The Capacity of MIMO Systems with Increasing SNR by Electromagnetic Analysis

Jie Xu, Member, IEEE, Dennis L. Goeckel, Senior Member, IEEE, and Ramakrishna Janaswamy, Fellow, IEEE

Abstract—The dependence of the communication capacity of multiple-input multiple-output (MIMO) wireless systems on the average received signal-to-noise ratio (SNR), assuming the channel is unknown at the transmitter and perfectly known at the receiver, is studied through full wave electromagnetic tools. Although it is commonly accepted that the capacity of a MIMO system increases linearly at high SNRs when plotted versus the SNR expressed in dB, the fact that the number of effective degrees of freedom (DOF) of the system increases with SNR in many practical environments calls this conclusion into question for reasonably high SNRs. Based on a full wave electromagnetic investigation, we are able to analytically predict and then confirm a significant region on the MIMO capacity curve where the capacity grows quadratically when plotted versus the SNR in dB. This gives analytical insight into a portion of the capacity curve that may previously be (incorrectly) attributed to the concavity of the logarithm function rather than the increase in electromagnetic degrees of freedom. The quadratic, rather than linear, growth of capacity suggests that it may be worthwhile to invest more transmit power to achieve higher performance gains. However, to fully take advantage of this second order benefit, the numbers of antennas at the transmitter and the receiver must be close to or slightly larger than the wavevector-aperture-product (WAP) of the corresponding EM system.

Index Terms—MIMO, capacity, electromagnetic analysis, two-slope eigenvalues, power pre-saturation.

I. Introduction

ULTIPLE-input multiple-output (MIMO) wireless communication systems employ multiple antennas at both ends of links to exploit the spatial diversity of the information-carrying electromagnetic sources and fields. Provided with a suitable environment, a considerable performance gain may be achieved from the spatial domain that results in a significant increase in the spectral efficiency of communications. Most of the past research on MIMO systems has been focused on determining the maximum achievable spatial multiplexing gains compared to their traditional single-input single-output (SISO) counterparts. This problem has been investigated from both the system and the electromagnetic perspectives [1]-[7], considering either unlimited or limited communication volumes. Although many useful results have been generated concerning the spatial multiplexing gains, little attention was given to the capacity benefit available from increasing the total transmit power or, equivalently, the average received signal-to-noise ratio (SNR).

Manuscript received October 4, 2008; revised April 28, 2009; accepted June 30, 2009. The associate editor coordinating the review of this paper and approving it for publication was G. Durgin.

J. Xu, D. Goeckel, and R. Janaswamy are with the Department of Electrical and Computer Engineering, University of Massachusetts Amherst, Amherst, MA 01003 (e-mail: {jixu, goeckel, janaswamy}@ecs.umass.edu).

Digital Object Identifier 10.1109/TWC.2009.081326

This lack of interest is not surprising, and the reason is more than the fact that it is the spatial domain that is primarily exploited by the MIMO systems. The classical Shannon capacity of a traditional SISO channel increases only logarithmically with the SNR in the high SNR region. Under many different sets of assumptions commonly employed in the research community, MIMO systems provide capacity expressions that increase linearly in the number of degrees of freedom (DOF), but still only logarithmically in the SNR. For example, considering the classic works of Telatar and Foschini et al. on uncorrelated Rayleigh channel systems, one sees in Figure 3 of [1] a nearly equal spacing of the ergodic capacity lines for increments of 5 dB in SNR, and, in Figure 1 of [2], a roughly equal spacing of the outage capacity curves for every 3 dB SNR increase above 10 dB. When correlation is considered, Khan et al. used replica analysis and Grassmann variables to arrive at a closed form capacity representation that also grows linearly with antenna numbers and logarithmically with SNR, supplemented by two additional factors corresponding to the effect of correlation at both ends of the link [8]. Using a high SNR approximation, Salo et al. separated the mutual information of a general MIMO system into three parts representing a supremum value, a channel fading factor, and a channel quality dispersion penalty [9]. The last two parts are SNR-independent, while the first part increases logarithmically with SNR. All of these results lead to the commonly held notion that increases in numbers of antennas result in linear gains in capacity, while increases in the SNR result in only logarithmic gains. Consequently, in order for the capacity to increase linearly, the received SNR must be increased exponentially. This slow growth of capacity with SNR is somewhat disappointing because it is very uneconomical for the communication capacity to grow only marginally when the total transmit power is raised by several orders of magnitude.

