifically, related to G, we have the matrix H, where the equation $\hat{P}^{1/2} \cdot G = P^{1/2} \cdot H$ defines the relationship so, $g(t) = (P/\hat{P})^{1/2} \cdot h(t)$. We say more about the explicit form of H for the important Rayleigh fading case in Section 4.

The following standard notation will be needed: $'$ for vector transpose, $\dagger$ for transpose conjugate, det for determinant, I_n for the $n \times n$ identity matrix, $E\{.\}$ for expectation and $*$ for convolution.

The basic vector equation describing the channel operating on the signal is

$$r(t) = g(t) * s(t) + \nu(t). \tag{1}$$

The two vectors being added are complex $n_R - D$ vectors ($2n_R$ real dimensions). Using the narrowband assumption, we simplify, replacing convolution by product and write

$$r(t) = (P/(\hat{P} \cdot n_T))^{1/2} \cdot h(0) \cdot s(t) + \nu(t). \tag{2}$$

3. Generalized Capacity Formula and Some Examples

The standard formula for the Shannon capacity [22–24] expressed in bps/Hz is

$$C = \log_2(1 + \rho \cdot |H|^2). \tag{3}$$

where the normalized channel power transfer characteristic is $|H|^2$. (in this 1-D case H is simply a complex scalar.) It is evident that for high SNRs a 3 dB increase in ρ gives another bit/cycle capacity.

We are interested in capacity computations having to do with space diversity in which the receiver takes full advantage of what the vector r conveys about the transmitted signal. Assume that, since the transmitter does not know the channel, the transmitted signal power is simply distributed equally over the n_T transmitting antennas. The n_T transmitted signal components are taken to be statistically independent gaussians. The virtue of a gaussian distribution for the transmitted signal is well established, see [25], and for more details involving matrix channels see [26]. (The collection of random variables comprising the n_T real, along with the n_T imaginary parts of all of the transmitted signal components are assumed to be a set of $2n_T$ equal power independent gaussians.) The reader may wonder: independence, why independence? After all, there are many signal paths between transmitter elements and receiver elements and in case a path turns out poorly should we not have redundancy to protect against such a weak signal path? Otherwise would not the capacity simply be n_T times the minimum of the capacities attained with the n_T component signals? To answer this, we note that, through the device of a one-time pad [27], shared by transmitter and receiver, it is not contradictory for the transmitted signal components to be, at the same time, statistically independent of each other and redundant.

To explain how this can be we give a very exaggerated illustration in this paragraph. Start with a bit stream, in the form of, say, an i.i.d. sequence of bits each equally likely to be a zero or a one. Moreover, we have n_T identical copies of this same stream. Furthermore, we have n_T transmitted signals conveying n_T identical bit streams all encoded in precisely the same way into n_T identical gaussian waveforms. Now say we want instead n_T independent gaussian waveforms but each encoding the same original bit stream. One can simply convert the n_T identical data streams into n_T statistically independent data streams with a one-time pad. We just subject each uncoded bit to an independent flip of a fair coin and change the bit or not according to a head or tail outcome. The list of n_T coin flip outcomes elaborated over time constitutes the one-time pad. The receiver is privy to the one-time pad and treats it as part of the encoding process. Encoding the bits streams after the one-time pad operation gives n_T statistically independent gaussian waveforms each containing precisely the same information content. For this example of exaggerated statistical dependence we see that the encoded signal has n_T-fold redundancy and n_T uncorrelated gaussian component waveforms, both at the same time.

Assuming that the transmitted signal vector is composed of n_T statistically independent equal power components each with a gaussian distribution, the relevant capacity expression can be derived (see Appendix). The convenient general capacity expression is

$$C = \log_2 \det[I_{n_R} + (\rho/n_T) \cdot HH^{\dagger}]\ bps/Hz. \tag{4}$$

In the rest of this section, and in Sections 4–7, we will specify (n_T, n_R) as well as certain constraints on the means of transmission or reception or both, and then determine capacity under the imposed constraints. Formula (4), or simple variants of the formula, will be used repeatedly.

We next apply (4) to some special cases. To hint at results to come, the example where $H = I_n$ is worth registering. We get

$$C_n = n \cdot \log_2(1 + (\rho/n)) \to \rho/\ln(2) \text{ as } n \to \infty. \tag{5}$$

Unlike in (3), capacity scales linearly, rather than logarithmically, with increasing SNR. This hints at a significant advantage to using parallel information transmission. In this simple example we have *orthogonal* parallel channels. However, as we shall see, with the parallelism offered by deploying MEAs at both sites, the signal component traversing different paths can interfere and consequently any possibility of advantage must be carefully assessed. We will do this in Section 4 analytically and then numerically in Sections 5–7. Drawing on (4) we will see, that in contrast to what is expressed by (5), with certain MEA arrangements there is a very significant capacity improvement with increasing n.

It is helpful to consider (5) in terms of a specific example of n uncoupled transmission lines ($H = I_n$). If we choose to send all our power on one of the lines, our capacity is only $\log_2(1 + \rho)$, which is much less than indicated in (5). A simple way in which the equality in (5) can be achieved is to divide the power equally between the n lines and send n equal rate independent signals. (We stress that we have been careful to account for the effect of equal power division on the received SNR.) Thus, one should use all the lines available. In effect, wireless communication has many lines available through radiation, and in many cases one should use these multiple lines. Of course the "radiation" lines are coupled and subject to fading, but as we will show, these "problems" can be dealt with through the use of redundant signals and detection schemes which take advantage of the multi-dimensionality of the channel.

Next we present the formula for optimum receive diversity, showing how the reception of the signal from one transmitter by many receivers increases capacity. Optimum refers to taking full advantage of what the received vector $r = (r_1(t), r_2(t), \cdots r_{n_R}(t))'$ tells us about the transmitted signal. We call this system an optimum combining, OC(n_R), system. Its capacity is

$$C = \log_2 \left[1 + \rho \cdot \sum_{i=0}^{n_R} |H_i|^2\right]. \tag{6}$$

Contrasting with Equation (2), we see that $|H|^2$ is replaced by a sum of squares and that the noise power level is just that of (any) one of the n_R channels. While we found this capacity imposing no constraint of linear combining of the antenna outputs, the capacity expressed by (6) turns out to be exactly that *linear* combination of the antenna outputs that maximizes the information that the output contains about the input signal. This optimum linear combiner is the well known "maximal ratio combiner". In what follows we will present applications

involving random channels where this n_R-fold diversity serves the dual role of capturing more of the transmitter power and stabilizing channel (spatial) fluctuations. This standard method of diversity will be contrasted with numerous other methods.

Selection diversity capacity is that offered by the best of the n_R channels, i.e.,

$$C = \max_m \log_2[1 + \rho \cdot |H_m|^2] = \log_2[1 + \rho \cdot \max_m |H_m|^2]. \tag{7}$$

The maximization is over $\{m \mid 0 \leq m \leq n_R\}$. While selection diversity is inferior to optimum combining, it can be simpler to implement. We will need additional capacity formulas for when MEA technology is more fully exploited, but first we require some notation for the channel of primary interest.

4. Capacities when the Matrix Channel is Rayleigh

A "quasi-static" analysis in which capacity is calculated as if a randomly selected channel is essentially not changing during a (long) burst is employed. The idea is that, whether for LAN or inter-building applications, channels change, but, the time constant of change is assumed negligible on the scale of a burst duration. E.g., an application might involve a burst rate of several Mbps so that while a burst only lasts for, say, a few ms, it spans such a large number of symbols that the standard infinite time horizon information theory view is a valuable idealization. For each channel realization we use (4) to get capacity. As already mentioned, we aim to plot graphs of capacity ccdfs.

The random channel model we treat here is the Rayleigh channel model. The Rayleigh channel model for H ($n_R \times n_T$ channel matrix) has i.i.d., complex, zero mean, unit variance entries:

$$H_{ij} = \text{Normal}(0, 1/\sqrt{2}) + \sqrt{-1} \cdot \text{Normal}(0, 1/\sqrt{2}), \tag{8}$$

consequently, $|H_{ij}|^2$ is a χ_2^2 variate but normalized so $E|H_{ij}|^2 = 1$. Regarding the i.i.d. assumption, independence is usually just an approximation, which improves as array element spacing becomes large compared to λ [15–16]. However, we expect that with the i.i.d. idealization capacity estimates can be useful even for spacings as small as $\lambda/2$, because good diversity performance maintains even for correlation as high as 0.5 [1].

While the H matrix is assumed to have been measured at the receiver, say using a training preamble, the channel matrix is not known at the transmitter. From [28], we have that, in a random channel context such as ours, a single code can be used by the transmitter even though H is not known at the transmitter and any one of an ensemble of channels might be realized. So remarkably, the results in [28] assure us that, in principle, robust codes are possible that can cope with a wide range of possibilities for H. Later, in Section 6 and especially in Section 7, in some simple contexts we will indicate ways that coded modulations can be structured to deal effectively with great uncertainty as to what the channel realization will be.

The capacities of some Rayleigh channels are now expressed using chi-squared variates. We will often not follow the usual custom of using one letter for a random variable and then adding a phrase to convey how the random variable is distributed. Instead, in a slight abuse of standard notation, we opt for a more expressive format of directly indicating the type of distribution of the random variable. We start with a simple example exhibiting this type of notation.