The logarithmical increase of capacity with SNR is based on the presumption that the number of effective equivalent SISO channels is independent of the received SNR. However, a recent study by the authors on the electromagnetic degrees of freedom (EM DOF) between communicating volumes in random multiple scattering environments suggested that the maximum achievable equivalent SISO channel number will generally increase with the received SNR, especially at low and medium SNR ranges [7]. The DOF numbers of practical MIMO systems demonstrate a similar property [10]. This SNR-dependent DOF number is due to the fact that the qualities of the parallel communication channels established by the orthonormal source and field modes (singular functions of the propagation operator) usually differ from channel to

channel. The ideal case when the equivalent SISO channels are of the same quality is believed to occur only in extremely limited situations. Since the final communication capacity of a MIMO system is the combined result of both the number of effective equivalent SISO channels and the average received SNR, a capacity-SNR dependence that is faster than a logarithmical law can be expected. The study of EM DOF in [7] considered a single cluster of scattering objects. Here, many more types of systems are investigated, and the relationship between the MIMO capacity and the received SNR is studied through a rigorous electromagnetic approach following the theory in [7], which includes the complete effects of multiple scattering of waves in the environments. Two-dimensional (2-D) systems are considered, and they are believed to be a good approximation when the communications take place primarily in a (e.g., azimuthal) plane.

The rest of the article is organized as follows. Section II briefly presents some relevant findings from the EM DOF study in [7] as the starting point of this paper. The relation between EM DOF and MIMO communications is introduced in Section III, and a lower bound expression for the MIMO capacity is derived. In Section IV, a number of MIMO systems are studied rigorously by electromagnetic simulations, and the increase of the ergodic MIMO capacity as a function of the received SNR suggested by the lower bound is evidenced numerically. A capacity upper bound due to Jensen's inequality is also considered. Finally, Section V concludes the presented work.

Specifically, this study comes to the following conclusions. Firstly, a two-slope property is observed for the eigenvalues of the propagation operators in various types of systems, and the transition point between the two slopes is predicted by a system's wavevector-aperture-product (WAP). Secondly, the eigenvalues of a MIMO channel matrix are completely characterized by those of the corresponding propagation operator when the numbers of antennas are slightly larger than the WAP. Thirdly and most importantly, a quadratic lower bound can be reached for the capacity of a MIMO system over a wide range of practical received SNR, and this second order increase is verified by full wave numerical electromagnetic simulations. Lastly, the ergodic capacity upper bound due to Jensen's inequality is observed to be tight for all the systems studied.

II. SINGULAR VALUES OF THE PROPAGATION OPERATOR

With the potential of being applied to MIMO communications, the number of electromagnetic degrees of freedom between two volumes can be thought of as the maximum number of equivalent parallel communication channels under certain operation conditions. The study in [7] considered the singular value decomposition (SVD) of the propagation operator $\mathcal G$ (Green's function integral representation of the EM fields) and defined the number of EM DOF as the maximum number of receive modes (left singular functions) that have power stronger than the system noise level when the total transmit power is fixed. In this definition, the important parameters are the singular values $\{|g_n|\}$ of the propagation operator, which represent the amplitude gains of the parallel

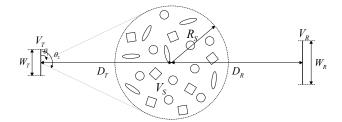


Fig. 1. Two dimensional system with linear transmit and receive volumes and a circular scattering region.