(A) NO DIVERSITY: $n_t = n_R = 1$

$$C = \log_2[1 + \rho \cdot \chi_2^2]. \tag{9}$$

So the parameter ρ multiplies a chi-squared variate with two degrees of freedom.

(B) RECEIVE DIVERSITY: OC(n), $n_T = 1$, $n_R = n$

$$C = \log_2[1 + \rho \cdot \chi^2_{2n}]. \tag{10}$$

Contrast this with the selection diversity formula $\log_2[1 + \rho \cdot \max\{n \text{ independent } \chi^2_2 s\}]$.

(C) TRANSMIT DIVERSITY: $n_T = n$, $n_R = 1$

$$C = \log_2[1 + (\rho/n) \cdot \chi^2_{2n}]. \tag{11}$$

Compare with $\log_2[1+\rho\cdot\chi^2_2]$ which is the formula for when all transmitted signal components are exactly the same. One often hears that with transmit diversity the SNR improves by $10 \cdot \log_{10} n$ as if the capacity were $\log_2[1 + n \cdot \rho \cdot \chi^2_2]$. We would expect that behavior if each of the n_T transmitted signal components was the same and each had power $\hat{P}$. That is not permitted since we require that the total transmitted power is held constant independently of n_T. In our setting, the average SNR (averaged over channel realizations) is ρ at each receiving antenna no matter how many transmitting antennas are used.

(D) COMBINED TRANSMIT-RECEIVE DIVERSITY: We will only need $n_T = n_R$, however, $n_T \geq n_R$ is easily accommodated so we include it as well. (For the interesting ($n_T < n_R$) case, namely $n_T = 1$ we already know the exact capacity formula.) In Section 5.1 we will prove the following lower bound on capacity:

$$C > \sum_{k=n_T-(n_R-1)}^{n_T} \log_2[1 + (\rho/n_T) \cdot \chi^2_{2k}]. \tag{12}$$

Contrast this bound with $\sum_{i=1}^{n_T} \log_2[1 + (\rho/n_T) \cdot \chi^2_{2n_R i}]$ where i is used to index statistically independent chi-squared variates each with $2n_R$ degrees of freedom.) This later formula is an upper bound on capacity. It represents the very artificial case when each of n_T transmitted signal components is received by a separate set of n_R antennas in a manner where each signal component is received with no interference from the others. In other words when the vector components are conveyed over n_T "channels" that are uncoupled and each channel has a separate set of n_R receive antennas. With such a contrivance there are a total of $n_R \cdot n_T$ receive antennas.

(E) SPATIAL CYCLING USING ONE TRANSMITTER AT A TIME: We will compare the right hand side of (12) with what is obtained when n_T transmitters as well as n_R receivers are used but only one transmitter is used at a time: we cycle through all n_T transmitters periodically with period n_T. Beside being associated with a simple implementation, this method has the feature that the cycling assures nontrivial dwelling on the better of n_T transmitters as well. This method avoids catastrophic interference, indeed, with this method there is no interference at all. This technique represents a minor form of combined transmit-receive diversity and since we are using only one transmitter at time, is a singular departure from the ongoing assumption in our examples of using all transmitters simultaneously. The capacity formula for spatial cycling is

$$C = (1/n_T) \cdot \sum_{i=1}^{n_T} \log_2[1 + \rho \cdot \chi^2_{2n_R i}]. \tag{13}$$

The transmitter does not know the channel so this formula needs proof (see Section 4.1.2).[1]

We will see (Section 4.1.1.1) that for (D) for the $(n,\ n)$ case, capacity grows linearly with large n. The spatial cycling method (E) is of much more limited merit as we will see in Section 6. Simple analysis shows that the capacity given by (13) (and (10) for B)) grows only logarithmically with n. Note that the variance of the random capacity that is given by Equation (13) goes to zero as $n \to \infty$ and it does so even more rapidly than the variance of the capacity given by (10).

4.1. Derivations Associated with Combined Transmit-Receive Diversity: Cases (D) and (E)

We now give the mathematical details associated with results we have stated on the more complex of the cases we are addressing, namely, cases (D) and (E). First we deal with the lower bound on capacity given by (12). We will also include an asymptotic, large n, analysis of this bound. Secondly, we give the mathematical justification for (13). This justification will prove quite useful since the essentials of the argument that we give will apply to several other communication systems that we will encounter in the numerical examples sections. (These are examples that use related forms of spatial cycling.)

4.1.1. *Derivation of the Lower Bound on Capacity*

To derive the lower bound we need a result on random matrices from [29] that builds on earlier work in [30]. We write m for n_T and n for n_R. We will employ the notion of *unitarily equivalent rectangular matrices*, see [31].

H is unitarily equivalent to an m by n matrix

$$\begin{bmatrix} x_{2m} & & & & 0 \cdots 0 \\ y_{2(n-1)} & x_{2(m-1)} & & & \cdot \quad \cdot \\ & \ddots & \ddots & & \vdots \quad \vdots \\ & & y_2 & x_{2(m-(n-1))} & 0 \cdots 0 \end{bmatrix} \tag{14}$$

where x_j^2 and y_j^2 are distributed as χ^2 variables with j degrees of freedom (i.e., χ_j^2).[2] *The x and y elements here are all non-negative and independent. All matrix entries below the y stripe and to the right of the x stripe are understood to be zero.*

[1] Note, that when the transmitter *does* know the channel, assuming that the same power, (P/n_T), is transmitted out of each transmitter, Equation (13) also expresses the capacity. In that case the proof of (13) follows easily from n_T applications of Equation (6), one for each statistically independent row of the H matrix. This is because the squared modulus of the entries of the H matrix are each independent chi-squared with two degrees of freedom. Each column of H corresponds to a different transmitter so an independent n_R-fold sum of chi-squared variates with two degrees of freedom is associated with each transmitter. The $((1/n_T)$ multiplier is because each of the transmitters is used $(1/n_T)$ of the time.

[2] The χ_j^2 distribution is usually defined as the distribution of j sums of squares of standard gaussian variates. As mentioned at the beginning of this section, we use a slightly modified form of this definition where the underlying *complex* gaussian variates are normalized so that the real gaussians have standard deviation $1/\sqrt{2}$.

We use the representation (14) to rewrite $I_{n_R} + (\rho/n_T)HH^\dagger$. First define $X_{2j} = (\rho/n_T)^{1/2}x_{2j}$ and $Y_{2j} = (\rho/n_T)^{1/2}y_{2j}$ so $I_{n_R} + (\rho/n_T)HH^\dagger$ is the $n_R \times n_R$ matrix of the form

$$\begin{bmatrix} 1+X_{2m}^2 & X_{2m}Y_{2(n-1)} & & & & \\ Y_{2(n-1)}X_{2m} & 1+X_{2(m-1)}^2+Y_{2(n-1)}^2 & X_{2(m-1)}Y_{2(n-2)} & & & \\ & Y_{2(n-2)}X_{2(m-1)} & 1+X_{2(m-2)}^2+Y_{2(n-2)}^2 & X_{2(m-2)}Y_{2(n-3)} & & \\ & & & \ddots & & \\ & & & & Y_2X_{2(m-(n-2))} & 1+X_{2(m-(n-1))}^2+Y_2^2 \end{bmatrix}. \tag{15}$$

We have displayed all nonzero entries in the first three and last rows.

In evaluating $\det(I_{n_R} + (\rho/n_T)HH^\dagger)$ we get, from the product of the n main diagonal terms in (15), a contribution of the form $L + Q$ where

$$L = \prod_{j=m-(n-1)}^{m} (1 + X_{2j}^2)$$

and Q is a positive number obtained by summing over only positive terms. The determinant calculation involves a signed sum over $n!$ permutations. Every permutation is a product of transpositions and the negatively signed contributions to the determinant involve an odd number of transpositions. Due to the structure exhibited in (15) it is evident that each such negative term is cancelled by a distinct positive contribution to Q. Moreover, Q contains more terms than needed to cancel all the negative terms. Therefore, $C > L$ with probability one.

4.1.1.1 Capacity Lower Bound Analysis for a Large Number of Antennas. We explore further the special case when $n_T = n_R = n$ deriving the large n asymptotics for this lower bound $L = L(n)$. We will look at bounding capacity on a per dimension basis, that is we will look at the lower bound $L(n)/n$ of $C(n)/n$. It is convenient to study the sequence of random variables $[L(n)/n]_{n>0}$ via a perturbation analysis with $(1/n)$ the perturbation parameter. We start by rewriting the right hand side of (12) in terms of centered (at zero) chi-squared variates. The centered variates are indicated with γs in the place of χ^2s

$$c(n)/n > L(n)/n = (1/n) \cdot \sum_{k=1}^{n} \log_2[1 + (\rho/n) \cdot (k + \gamma_{2k})]. \tag{16}$$

The constant term, that is the term that is asymptotically independent of the parameter $(1/n)$, is the term of primary interest to us. Note that $\lim_{n\to\infty} n^{-1}\gamma_{2k} = 0$ uniformly in k, since each $n^{-1}\gamma_{2k}$ has an extra divisor of $n^{1/2}$ over what the Central Limit Theorem requires. (See, for example, [32] for the Central Limit Theorem in communication theory contexts.) To find the constant we write

$$L(n)/n \approx (1/n) \cdot \sum_{k=1}^{n} \log_2[1 + (k \cdot \rho/n)] \ (n \text{ large}). \tag{17}$$

Using the trapezoidal rule we get

$$L(n)/n \to \int_0^1 \log_2[1 + x \cdot \rho]dx \ (n \text{ large}). \tag{18}$$

Consequently,

$$L(n)/n \to (1+\rho^{-1}) \cdot \log_2(1+\rho) - \log_2 e \text{ as } n \to \infty. \tag{19}$$

Noticing how (17) departs from the trapezoidal rule, we see that the integral (18) involves an error relative to (17): More precisely we write

$$L(n)/n \to (1+\rho^{-1}) \cdot \log_2(1+\rho) - \log_2 e + n^{-1} \cdot \log_2(1+\rho)^{1/2} + o(1/n) \text{ as } n \to \infty. \tag{20}$$

The next term in the expansion of $L(n)/n$ can be shown to be normally distributed with zero mean and variance vanishing as n increases. This vanishing stems from the fact that the outside n^{-1} multiplier appearing in the argument of the logarithm in (16) is more of a normalization than the Central Limit Theorem requires. Specifically, the next term in the large n expansion of (16) can be shown to be the stochastic integral with respect to standardized white Gaussian noise

$$\frac{\rho}{n \cdot \ln 2} \int_0^1 \frac{x^{1/2}}{1+\rho x} dW_x. \tag{21}$$

W_x here is the standard Wiener process which is a zero mean Gaussian process with Probability $[W_0 = 0] = 1$ and $E(W_{x_1} \cdot W_{x_2}) = \min(x_1, x_2)$. See, for example, [33] for representing an expression such as (21) using a sequence of equipartitions of the unit interval to enable us relate to the first nonconstant term in the expansion the right hand side of (16). Reference [33] is also used to conclude that the integral (21) is a zero mean Gaussian variate with variance equal to the squared L_2 norm of the integrand times $[\rho/(n \cdot \ln 2)]^2$. We can now write

$$C(n)/n > (1+\rho^{-1}) \log_2(1+\rho) - \log_2 e + \varepsilon_n + o(n^{-1}) \tag{22}$$

where the random variable ε_n has a Gaussian distribution with

$$\text{Mean } \varepsilon_n = (1/n) \cdot \log_2(1+\rho)^{-1/2} \tag{23a}$$

and

$$\text{Var } \varepsilon_n = \left(\frac{1}{n \cdot \ln 2}\right)^2 \cdot \left[\ln(1+\rho) - \frac{\rho}{1+\rho}\right]. \tag{23b}$$

We see that for large n the dominant $C(n)$ term scales at least linearly with increasing n. The slope lower bound is given by the right hand side of (19) which for ρ large is $\log_2(\rho/e)$.

4.1.2. (n_T, n_R) *Capacity Derivation: One Spatially Cycled Transmitting Antenna/Symbol*

To derive (13),[3] instead of cycling the choice of transmit antenna, we will look at the capacity of a system in which a one time-pad based on a fair die toss, one per transmitted symbol, is used to choose the transmit antenna. Since the symbol transmission rates out of each of the transmitting antennas using die tossing is the same as when cycling, the transmitter and receiver can map the symbols from the die tossing format to the cycling format (or vice versa). It is convenient to compute capacity with the equivalent die tossing format.

Channel capacity is defined in terms of mutual information between input and output, I(*input, output*). (We are not concerned with optimizing the input alphabet density since we

[3] Reference [34] (also [28]) provides a powerful abstract theory for treating time-varying channels that we could look to use to produce (13). However, justification of (13) lends itself to a relatively straight-forward argument.

are constrained to use a gaussian.) Reference [24] gives the following expression for the mutual information in the case when the output is a multiple of the transmitted signal plus independent additive noise:

$$I(input,\ output) = \varepsilon(output) - \varepsilon(noise). \tag{24}$$

In this formula $\varepsilon(.)$ represents entropy. Recall that for a random variable ξ the entropy, $\varepsilon(\xi)$, is defined as the expectation of $-\log_2 \xi$, and entropy is especially easy to compute for gaussian variates for which it turns out to be the base two logarithm of a multiple of the variance.

For $n_T = 1$, we get from (24) the classical Shannon formula $\log_2(1 + \text{SNR})$ for capacity of an AWGN channel. For (13), by the immediately following conditioning argument, we will get an average of n_T such differences. Specifically, with die tossing, for each of the two entropy terms compute the expectation by conditioning on the n_T equally likely outcomes. Then, since $\varepsilon(output)$ and $\varepsilon(noise)$ are each expressed as the sum of n_T conditional entropies we obtain

$$\begin{aligned} I(input,\ output) = &\frac{1}{n_T} \cdot \sum_{i=1}^{n_T} \varepsilon(output|i\text{-}th\ outcome) \\ &-\frac{1}{n_T} \cdot \sum_{i=1}^{n_T} \varepsilon(noise|i\text{-}th\ outcome). \end{aligned} \tag{25a}$$

Rearrange this summation by collecting the n_T pairs of identically conditioned terms to get

$$I(input,\ output) = \frac{1}{n_T} \cdot \sum_{i=1}^{n_T} [\varepsilon(output|i\text{-}th\ outcome) - \varepsilon(noise|i\text{-}th\ outcome)]. \tag{25b}$$

The summands are just n_T entropy differences each just like the difference encountered in deriving the classical Shannon formula except that the SNRs depend on i. Consequently, the rearranged sum is the average given by Equation (13).

5. Capacity ccdf Comparisons for some (n, n) Cases

For fixed ρ and $n_T = n_R = n$ we will look at some ccdfs often focusing on $P_{\text{out}} = 1\%$. We will also find the SNR and n to meet a required P_{out} requirement.

The ccdf of capacity, when H is a Rayleigh matrix can be obtained from (4). In the more ambitious examples reported in this section H is a large square matrix – up to 64 by 64. We judged the random determinant to be a complicated function whose distribution was best estimated by simply generating the complex gaussian variates by Monte-Carlo methods as opposed to seeking to analytically determine the distribution. Having generated the underlying gaussian variates for (4), it was advantageous to use these variates for estimating the ccdfs of the other capacities as well.

For very small outages, say much less than 1%, it is worth noting that it is possible to compute nearly all of the ccdfs that we will present using either closed form or standard convolution routines utilizing the fast fourier transform (FFT). Indeed, one can see from the various formulas that we presented for the Rayleigh channel, that the typical computation involves simple functions of chi-squared variates or additions of simple functions of independent chi-squared variates. For the latter, convolution methods based on the FFT are well known to be extremely useful for computing probability distributions stemming from sums

of independent variates. Refined analytical methods specifically targeting ultra-small outages should also be considered to expedite computation of extreme tail probabilities.

Most runs used 10,000 realizations of H. For $n = 2$ generating a typical ccdf only takes tens of seconds, whereas generating a ccdf for (64, 64) took several hours. Most examples involve small values of n: as we will see, $n = 2$ was, by far, the most common case. Numerical Analysis Group's (NAg®) routines [35] were used for random number generation, complex determinants, etc.

For the most part, the systems that we report on are covered by the equations already given. However, in addition, some standard computations with vectors will be required. Specifically, we will refer to receiver processing involving nulling out all signal components but one. One needs only to form a linear combination of $n - D$ vectors that is perpendicular to $n - 1$ distinct vectors where the n vectors involved have, with probability one, full rank. An optimization problem needs to be solved for the numerical work reported in Section 7: this problem and its solution are presented there.

5.1. Marked Improvement of Outage Probability

Figures 1a and 1b depict capacity ccdf tails when the number, n, of both transmitter and receiver antennas is two and four respectively. Capacity is measured in units of bits/cycle. The average received SNR is a parameter covering the range 0–21 dB in steps of 3 dB. For contrast each figure also includes the reference case of only one transmit and one receive antenna with the corresponding eight curves for this reference case drawn using thin lines. We emphasize that, regardless of the number of transmitters, the total power radiated by all transmitters is the same as for the baseline case of only one transmitter. In the two figures the comparison with the baseline case makes evident a highly significant improvement of the low P_{out} tails by using more antennas.

To illustrate how great this capacity is, even for small n, take the cases $n = 2$ and 4 at an average received SNR of 21 dB. Look at Figures 1a–b. For over 99% of the channels the capacity exceeds about 7 and 19 *bits/cycle* respectively, whereas, if $n = 1$ there is only about 1 bit/cycle at the 99% level. For this same SNR and values of n, at the 95% there is again very substantial improvement in using more antennas at transmitter and receiver, although, not as much as at the 99% level.