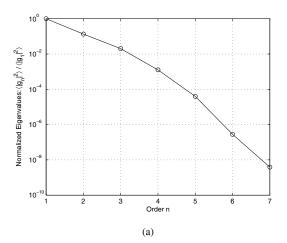
channels. Equivalently, the eigenvalues $\{|g_n|^2\}$ of the self-adjoint operator $\mathcal{G}^{\dagger}\mathcal{G}$, where \mathcal{G}^{\dagger} is the adjoint of \mathcal{G} , may be used, and they correspond to the power gains accordingly. The singular values or the eigenvalues are completely determined by the communicating volumes and the surrounding environments. For convenience, the eigenvalues are primarily considered in the following discussions.

One of the conclusions of [7] is that, instead of possessing an ideal step-like behavior, the ordered eigenvalues will generally decay exponentially, especially when the sizes of the communicating volumes are not electrically large. Accordingly, the EM DOF number exhibits a logarithmical dependence on the transmit power. However, further study can show that the decrease of the eigenvalues can actually be divided into two stages, one corresponding to the first few eigenvalues and the other corresponding to the remaining ones, respectively. Both stages exhibit exponential decay properties, but the second stage is at a much faster rate than the first one. As will be illustrated below, when plotted on a logarithmic vertical scale against the order index, the two stages of the eigenvalues form two linear sections of different slopes, which will be called the two-slope property. The slope of the first section will be called the primary slope, and that of the second section will be termed the secondary slope. For systems with electrically small volumes, the distinction between the two slopes is usually observable but not remarkable. However, when the communicating volumes are large enough (e.g., 5λ or larger), a weak knee in the eigenvalue curve can be clearly identified. The term weak knee is used to emphasize its difference from the sharp knee when the eigenvalues are assumed roughly constant up to some order and zero thereafter, for which the primary slope is roughly zero, suggesting constant instead of exponentially decaying eigenvalues.

As an example, two 2-D systems configured as shown in Figure 1 are considered in the following. In the figure V_T , V_R , and V_S are the linear transmit volume (of length W_T), receive volume (of length W_R), and the circular scattering region, respectively. The detailed configurations are summarized in Table I and labeled as System I and System II. The scattering region contains randomly distributed perfectly electric conducting (PEC) cylinders of perimeter $6\pi\lambda$, but of various shapes (circle, square, and ellipse of eccentricity 0.968), where λ is the wavelength of operation. It is assumed that ideally the scattering cylinders are distributed in a uniformly random way, with the only requirement being that they do not overlap or contact with each other. To simplify the cylinder deployment in simulations, a slightly different approach is adopted. First the scattering region is segmented into a group of hexagons of

TABLE I SYSTEM CONFIGURATIONS FOR FIGURE 1. N_c is the number of Cylinders, and WAP is the wavenumber-aperture-product as defined in (1).

System	D_T	D_R	W_T	W_R	R_S	N_c	WAP
I [Fig. 2(a)]	60λ	80λ	2λ	2λ	50λ	20	2.7
II [Fig. 2(b)]	95λ	110λ	10λ	10λ	80λ	54	16.0



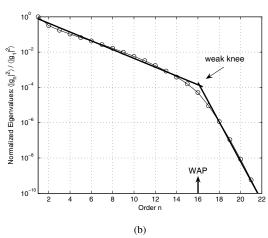


Fig. 2. Normalized eigenvalues of propagation operator. (a): System I in Table I; (b): System II in Table I.

side length $10/\sqrt{3}\lambda$. Each hexagon may or may not contain a cylinder. However if a hexagon does contain a cylinder, the cylinder can take a uniformly random position and orientation therein, but its whole body must be within the boundary of the hexagon. Cylinders of different shapes are of roughly the same number. The underlying EM scattering problems are solved by the integral equation technique using the Nyström method and Generalized Conjugate Residual (GCR) iterative solver. The single layer fast multipole method (FMM) is used to accelerate the matrix-vector multiplication in the iteration. The simulation algorithm is implemented in FORTRAN. Detailed mathematical groundwork and algorithm can be found in [7] and the references therein.

Figure 2 shows the normalized average eigenvalues of the operator $\mathcal{G}^{\dagger}\mathcal{G}$ for the two systems, whose most significant distinction is that the communicating volumes in System II are much larger than those in System I. It can be seen in Figure 2(a) that the decay rate becomes faster as the order index of the eigenvalues increases. However, Figure 2(b) illustrates the

transition from the primary slope to the secondary slope very clearly. More interestingly, it is noted that the position of the weak knee, as characterized by the intersection of two linear fittings (bold straight lines on logarithmic scale) in Figure 2(b), is very close to the smaller wavenumber-aperture-product (WAP) between the transmit and the receive sides [6]. For the system depicted in Figure 1, the WAP on the transmit side is determined as the normalized projection length (projection area for 3-D cases) of the transmit volume

$$WAP_T = \int_{\theta_1}^{\theta_2} \frac{W_T}{\lambda} \sin \theta d\theta, \tag{1}$$

which can be reduced to $WAP_T \approx \frac{W_T}{\lambda}\Omega_T = \frac{W_T}{\lambda}(\theta_2 - \theta_1)$ if the scattering region is on the broadside of the transmit volume and Ω_T is not too large. The WAP on the receive side is computed in a similar way. It was concluded in [7] that the smaller WAP may not be sufficient to characterize the EM DOF number of a system with exponentially decaying eigenvalues. Nevertheless the result in Figure 2(b) suggests that it may be used to identify the position of the weak knee for systems where the sizes of the communicating volumes are not electrically small. This notion will be further verified in Section IV. For the convenience of the following discussion, whenever the term WAP is mentioned alone, it refers to the smaller WAP between the transmit and receive sides.

The exponential decay of the eigenvalues of the propagation operator observed in Figure 2 is not a coincidence. This property has been frequently verified by measurement and believed to be true for general systems and environments. The measured mean eigenvalues of a MIMO system consisting of uniform linear arrays are reported in [11], and the temporal variation of the eigenvalues of other MIMO systems are presented in [12], [13]; in each case, the exponential decay can be clearly identified. However, the two-slope property described above is not observed in these studies, which is primarily because the antenna arrays in all of the measurement campaigns occupy only small spatial volumes. Even though these reported results are for MIMO channel matrices instead of propagation operators, it will be shown in the next two sections that the two sets of eigenvalues are closely related.

III. MIMO CHANNEL EIGENVALUE MODEL AND CAPACITY LOWER BOUND

The study of EM DOF in [7] did not involve any specific antennas. The communication was considered between two disjoint volumes, which in some sense can be thought of as two continuous arrays of antennas. Due to this antennaindependence, the study of EM DOF is expected to result in bounding limits for the optimal performance of any MIMO system involving actual antennas in the same volumes. It is further expected and will be verified by simulations later in this study that, when the antennas are placed densely enough, the singular values of the MIMO channel matrix can be fully described by those of the propagation operator, which corresponds to the antenna saturation effect [14]. Therefore the property of the singular values of the propagation operator can be used to estimate the capacity behavior of such a MIMO system.

Suppose a MIMO system consists of N_T transmit antennas and N_R receive antennas that occupy two volumes of considerable sizes, The antenna numbers are assumed to be slightly above saturation to fully take advantage of the performance gain from the spatial domain, i.e., $N \stackrel{\triangle}{=} \min\{N_T, N_R\} \gtrsim N_{\rm knee}$, where $N_{\rm knee}$ indicates the eigenvalue weak knee of the propagation operator between the volumes. Also suppose the squares of the ordered singular values $\{\sigma_n^2\}$ (referred to as eigenvalues hereafter) of the complex channel matrix $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$ exhibit a two-slope exponential decay property, which is dictated by the propagation operator and further modeled as

$$\sigma_n^2 = \begin{cases} \gamma 2^{-\frac{n-1}{\alpha}} & \text{if } n \le N_{\text{knee}} \\ \zeta 2^{-\frac{n-N_{\text{knee}}}{\beta}} & \text{if } n > N_{\text{knee}} \end{cases}, \tag{2}$$

where $\gamma \stackrel{\triangle}{=} \sigma_1^2$ and $\zeta \stackrel{\triangle}{=} \sigma_{N_{\rm knee}}^2 = \gamma 2^{-\frac{N_{\rm knee}-1}{\alpha}}$ are the first eigenvalue and that at the weak knee position $N_{\rm knee}$, respectively. Clearly the two stages of decay require that $\alpha \gg \beta > 0$. It is also assumed that $N_{\rm knee}$ (integer) can be approximated by the smaller WAP (real) between the transmit side and the receive side.