5.2. Reasonability of Number of Bits per Cycle per Dimension

Figures 1–b inform us that for a fixed P_{out} the capacity improvement increases markedly with the number of antennas. For $n = 4$ the capacities might at first seem too large to be reasonable, and, as already indicated, in many applications involving multi-GHz carriers there can be room for an order of magnitude or so more antennas at both transmitter and receiver. Out of context, if the symbol rate is about equal to the channel bandwidth, striving to attain a hefty fraction (say 50%) of a capacity of several hundred bits per symbol would very likely evoke concerns about astronomical constellation sizes. However, the perspective for assessing constellation size considerations is the *per dimension* perspective where we are counting one (complex) dimension for each of the n vector components of the transmitted signal. Figure 2 depicts the per dimension capacities for $n = 16$, 32 and 64, again for the parameter range of 0 to 21 dB in steps of 3 dB. For an SNR of 21 dB the 90% point (10% outage) for $n = 32$

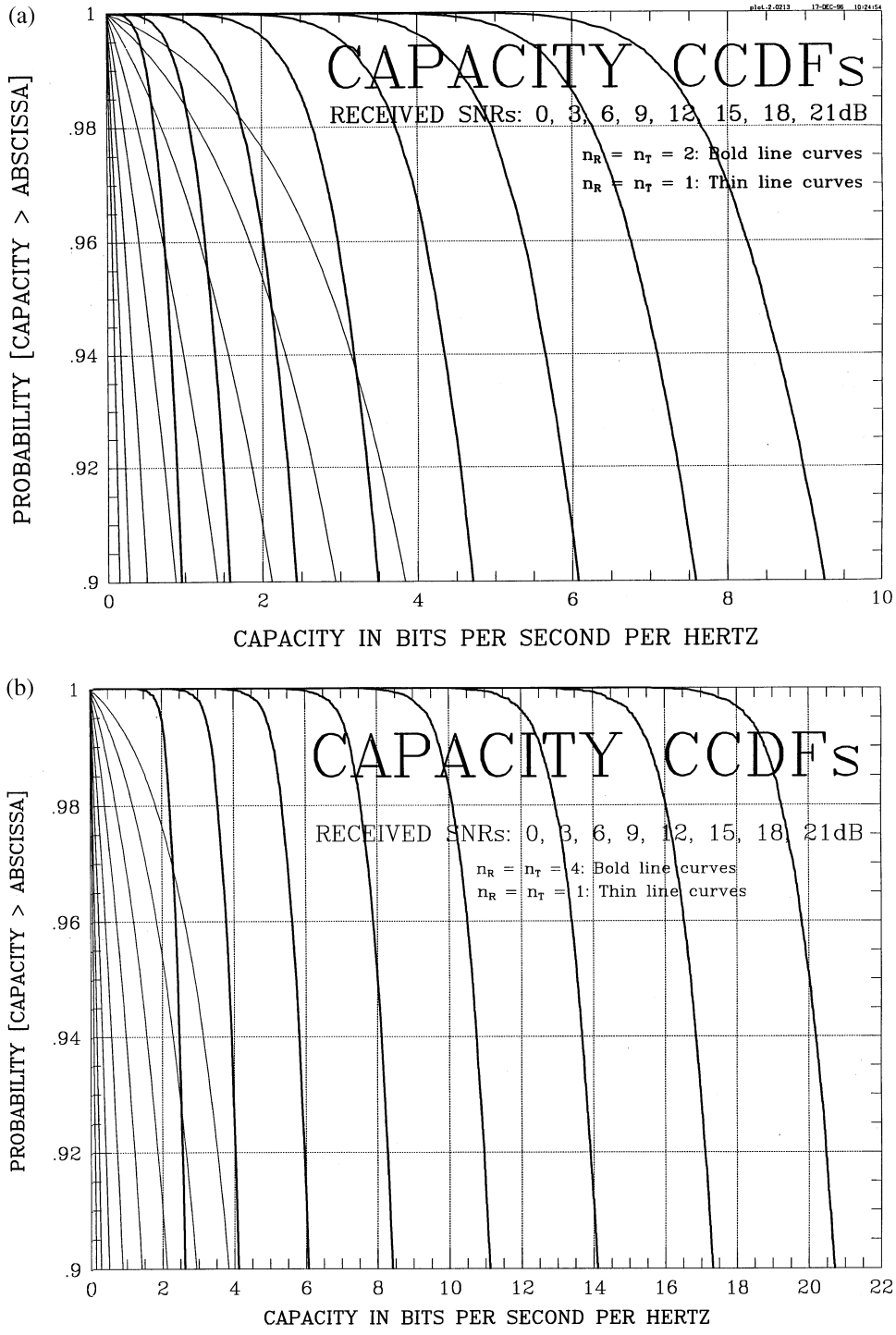


Figure 1. Capacity: Complementary Cumulative Distribution Functions. Assumes statistically independent Rayleigh faded paths. Average received SNR is a parameter ranging from 0 to 21 dB in steps of 3 dB. (a) Two antennas at both transmitter and receiver (bold line curves). Single antenna at both transmitter and receiver shown for reference (thin line curves). (b) Same as (a) except four antennas depicted by bold line curves.

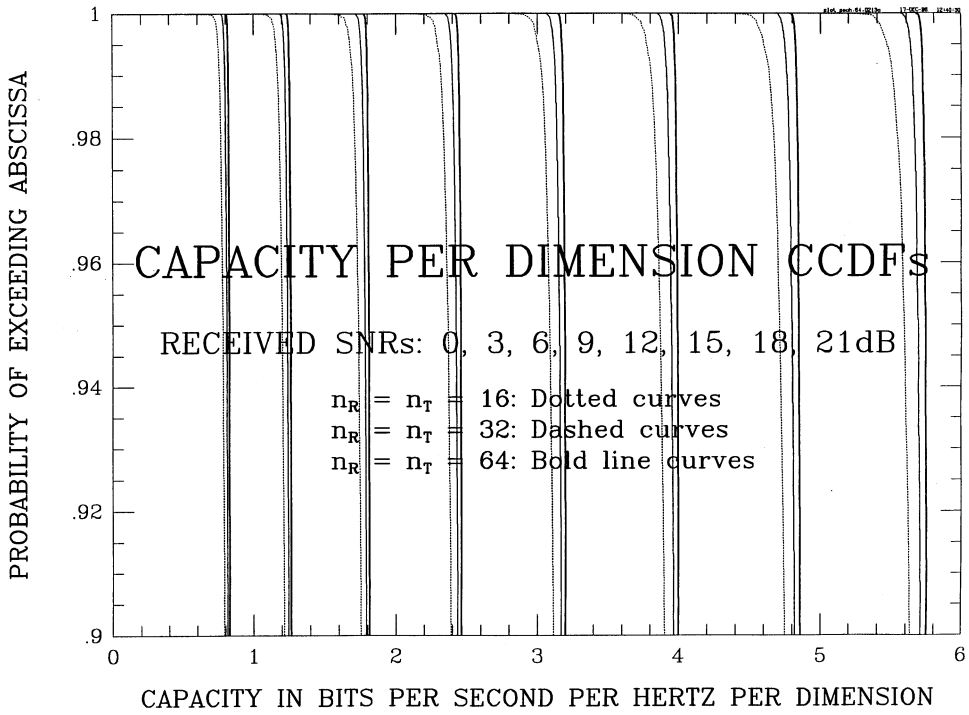


Figure 2. Capacity: Complementary Cumulative Distribution Functions. Capacity represented in terms of bits/Hertz/spatial dimension. Assumes statistically independent Rayleigh faded paths. Average received SNR is a parameter ranging from 0 to 21 dB in steps of 3 dB. Number of antennas at both transmitter and receiver: sixteen (thin line curves), thirty-two (medium line curves) and sixty-four (bold line curves).

Figure 2 shows 5.7 bits per symbol per (complex) dimension which is equivalent to 182.4 bits per symbol assuming a symbol rate equal to the channel bandwidth.

The linear capacity growth hinted at by the lower bound analysis in Section 5 is apparent in the strong tendency toward asymptotic constancy of the per dimension capacities illustrated in Figure 2. Figure 2, as well as Figures 1a–b, also serve to illustrate that considerable capacity is available even at an SNR of 0 dB. For example, look at the 99% point in Figure 2 for n of 32 and 64 we see about 0.8 bits/cycle/dimension. For $n = 16$ the per dimension capacity is very slightly less than 0.8.

6. Contrasting System Architectures

We will next explore the relative merits of various architectures including (1, n_R) and (n_T, 1) systems, but most interestingly a number of (n, n) systems. Then we will also very briefly mention multicarrier systems as well as systems designed for bulk data transfer.

6.1. The ccdfs of Some (2, 2) and (8, 8) Systems

Comparisons of the ccdfs of some architectures are shown in Figures 3a and 3b for two and eight antennas respectively. For illustrative purposes a 21 dB average received SNR is used for each figure. Look, for example, at Figure 3b (3a is similarly described). The boldest line rightmost curve is the ccdf of capacity for the (8, 8) case. The thin line curve very close to it on the left represents the lower bound approximation which is seen to be quite accurate. The thin

line curve on the extreme left is the (1, 1) case which, as we have already learned to expect, is very significantly surpassed by the (8, 8) case. Disposed just to the right of the (1, 1) curve are three nearly parallel, predominantly vertical thin line curves. The leftmost, most inferior of the three, is the (8, 1) curve. Its outage performance is seen to far surpass that of (1, 1). This curve is followed on the right by two (1, 8) curves. The middle of the three curves is selection diversity while the rightmost (best of the three) is OC(8).

At the low outage tail (ccdf near 100%) somewhat superior to OC(8) is an (8, 8) curve corresponding to using all eight transmitters by simply cycling through all eight in sequence, using only one transmitter at a time. This curve is nearly vertical at an abscissa of about 10 bits per cycle. The formula for this curve is given by Equation (13) in Section 4, a simple analysis of which reveals this vertical asymptote to be $\log_2(1 + n \cdot \rho)$. For $n = 8$ and an SNR of 21 dB this evaluates to $\log_2(1 + 8 \cdot 10^{2.1})$ which is about 10 bits per cycle which agrees with Figure 3b. A straightforward perturbation analysis shows a much more severe decrease of the variance with increasing n than for the full (n n) capacity which requires all n transmitters to be used simultaneously. This comparative variance behavior is evidenced by the two boldest curves in Figure 3b.