The equal power transmission strategy is often used when the channel state information (CSI) is available at the receiver side only, and it has been shown in [15] that this strategy will maximize the minimum (minimax) average mutual information of a MIMO system with arbitrary correlated fading channels under such an operation condition. The corresponding communication capacity of the system is

$$C(\rho) = \log_2 \left| \mathbf{I} + \frac{\rho}{N_T} \mathbf{H} \mathbf{H}^{\dagger} \right| = \sum_{n=1}^{N} \log_2 \left(1 + \frac{\rho}{N_T} \sigma_n^2 \right), \quad (3)$$

where I is an $N_R \times N_R$ identity matrix, ρ is the average SNR per receive antenna, the symbol † denotes the Hermitian of complex matrices, and $|\cdot|$ represents the determinant of a square matrix. In order to analyze the behavior of capacity versus the received SNR, the *non-low* SNR range $(\frac{\rho\gamma}{N_T}>1)$ is first divided into two sub-ranges according to the following definition.

Definition 3.1: The received SNR is directly determined by the total transmit power provided. When the transmit power is such that none of the equivalent SISO channels corresponding to the secondary slope eigenvalues is effective, the system is said to be operating under the *power pre-saturation* condition; otherwise the system is under the *power post-saturation* condition. Specifically, the power pre-saturation region is defined mathematically as

$$\frac{\rho}{N_T} \sigma_{N_{\text{knee}}}^2 \le 1. \tag{4}$$

A lower bound for the MIMO capacity when the system is under power pre-saturation condition can be obtained in a straightforward way.

Theorem 3.2: When the eigenvalues of the MIMO channel exhibit the two-slope exponential decay property in (2) and the received SNR is in the power pre-saturation sub-range $\rho \leq \frac{N_T}{\sigma_{N_{\rm knee}}^2}$, a lower bound for the capacity can be found as

$$C(\rho) > \frac{1}{2} \log_2 \left(\frac{\rho \gamma}{N_T} \right) \cdot \left[\alpha \log_2 \left(\frac{\rho \gamma}{N_T} \right) - 1 \right]. \tag{5}$$

Proof: Define an integer-valued function $M(\rho)=\lceil\alpha\log_2\frac{\rho\gamma}{N_T}\rceil$, where $\lceil\cdot\rceil$ is the ceiling operator. Alternatively, $M(\rho)$ is such that $\frac{\rho}{N_T}\sigma_M^2>1$ and $\frac{\rho}{N_T}\sigma_{M+1}^2\leq 1$. The capacity of the system is lower bounded, using the inequality $\log(1+x)>\log(x)>0$ for x>1, by

$$\begin{split} C(\rho) &> \sum_{n=1}^{M(\rho)} \log_2 \left(\frac{\rho}{N_T} \gamma 2^{-\frac{n-1}{\alpha}} \right) \\ &= M(\rho) \log_2 \left(\frac{\rho \gamma}{N_T} \right) - \frac{\left[M(\rho) - 1 \right] M(\rho)}{2\alpha} \\ &> \frac{1}{2} \log_2 \left(\frac{\rho \gamma}{N_T} \right) \cdot \left[\alpha \log_2 \left(\frac{\rho \gamma}{N_T} \right) - 1 \right], \end{split}$$

where the last inequality is due to the relation $\alpha \log_2 \frac{\rho \gamma}{N_T} \leq M(\rho) < 1 + \alpha \log_2 \frac{\rho \gamma}{N_T}$.
Clearly, Theorem 3.2 suggests that under the power pre-

saturation condition, the capacity as a function of the received SNR of a MIMO system will grow as the product of two logarithmically increasing functions, which will be termed square-logarithmically in the following. Alternatively, the capacity grows quadratically with SNR in dB, rather than linearly as generally accepted. It is easy to prove that the capacity-SNR relationship in the power post-saturation range will be dominated by a logarithmical dependence of the form $C(\rho) \sim N_{\rm knee} \log_2 \left(\frac{\rho \gamma}{N_T}\right) - \frac{N_{\rm knee}^2 - N_{\rm knee}}{2\alpha}$ because $\alpha \gg \beta$ and $N = \min\{N_T, N_R\} \approx N_{\rm knee}$. The squarelogarithmical increase of capacity with SNR in the power pre-saturation range achieves an increase rate faster than logarithmical dependence simply because the number of effective DOF (equivalent SISO channels) of the MIMO channel also increases with SNR, which is suggested to be true for most general systems. This result highlights the fact that not all the equivalent SISO channels might be effective in the power presaturation range, and it is beneficial to increase the received SNR as much as possible to achieve a second order increase rate in capacity. Furthermore, it is worth pointing out that although the capacity lower bound in Theorem 3.2 assumes a power pre-saturation condition, it actually covers a very wide range of transmit power. As can be seen from Figure 2(b) and more detailed studies that follow next, the primary slope of the eigenvalues will usually span quite a few orders of magnitude. The resulting power pre-saturation sub-range is believed to be wide enough to include all the practically affordable transmit power values.

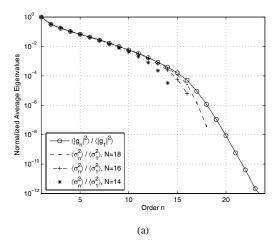
IV. NUMERICAL SIMULATIONS

The dependence of MIMO capacity on the received SNR is investigated numerically in this section. In particular, the ergodic capacities of a series of MIMO systems are calculated by full wave electromagnetic techniques described in Section II through Monte Carlo simulations (300 realizations each). However, the focus will be on the power pre-saturation range, considering that most practical MIMO systems are unlikely to be able to afford a post-saturation transmit power. The square-logarithmical capacity increase predicted by Theorem 3.2 as well as all the assumptions and statements made to reach Theorem 3.2 will be verified by the results. The implication for some aspects of the MIMO design will also be discussed.

Systems similar to Figure 1 are considered first in the numerical study. But instead of using continuous transmit and receive volumes, discrete antennas are placed across them. Since both the transmit volume and the receive volume are linear regions, when the antenna elements are equally spaced, uniform linear arrays (ULA) are employed at the transmitter and receiver. The transmit antennas are assumed to be infinitesimal line sources (infinitely long in the direction perpendicular to the 2-D plane), and the receive antennas are assumed to be infinitesimal sensors whose outputs are proportional to the electric fields at individual points of the antenna locations. If the complex amplitudes of the source currents and electric fields at the antennas are considered to be the transmit and receive signals, respectively, the MIMO channel matrix can be completely described by the Green's function of the EM system in the presence of scattering. In addition, the channel matrix H is normalized such that its mean squared Frobenius norm is the number of elements in the matrix, i.e., $\mathcal{E}\{\|\mathbf{H}\|_F^2\} = N_T N_R$, where $\mathcal{E}\{\cdot\}$ represents the statistical expectation.

One statement made in the previous section is that when the antennas are dense enough, the singular values of the MIMO channel matrix are expected to be fully characterized by those of the propagation operator. Accordingly, in deriving Theorem 3.2, the transmit and receive antenna numbers were chosen such that the minimum of the two is slightly larger than the smaller WAP, which is considered to approximately indicate the position of the eigenvalue weak knee. Figure 3(a) compares the normalized average eigenvalues of the propagation operator for System II in Table I and those of the corresponding MIMO channel matrices when 18, 16, or 14 antennas are uniformly placed in each of the two communicating volumes. The separation between adjacent antennas is 0.588λ , 0.667λ , or 0.769λ , respectively. Notice that the WAP of the EM system is roughly 16. It is observed from Figure 3(a) that even though the MIMO system only involves the source/field details at a small number of discrete points, most of its significant normalized eigenvalues (primary slope) are practically indistinguishable from those obtained from continuous volumes, provided that the antenna number is slightly larger than the WAP (N = 18 case). In contrast, obvious differences are introduced in the weaker eigenvalues if the antenna number is smaller than the WAP. Therefore it is reasonable to use (2) to model the eigenvalues of an antenna saturated MIMO channel matrix when the propagation operator is known to have twoslope exponentially decaying eigenvalues. Although average eigenvalues are used to illustrate this point in Figure 3(a), it is applicable for each realization, which can be seen from Figure 3(b). The eigenvalues of different realizations form a narrow band around the average values, and the corresponding γ and α in (2) are random variables. Determining the distributions of the two parameters is beyond the scope of the current study. However, a simple test was done by analyzing the parameters obtained from simulations; the results suggest that γ and α are generally correlated, which means that a joint probability distribution density (pdf) of the two is needed to fully characterize the statistical behavior of the eigenvalues.