Two (8, 8) curves remain to be described. These are the two ccdfs that exhibit, by far, the most substantial variances in the figure. For the leftmost of the remaining (8, 8) curves the transmitted signal components are independently encoded so that there is no redundancy *between* the components of the vector. Also, the received vector is linearly processed in an eightfold manner. Specifically, the ith processing nulls out all transmitted vector signal components except that from component i ($i = 1, 2, \ldots 8$). In this way there are eight linear processors – amounting to eight linear combiners. In other words, the process involves transmission and reception of eight separate data streams.

This last transmission process is improved by simply cycling each one of the eight separately encoded data streams over the eight transmitters. The reason that cycling is superior is that, when there is no cycling, one only obtains eight times the minimum of the eight signal component capacities instead of the sum of the eight. The receiver is assumed privy to this cycling. The heavier curve (8, 8) predominantly to the right represents the capacities when we require that cycling be employed. Then there are eight subchannels each of which changes state in a periodic fashion. By symmetry, the capacities of the eight subchannels are identical to each other. With equal time in each state, a state known to the receiver, the capacity of each subchannel is the average of the capacities associated with each state. The result that one can simply average capacity in a case like this was treated in Section 4. Note at low outage levels cycling is inferior to (1, 8) optimum combining.

We also looked into more elaborate ways than just simply cycling to avoid vulnerability to the "worst received" of the signal components. We randomized the presentation of the eight constituent modulated data streams to the eight transmitters with a highly volatile time varying rotation (unitary to be precise) matrix valued process. The purpose of a volatile presentation of the orientation of the signal vector is to move quickly away from a poor conjunction of the matrix channel with the transmitted signal vector. The receiver was privy to the volatile signal presentation process, which can be thought of as a spatial one-time-pad. We will not explain this signal launching in further detail since its performance gain was negligibly different from simply cycling. In fact, if the reader looks very closely at Figures 3a–b, the cycling curve that we have described is really two nearly indistinguishable curves corresponding to these two ways we devised to avoid vulnerability to the worst subchannel. There are better ways of improving performance as we will see in Section 7.

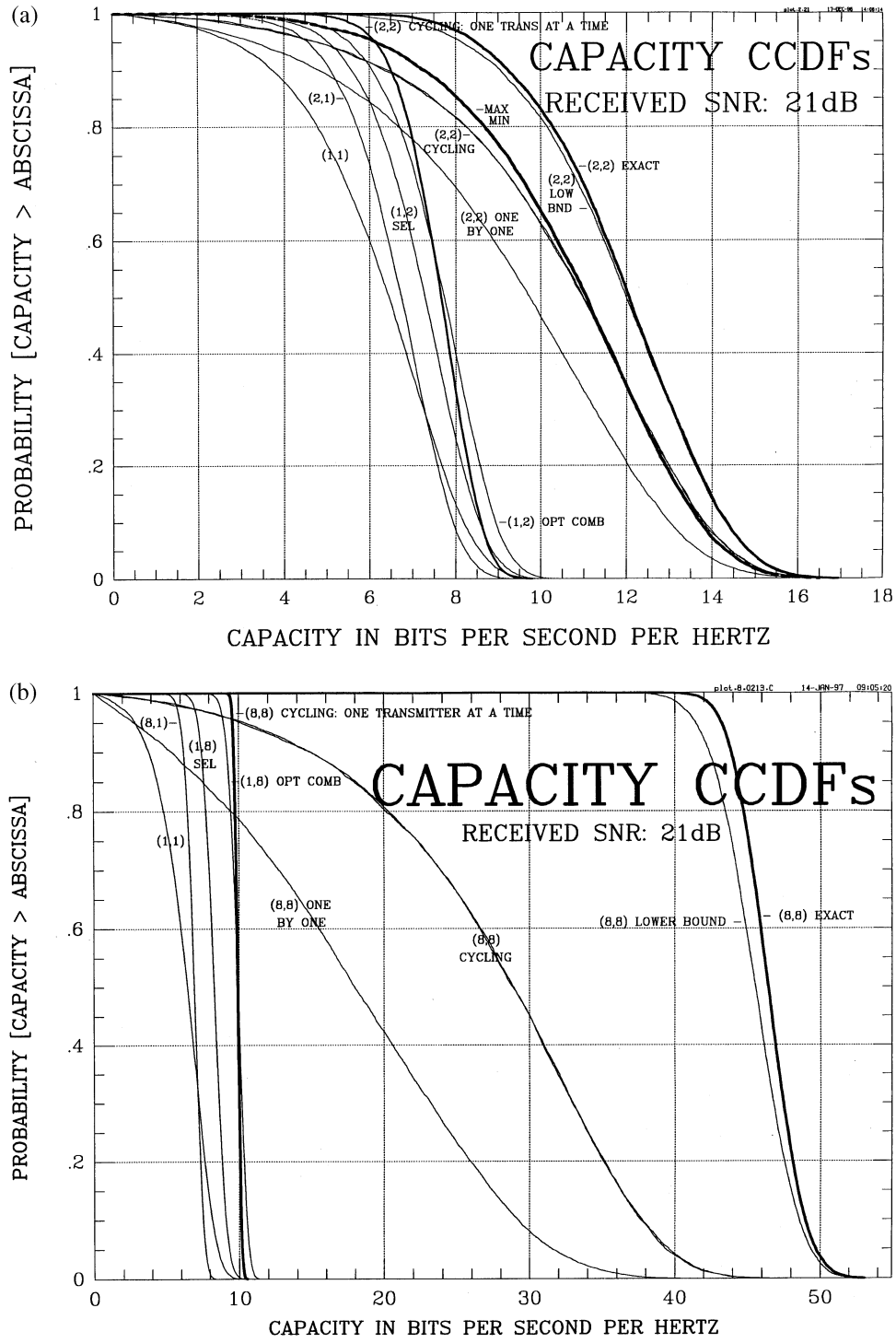


Figure 3. Contrasting Various Complementary Cumulative Distribution Functions of Capacity. Assumes statistically independent Rayleigh faded paths. Average received SNR = 21 dB. Parameter (number transmit antennas, number receive antennas). (a) (2,2) and related cases (b) (8,8) and related cases.

6.2. Multi-Carrier Communication (MCC) Architectures

Readers familiar with MCC [36] (also termed Multi-Tone Transmission) where broadened bandwidth increases signal dimensionality, may wonder about its relationship to $(n,\ n)$ systems. While it is beyond our scope here to quantitatively compare capacities with MCC we do offer some brief qualitative comments. Assume that the MCC system has n frequency bins and is a (1, n') system with $n' \leq n$ and with the same *fixed total transmitted power* as the $(n,\ n)$ system to which it is compared. The $(n,\ n)$ system is operated so (4) expresses capacity, so for large n capacity grows linearly with n.

With MCC the n frequency bins do *not* play a role corresponding to n distinct spatial modes in an $(n\ n)$ system. So, for a Rayleigh fading environment, fixing the total transmit power, capacity scales very differently with space in an $(n,\ n)$ system than with bandwidth in an $(1,\ n')$ MCC system. With MCC, as bandwidth increases, capacity, measured in bps/Hz, does not have the linear growth with n feature of an $(n,\ n)$ system. A simple extension of the argument for (5) using n adjacent bands instead of n modes shows this. Note that with Rayleigh fading, with $(1,\ n')$ MCC, if you have power transmitted in a (frequency) bin it is received in just that bin. However, with an $(n,\ n)$ system, launch power in a spatial mode and it gets received in *all n* "modes" (these are not eigenmodes since the transmitter lacks channel knowledge). Thereby, the $(n,\ n)$ system derives its advantage over MCC. True, in an $(n,\ n)$ system there is the random "destructive" cancellation of waves launched from different antennas, whereas there is no interference from neighboring bins with (ideal) MCC. However, the "destructiveness" was accounted for in our proof that for large n, $(n,\ n)$ has linear in n capacity growth (until the available surface is saturated with $\lambda/2$ spaced antenna elements).

6.3. Performance Criterion for Bulk Data Transfer

We have taken burst communications as the perspective for evaluating performance and we featured the capacity ccdf, denote here $\Gamma(x)$. We mention in passing another view that is appropriate for the scheduled communication of large files. Say that q is the fraction of time that is allotted to a user. Then if the user attempts to transmit at capacity[4] x, capacity x is achieved with probability $\Gamma(x)$ and with probability $1 - \Gamma(x)$ the transmission is completely lost (zero capacity). Then, assuming communications are acked and nacked, over the long term a rate $q \cdot \max_{x>0}[x \cdot \Gamma(x)]$ is clearly the best rate such a user can obtain. (We are assuming the user does not process ack and nacks to strive to match transmissions to the channel). For such an application a simple scheme like $(n,\ n)$ with cycling when all transmitting antennas are used simultaneously will be far improved over $(n,\ n)$ with cycling when only one antenna is used at a time. While we will not report any quantitative results here, this improvement is evident from Figures 3a–b where we observe the great probability mass that the former technique puts at high capacities where the latter method has no significant probability mass.