The ergodic capacity of the above MIMO system with 18 transmit and 18 receive antennas is plotted against the



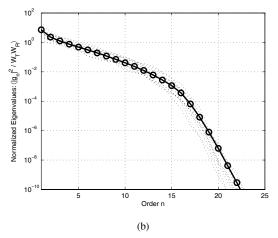
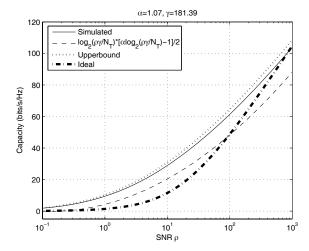


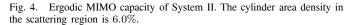
Fig. 3. Normalized eigenvalues for System II. The cylinder area density in the scattering region is 6.0%. (a): Comparison between EM and MIMO systems; (b): Comparison between average and realization samples (dotted lines).

received SNR on a logarithmic horizontal scale in Figure 4. The simulated result is the average of the capacities from 300 realizations of random scatterer distributions. It is first compared with an analytical curve (dashed line) following (5), where γ is taken to be the first average eigenvalue and α is determined by a linear fitting of the primary slope of the average eigenvalues on a logarithmic vertical scale. It must be noted that a curve generated in such a way is not the lower bound of the ergodic capacity of the MIMO system because the ergodic capacity is not obtained from the average eigenvalues of the channel matrix. Therefore in principle this curve could go above that of the actual ergodic capacity. Even so, it is still observed that the actual ergodic capacity is lower bounded by the square-logarithmical rate in the power presaturation range. An upper bound of the ergodic capacity (dotted line) is also shown in Figure 4. This upper bound is based on the concavity of the logarithmic functions and the Jensen's inequality [16]. Specifically,

$$\mathcal{E}\{C(\rho)\} = \sum_{n=1}^{N} \mathcal{E}\left\{\log_{2}\left(1 + \frac{\rho}{N_{T}}\sigma_{n}^{2}\right)\right\}$$

$$\leq \sum_{n=1}^{N} \log_{2}\left(1 + \frac{\rho}{N_{T}}\mathcal{E}\{\sigma_{n}^{2}\}\right). \tag{6}$$





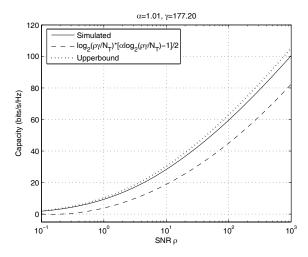


Fig. 5. Ergodic MIMO capacity of System II. The cylinder area density in the scattering region is 9.8%.

It is evidenced that this upper bound is actually very tight for the system considered. Furthermore, since the upper bound is in terms of the average eigenvalues, which also exhibit a two-slope exponential decay property, its curve must be above the square-logarithmical curve just described (dashed line in Figure 4).