7. 1-D Codecs for 2-D Systems

Included in the comparisons in the previous section are architectures, that, employ, at bottom, 1-D codecs. The curves in Figure 3a–b for selection diversity and OC(n), the two curves having to do with cycling and the curve labeled one by one all correspond to such architectures. These

[4] We say "transmit at capacity x" to mean transmit at some hefty fraction of capacity x at some low bit error rate.

systems, along with others that we discuss, are structured so that a redundant coded message is prepared for a burst in a manner that does not utilize any information about the specific H outcome actually realized. The codec technology for achieving a hefty fraction of capacity for the 1-D AWGN channel is very advanced [36–39] so it is worthwhile to look to leverage this technology[5] when seeking to exploit MEAs.

The diversity schemes we looked at, that are based on adapting one (complex) dimensional architectures perform far better than (1, 1) but they are far from the $(n,\ n)$ promise. These inherently 1-D architectures can be surpassed by more complex inherently 1-D architectures. E.g., instead of nulling for $(n,\ n)$ with cycling, we could have used the eight linear receivers giving the best SNR. Then for each of the signal components the other $n-1$ components are treated as part of the noise. We can even do better as we will show for the lowest interesting value of n here, namely $n=2$.

First we discuss the maximization of the SNR with a linear operation on the received vector when the AWGN vector involves (spatially) correlated components. We will encounter instances of this maximization problem in Section 7.1 in analyzing the so-called MAX-MIN communication system. The spatial "noise" correlation occurs for these systems, because, as will be apparent, interference from other transmitted signal components than the one being detected contribute to what we term "noise". All transmitted signal components are eventually detected and the spatial correlation does not thwart our capacity analysis since ultimately the interference into each processed signal component is 1-D AWGN. The problem of computing the optimum linear operation is that of maximizing a ratio of Hermetian quadratic forms. Let ξ be the maximizing vector we seek. Use $\langle\cdot,\cdot\rangle$ for complex scalar product. The numerator form (signal power) will turn out to be the absolute value squared of a linear functional, denote $|\langle\xi,\nu\rangle|^2$ and the denominator form ("noise" power) is $\langle\xi, Q\xi\rangle$ where Q is Hermetian and positive definite. To do this optimization in a simple way note

$$\max_{\xi\neq 0}\frac{|\langle\xi,\nu\rangle|^2}{\langle\xi,Q\xi\rangle}\overset{a}{=}\max_{\xi\neq 0}\frac{|\langle\xi,\nu\rangle|^2}{\langle\xi,qq^\dagger\xi\rangle}\overset{b}{=}\max_{\xi\neq 0}\frac{|\langle\xi,\nu\rangle|^2}{\langle q^\dagger\xi,q^\dagger\xi\rangle}\overset{c}{=}\max_{\eta\neq 0}\frac{|\langle(q^\dagger)^{-1}\eta,\nu\rangle|^2}{||\eta||^2}\overset{d}{=}$$
$$\max_{\eta\neq 0}\frac{|\langle\eta,q^{-1}\nu\rangle|^2}{||\eta||^2}\overset{e}{=}\langle q^{-1}\nu,q^{-1}\nu\rangle\overset{f}{=}\langle\nu,Q^{-1}\nu\rangle. \tag{26a–f}$$

Equation (26a) is because q is defined to be the unique positive definite square root of Q. Equation (26c) is because the Cauchy-Schwartz inequality can be attained with equality when η is a scalar multiple of the other vector in the scalar product. The other three (26·) equations are for obvious reasons.

The line of reasoning represented by (26a–f) appears in a different detection problem in reference [40] and holds for any dimension. In the cases we will compute here $n=2$. Then, for detecting signal i with signal j as the interferer, the right hand side of (26f) becomes

$$\langle\nu,Q^{-1}\nu\rangle=\tfrac{1}{2}\cdot\rho\cdot[||(H_{1i},H_{2i})'||^2+\tfrac{1}{2}\cdot\rho\cdot|det\ H|^2]/[1+\tfrac{1}{2}\cdot\rho\cdot||(H_{1j},H_{2j})'||^2]. \tag{27}$$

7.1. THE MAX-MIN COMMUNICATION SYSTEM

We include in Figure 3a a ccdf for another communication system that only uses 1-D codecs. It is called the MAX-MIN system and its performance is given by the bold heavy dashed curve.

[5] When cycling is involved, the 1-D decoder is nonstandard because of the cyclic nature of the SNR.

We describe now the system giving this curve by detailing constriants that we impose to force the use of 1-D codecs and at the same time yield a setup suited to performance assessment. We constrain the encoding techniques for the two signal components to be identical and we require each signal component to be responsible for half the payload. Each encoder, (one for each signal component) is to operate without any knowledge of the information bits the other is encoding – no redundancy between the two encoded signals. Next the constraints we impose on detection. Detection is described in terms of two underlying detectors. Each of the two underlying detectors is comprised of two levels of "optimum combining" one level for detecting each of the transmitted signal components 1 and 2. The detector we shall use is defined as the better of these two detectors each of which involves two stages of optimum combining.

- DETECTOR 1: Detect signal component 1 first, using optimum combining, but treating the "interference" contributions from signal component 2 as part of the gaussian noise. Assuming 1 is detected successfully, its contribution is subtracted from the received signal vector and then signal component 2 is extracted again by optimum combining. Component 2 is impaired only by thermal noise since the effect of component 1 has been removed.
- DETECTOR 2: The description is the same as for detector 1 except 1 and 2 are interchanged.

Consider the capacity of four "channels" corresponding to that obtained when using the detector i to detect the jth signal component for the four (i, j) pairs. Therefore, C_{ij} is the capacity that a hypothetical "(i, j) channel" would have if we detected signal j in a process in which signal i is detected first. So there are four statistically dependent random capacities denoted C_{ij}. Thus, for example, C_{22} is the capacity when detecting the signal transmitted from the second transmit antenna by optimally combining the two receive antenna signals in the presence of the noise power and the other interfering signal. C_{12} is the capacity when detecting the signal transmitted from the second transmit antenna after having (a) first detected the signal transmitted from the first transmit antenna, (b) subtracted it out of the outputs of the two receiving antennas before (c) optimally combining them in the presence of noise, without interference from the previously detected signal.

The capacity of the matrix channel when constrained to using only the first detector is $2 \cdot \min[C_{11}, C_{12}]$. Note that the minimum, not the sum, gives the correct capacity because any excess capacity of one component over another is worthless. The capacity when constrained to using the second detector is $2 \cdot \min[C_{21}, C_{22}]$. The capacity of the better of these two detectors is $2 \cdot \max\{\min[C_{11}, C_{12}], \min[C_{21}, C_{22}]\}$. We note that given the outcome of the random channel, H, it is a simple straight-forward exercise to compute the four C_{ij}s to produce the bold heavy dashed curve labeled MAX-MIN in Figure 3a. Around the 91% level (but not at 95% and above) the capacity is seen to surpass the best of all the other methods of using 1-D codecs that are represented in the Figure 3a.

It is interesting to establish the very best that can be done using only 1-D codecs when communicating over an n-D channel. Our simple MAX-MIN architecture served to highlight the issue as well as to illustrate that nontrivial performance is in the offing.

8. Summary

From Equation (4) we were able to derive a unified set of closed form capacity expressions for key cases. These are Equations (5, 6, 9–11, 13). Equations (9–11, 13) apply to the important case of a stochastic quasi-static channel characterized as having independent Rayleigh faded paths between antenna elements. We derived and interpreted a capacity lower bound, inequality (12) in Section 4 and we expressed its asymptotic behavior for a large number of antennas, n, at both sites. In Section 6, in numerical work we highlighted the contrast with the baseline case of a single transmit and a single receive antenna. In analyzing the case of independent Rayleigh faded paths between antenna elements we again emphasized the (n, n) case. From the aforementioned capacity lower bound asymptotics as expressed by Equation (22) in Section 4.1.1.1 and again in the numerical work in Sections 5 and 6 for the large SNR realm we observed the scaling to be like n more bits/cycle for each 3 dB increase in SNR.

While, as is well known, receive diversity alone, especially optimal combining, offers significant capacity improvement over single antenna reception, in contrast, the simultaneous use of transmit and receive diversity greatly increases the capacity over what is possible with just receive diversity. Using the equations that we derived in Section 5, in Section 6 curves that quantified this benefit were presented. Specifically, under the constraint of fixed overall transmitted power, we computed many examples to illustrate how great the capacity can be relative to the $n = 1$ baseline (Figures 1 and 2) and also relative to other diversity schemes (Figure 3). We showed how great this capacity is, even for small n. For example, in Figures 1a–b we saw the cases $n = 2$ and 4 at an average received SNR of 21 dB (other dB values are also given). For over 99% of the channels the capacity exceeds about 7 and 19 *bits/cycle* respectively, whereas, if $n = 1$ there is only about 1 bit/cycle at the 99% level.

Note, for example, for a symbol rate equal to the channel bandwidth, 19 bits/cycle for $n = 4$ amounts to 4.75 bits/symbol/dimension. We have been careful to emphasize that, in exploiting MEA technology, while the number of bits per (vector) symbol can be unusually large, viewed in terms of bits/symbol/signal dimension the per component numbers can be quite reasonable. In Figure 2 curves for $n = 16$, 32 and 64 were presented in terms of bits/symbol/dimension. We see the strong tendency of the probability mass of the per dimension capacity to coalesce to the right of $\log_2(\rho/e)$ when ρ and n are large. This is expressed by the lower bound (Equation (22)), which, for example, for $n = 64$ and an SNR of 21 dB evaluates to about 5.6 bits/cycle/dimension. The more precise capacity calculation for these values is shown in Figure 2 and puts the great predominance of the probability mass at about 5.75 bits/cycle/dimension. For a 21 dB SNR Figure 2 also showed that even the results for $n = 16$ and 32 were close to the asymptote. Figure 2 also demonstrated that MEA technology offers considerable capacity at low outage even at an SNR of 0 dB.

In Section 6 we explored various simple architectures that used 1-D codecs in a multidimensional setup. For $n = 2$ and 8 we showed in Figure 3a and 3b respectively, how various inherently 1-D architectures fared. For the low P_{out} tail we learned that none of these were competitive with the best (n, n) performance corresponding to fully exploiting MEA technology. However, one can obtain interesting levels of performance using only 1-D codecs. In the context of very low P_{out} we found that while optimum combining performed well within the class of such architectures, a variant of optimum combining that included cycling the transmitted signal over different transmitters using only one transmitter at a time did even better. We indicated how, with somewhat more advanced architectures (e.g., in two dimensional systems the MAX-MIN) one could do still better at some outage levels.

It is evident from the great capacity improvements that we presented, that exploitation of MEAs can greatly advance indoor (and interbuilding) wireless data communications. Antenna lattices can pave the surface of transmitter and receiver spaces and two states of polarization are available. E.g., for indoor wireless LANs, we can anticipate bit rates far beyond where they are today where very little of the diversity that is possible is used. Where LAN systems may achieve 10's of Mbps in 20 MHz in the near term, in the long term, with the large n values that are possible. Gbps bit rates may be possible in such a bandwidth. It is also clear from our results that, by far, most of the great improvement in bit rate will come from processing spatially (compared to the incremental but none-the-less worthwhile advances in spectral efficiency employing existing antenna systems).

9. Future Work Items Related to Antenna Theory

We have idealized the channel model and focused on information transmission aspects, but the other important side of the coin in determining fundamental limits has to do with electromagnetic theory considerations. One wants to cram as many basic antenna elements as can be productively used into the transmit and receive spaces. While we have succeeded here in taking a first step, at this early stage the understanding of antenna theory associated with packing of array elements is not at all adequate for us to provide the last word on ultimate limits. Say the receiver (transmitter) surface is that of a box. How do the opposing faces shadow one another? What is the interactive effect of arrays on surfaces at right angles to each other? Advancing antenna theory to assess what can be achieved by packing antenna elements in the transmit and receive spaces to maximize capacity is a most important area of research. In the next subsection we discuss some practical antenna design considerations.

9.1. Some Practical Limitations of Antennas

The capacity has been shown to increase drastically with the addition of more antennas. This suggests cramming in as many antennas as space will allow. Indeed, if increased capacity is essential, using antenna spacings smaller than $\lambda/2$ to provide more antennas can provide a significant capacity improvement. For the case of linear arrays or square arrays of equally spaced antenna elements and incident rays uniformly distributed in azimuth, diversity advantage improves even for antenna spacings of $\lambda/10$, compared to fewer elements in the same overall array size. [Reference 1, Figures 5.3–5 and 5.3–6]. There are some limitations, however.

First, the elements of the field transmission matrix become strongly correlated as the spacing between antennas drops below $\lambda/4$, assuming the significant rays are widely distributed (90 degrees or more) in angle at the antenna array. When the angular spread of the significant rays is very narrow (30 degrees or less) the correlation between elements becomes strong at even larger antenna spacings than λ. (For narrow angular spreads, the correlation exceeds 0.5 for broadside antenna spacings that are less than $1.9/[\pi \cdot (angular\ spread\ in\ radians)]$, (see [1] Chapter 1). However, even for fairly high correlations (less than 0.95), significant diversity gains are still available (see Figure 5.2–10 in [1]).

Second, close spacing of antennas introduces mutual coupling between antennas. This coupling makes it difficult to match the antenna impedance for efficient energy transfer to a receiver or from a transmitter. Also the coupling causes a further increase in the correlation between antenna signals. The amount of mutual coupling is a function of the spacing between antennas, the number of antennas, and the direction of each ray relative to the array plane.

In general the mutual coupling effect becomes greater as the antenna spacing decreases, as the number of antennas increases, and as the ray direction moves from broadside to end fire (shadowing). Typically, for rays restricted to near broadside, mutual coupling is small for $\lambda/2$ spacing for any number of antennas. Because of the mutual coupling effect, there is minimal advantage to be obtained by filling a volume with antennas, as compared to only placing antennas on the bounding surface. The surface of the volume could be a case containing the electronics of the communications system, and that case will shadow some of the antennas for various ray directions, making those antennas of little use for those ray directions. Finally, decreasing the spacing between antennas may require the antennas themselves to be reduced in size beyond that appropriate for the carrier frequency, which again makes it difficult to match the antenna impedance.

Section 5.3 of Reference 1 shows the diversity performance effects of correlation and mutual coupling for linear and square arrays of half wave dipoles in the presence of rays arriving uniformly over 360 degrees in azimuth. Optimum reception in the presence of Gaussian noise is achieved as prescribed by the techniques of [41]. The significant result is that worthwhile improvement over a single antenna can be obtained even with strong correlation and mutual coupling. In cases where size is limited, as in pocket or wrist watch communicators, or where higher capacity is needed for a desktop communicator of fixed size it may be necessary to use multiple antennas with strong mutual coupling.

Appendix. Derivation: Formula for Capacity when MEAs Are Used

Assuming that the transmitted signal vector is composed of n_T statistically independent equal power components each with a gaussian distribution, the capacity expression (4) can be derived from a general basic formula[6] which appears in [25] and [43–44]. In our case the received signal is linearly related to the transmitted signal as represented by Equation (1) in the text. Then capacity takes the form

$$C = \log_2 \frac{\det A_s \cdot \det A_r}{\det A_u} \,. \tag{A1}$$

In this expression $A_s = E(ss^\dagger) = \hat{P}/n_T \cdot I_{n_T}, A_r = E(rr^\dagger) = N \cdot I_{n_R} + \hat{P}/n_T \cdot GG^\dagger$, and $A_u = E(uu^\dagger)$ where u is the $n_T + n_R$ dimensional vector $(s, r)'$. So A_u has A_s in the northwest corner and A_r in the southeast corner. The remaining two corners are transpose conjugates of each other. The northeast of these is $\hat{P}/n_T \cdot G^\dagger$. The statistical independence among all the components of the $n_T + n_R$ dimensional vector (s, ν) is what facilitated the explicit computation of A_s, A_r and A_u. This was because all matrix entries are variances and covariances of gaussians. Some tedious algebra reduces this function of three determinants into a usable form. The algebra to do so is simplified by the following identity from [31]

$$\det \begin{bmatrix} A & C \\ B & D \end{bmatrix} = \det A \cdot \det(D - CA^{-1}B) \tag{A2}$$

[6] The references have a version of (A1) for the real vector case from which the complex case follows. For real signals and channels the formula looks just like (A1) except $\log_2[.]$ is replaced by $(1/2) \cdot \log_2[.]$. The vanishing of the $\frac{1}{2}$ in the complex case is explained as follows. Let K be the covariance matrix of any $m - D$ complex gaussian vector and let κ denote the corresponding covariance of this vector expressed as a $2m - D$ real vector (the m real components, followed by the m imaginary components). As [42] points out $\det[\kappa] = (\det[\mathrm{K}])^2$. This squaring of the determinants of the covariances of the complex gaussian vectors serves to cancel the $\frac{1}{2}$ multiplier since $(1/2)\log_2(\det \mathrm{K})^2 = \log \det \mathrm{K}$.

which requires that A be nonsingular and while A and D need to be square, they do not need to be of the same size. They are not the same size in some of our examples where $A = A_s$, $D = A_r$, etc. The identity (A2) facilitates expressing det A_u as a multiple of det A_s so det A_s can be cancelled in (A1). After the cancellation, the numerator in the argument of the logarithm becomes det$[N \cdot I_{n_R} + (\hat{P}/n_T) \cdot GG^\dagger]$. The denominator becomes det$[N \cdot I_{n_R}]$ and can be moved to the numerator as det$[N^{-1} \cdot I_{n_R}]$. Recalling that the product of determinants is the determinant of the product and that $\hat{P}^{1/2} \cdot G = P^{1/2} \cdot H$ and $\rho = P/N$, the convenient formula for generalized capacity is

$$C = \log_2 \det[I_{n_R} + (\rho/n_T) \cdot HH^\dagger] \; bps/Hz \, . \tag{4}$$

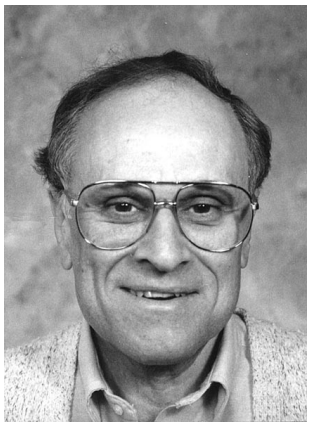
Acknowledgements

We had valuable discussions with I. Bar-David, S. Hanly, J. Salz, G. Vannucci and J.H. Winters. C.L. Mallows alerted us to reference [29] which proved very useful. A.D. Wyner first informed us of the information theory results in [34]. In discussions with P.F. Driessen the authors became aware that a truly fundamental determination of capacities for the multipath wireless environment was an important open research topic.

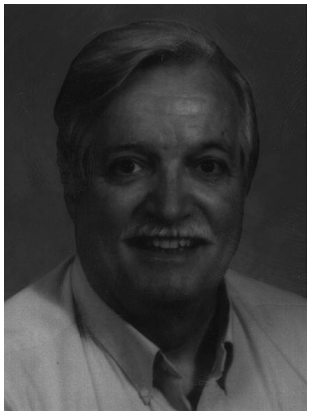
References

1. W.C. Jakes, Jr., "Microwave Mobile Communications", John Wiley and Sons, New York, Chapters 1 and 5, 1974.
2. B.J. Tuch, "An ISM Band Local Area Network: WaveLAN", IEEE Workshop on Local Area Networks, Worcester Polytechnic Institute, Worcester Massachusetts, pp. 103–111, 1991.
3. W. Diepstraten and H.J.M. Stevens, "WaveLAN Systems Test Report", NCR Corporation TR No. 407-0023871 rev A.
4. J. Kruys, "HiperLAN, Applications and Requirements", *PIMRC '92*, Boston Massachusetts, pp. 133–135, 1992.
5. M.V. Clark, L.J. Greenstein, W.K. Kennedy and M. Shafi, "Matched Filter Performance Bounds for Diversity Combining Receivers in Digital Mobile Radio", *IEEE Trans. Veh. Technol.*, Vol. 41, No. 4, pp. 356–362, 1992.
6. M.V. Clark, M. Shafi, W.K. Kennedy and L.J. Greenstein, "Optimum Linear Diversity Receivers for Mobile Communications", *IEEE Trans. Veh. Technol.*, Vol. 43, No. 1, pp. 47–56, 1994.
7. M.V. Clark, M. Shafi, W.K. Kennedy and L.J. Greenstein, "MMSE Diversity Combining for Wide-Band Digital Cellular Radio", *IEEE Trans. Commun.*, Vol. 40, No. 6, pp. 1128–1135, 1992.
8. M.V. Clark, L.J. Greenstein, W.K. Kennedy and M. Shafi, "Optimum Linear Diversity Receivers in Digital Cellular Radio", *PIMRC '92*, Conference Proceedings, IEEE Communications Society, Boston Massachusetts, pp. 174–178, 1992.
9. D.L. Noneaker and M.B. Pursley, "Error Probability Bounds for M-PSK and M-DPSK and Selective Fading Diversity Channels", *IEEE Trans. Veh. Technol.*, Vol. 43, No. 4, pp. 997–1005.
10. J.H. Winters, "On the Capacity of Radio Communications Systems with Diversity in a Rayleigh Fading Environment", *IEEE J. Select. Areas Communic.*, pp. 871–878, 1987.
11. J.H. Winters, "The Diversity Gain of Transmit Diversity in Wireless Systems with Rayleigh Fading", *ICC '94*, pp. 1121–1125, 1994.
12. J. Salz and P. Balaban, "Optimum Diversity Combining and Equalization in Digital Data Transmission with Applications to Cellular Mobile Radio – Part I: Theoretical Considerations", *IEEE Trans. Communic.*, Vol. 40, No. 5, pp. 885–894, 1992.
13. J. Salz and P. Balaban, "Optimum Diversity Combining and Equalization in Digital Data Transmission with Applications to Cellular Mobile Radio – Part II: Numerical Results", *IEEE Trans. Communic.*, Vol. 40, No. 5, pp. 895–907, 1992.
14. R.D. Gitlin, J. Salz and J.H. Winters, "The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems", *IEEE Trans. Commun.*, Vol. 42, No. 4, pp. 1740–1751, 1994.
15. J. Salz and J.H. Winters, "Effects of Fading Correlation on Adaptive Arrays in Digital Mobile Radio", *IEEE Trans. Veh. Technol.*, Vol. 43, No. 4, pp. 1049–1057, 1994.

16. P.C.F. Eggers, J. Toftgard and A. Oprea, "Antenna Systems for Base Station Diversity in Urban Small and Micro Cells", *IEEE J. Select. Areas Communc.*, Vol. 11, No. 7, pp. 1046–1047, 1993.
17. P. Dent, B. Gudmundson and M. Ewerbring, "CDMA-IC: A Novel Code Division Multiple Access Scheme Based on Interference Cancellation", *PIMRC 1992*, pp. 98–102, 1992.
18. B. Suard, A. Naguib, G. Xu and A. Paulraj, "Performance Analysis of CDMA Mobile Communication Systems Using Antenna Arrays", *Proc. ICASSP 93*, Vol. VI, Minneapolis, MN, pp. 153–156, 1993.
19. S. Talwar, A. Paulraj and M. Viberg, "Reception of Multiple Co-Channel Digital Signals Using Antenna Arrays with Applications to PCS", *Proc. ICC 94*, Vol. II, pp. 700–794, 1994.
20. J. Blanz, A. Klein, M. Nasshan and A. Steil, "Performance of a Cellular Hybrid C/TDMA Mobile Radio System Applying Joint Detection and Coherent Receiver Antenna Diversity", *IEEE J. Select. Areas Communic.*, Vol. 12, No. 4, pp. 568–574, 1994.
21. W. Honcharenko, H.L. Bertoni and J.L. Dailing, "Bilateral Averaging over Receiving and Transmitting Areas for Accurate Measurements of Sector Average Signal Strength Inside Buildings", *IEEE Trans. Antennas Propag.*, Vol. 43, No. 5, pp. 508–511, 1995.
22. J.R. Pierce and E.C. Posner, "Introduction to Communication Science and Systems", Plenum Press, New York, Chapter 11, 1980.
23. R.M. Fano, "Transmission of Information", John Wiley and Sons, New York, pp. 168–178, 1961.
24. R.G. Gallager, "Information Theory and Reliable Communcation", J. Wiley and Sons, New York, Chapter IV, 1968.
25. M.S. Pinsker, "Information and Information Stability of Random Processes", Holden Bay, San Francisco, Chapter 10, 1964.
26. G.J. Foschini and R.K. Mueller, "The Capacity of Linear Channels with Additive Gaussian Noise", *Bell Syst. Tech. J.*, pp. 81–94, 1970.
27. R.E. Blahut, "Digital Transmission of Information", Addison Wesley, New York, p. 503, 1990.
28. J. Wolfowitz, "Coding Theorems of Information Theory", Springer-Verlag, New York, 1978.
29. A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices", M.I.T. Doctoral Dissertation, Mathematics Department, 1989.
30. T.W. Anderson (ed.) and S.S. Wilks, "Collected Papers Contributions to Mathematical Statistics", John Wiley and Sons, New York, 1967.
31. P. Lancaster and M. Tismenetsky, "The Theory of Matrices", Academic Press, p. 46, 1985.
32. J.M. Wozencraft and I.M. Jacobs, "Principles of Communication Engineering", John Wiley and Sons, New York, pp. 106–111, 1965.
33. B. Hajek and E. Wong, "Stochastic Processes in Information and Dynamical Systems", McGraw Hill, New York, 1989.
34. E. Csiszar and J. Korner, "Information Theory: Coding Theorems for Discrete Memoryless Systems", Academic Press, New York, 1981.
35. Fortran Library Mark 16, "Introductory Guide", The Numerical Algorithms Group Limited, Downers Grove, Illinois, 1993.
36. R.D. Gitlin, J.F. Hayes and S. Weinstein, "Data Communication Principles", Plenum Press, New York, Chapters 5 and 7, 1992.
37. G. Ungerboeck, "Channel Coding with Multilevel/Phase Signals", *IEEE Trans. Inform. Theory*, Vol. IT-28, No. 1, pp. 55–67, 1982.
38. G.D. Forney, Jr., R.G. Gallager, G.R. Lang, F.M. Longstaff and S.U. Qureshi, "Efficient Modulation for Band-Limited Channels", Vol. SAC-2, No. 5, pp. 632–647, 1984.
39. C. Berrou, A. Glavieux and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbocodes", *ICC '93*, Conference Record, Geneva, pp. 1064–1070, 1993.
40. G.J. Foschini and A.C. Salazar, "Data Pulse Design – Intersymbol Interference Aspects", *J. Institute Mathem. Applic.*, pp. 225–237, 1975.
41. J.A. Stratton, "Electromagnetic Theory", McGraw-Hill Book Company, Inc., New York, 1941.
42. K.S. Miller, "Complex Stochastic Processes", Addison-Wesley, Massachusetts, 1974.
43. S. Kullback, "Information Theory and Statistics", John Wiley and Sons, New York, 1959.
44. D.B. Osteyee and I.J. Good, "Information Weight of Evidence, the Singularity between Probability Measures and Signal Detection", Springer-Verlag, New York, 1970.



Gerard J. Foschini is on the staff of the Wireless Communications Research Department at Bell Labs in Holmdel, New Jersey. Currently he is conducting communication and information theory investigations in wireless communications at both the point-point and systems level. Mr. Foschini has done research in many communication areas. He has taught at Princeton University, Rutgers University and Stevens Institute of Technology. He holds a B.S.E.E. from the New Jersey Institute of Technology, and M.E.E. from New York University and a Ph.D. in mathematics from Stevens.



Michael J. Gans is on the staff of the Wireless Communications Research Department at Bell Labs in Holmdel, New Jersey. He received his Ph.D. degree in electrical engineering from the University of California at Berkeley, in 1965. He has been at Bell Laboratories since 1966. His primary technical areas include mobile radio, antennas, satellites, fiber optics and infrared communications.