The capacity-SNR curves in Figure 4 demonstrates a quadratic shape. Similar behaviors have been observed in results from various measurement campaigns such as [17], [18]. Without taking into account the fact that the effective DOF number of a MIMO system usually increases with the received SNR, the curvature of these capacity-SNR lines can be easily interpreted as being due to the concavity of the logarithmic function when the horizontal axis is on a logarithmic scale. To appreciate this point, one may plot the capacity versus the received SNR for an ideal MIMO system, where $\mathbf{H}\mathbf{H}^{\dagger}$ is of full rank and has identical unity eigenvalues $[C(\rho) = N \log_2(1 + \rho/N), N = N_T = N_R = 18]$. This result is also shown in Figure 4 with a thick dash-dotted line. It can be easily observed that the capacity for this system increases at some high order rate before flatting out to a first order linear law from roughly $\rho = 100$. The initial high order increase is the result of the concavity of the logarithmic function, a completely different reason from that of solid line in Figure 4, which is caused by Theorem 3.2. Therefore, for a general MIMO system, the characteristics of its DOF may provide a more appropriate explanation for its capacity-SNR behavior, and Theorem 3.2 gives very useful analytical insight into the power pre-saturate portion of the capacity curve.

System II has a scattering region containing 54 cylinders, which corresponds to a scatterer area density of 6.0%. In order to study the effect of the area density, the number of cylinders is increased to 90, and the simulations are repeated at a density of 9.8%. The normalized average eigenvalues of this system behave very similarly to those in Figure 3(a), which suggests that the details about the scattering medium may not be very crucial in determining the multipath richness, and hence the relative relationship among the equivalent SISO channels, provided the area density is already large enough.

As a matter of fact, the effect of the scatterer area density is embodied in the overall path loss. Due to the normalization of the channel matrix, this path loss is integrated into the received SNR ρ . The ergodic MIMO capacity for this higher density system is shown in Figure 5, which demonstrates a behavior very similar to that in Figure 4. But it is to be recognized that more transmit power is required in order to reach the same received SNR in this higher density case because of the larger path loss introduced by the more severe wave blockage from the obstacles. Small loss in the ergodic capacity is observed when the area density of the scattering region is increased, which may be due to the details of the medium.

Some implications for MIMO design can be obtained by a more careful observation of Figure 3(a). Firstly, WAP is seen to be an important parameter in that it estimates the number of the primary slope eigenvalues and gives a guideline on how many antennas to be employed. It may not be worthwhile to exploit the secondary slope eigenvalues due to the diminishingly small return. Therefore, keeping the antenna number below or slightly over the WAP will probably result in a more reasonable cost-benefit rationale. Reducing the number of antennas will not introduce severe performance degradation as long as only the first few significant equivalent SISO channels are intended for use in the first place. Secondly, it is observed that a uniform linear array structure, the most dominating structure adopted in practice, is well suited for linear communication volumes. The element separation in a slightly antenna saturated system is primarily determined by the span of scattering angles, irrespective of the volume size. It is easy to deduce from (1) that the wider the scattering angle is, the smaller the element separation could be. A rule-ofthumb generally accepted in rich multipath communications is that fields at points separated roughly half a wavelength apart tend to be uncorrelated. This separation can be theoretically reached when the scattering takes place in the half angular domain on one side of a linear array and the antenna number is taken to be the corresponding WAP. It is also expected that a ULA is a very appropriate array geometry when wave scattering is concentrated in an angular region less than π

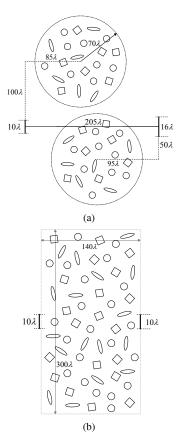


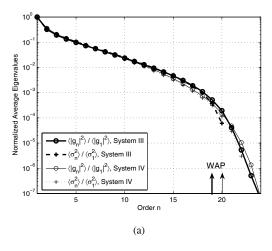
Fig. 6. System diagrams. The cylinder area density in the scattering region is 6.0%. (a): System III, a system with two circular scattering clusters, WAP=18.8; (b): System IV, a system with rectangular scattering slab, WAP=20.0.

because as shown in the studies of [6], [7], [19], the depth of the communicating volumes has only limited influence on the eigenvalue distribution of the propagation operator.

A third system (System III) as depicted in Figure 6(a) is also studied. In this system there are two disjoint circular scattering regions of identical size. Each of them contains 45 cylinders, and the same hexagon segmentation technique as before is used for their deployment. The length of the transmit volume is still 10λ ; but the receive volume is increased to 16λ long. The WAPs on the transmit and receive sides are calculated from the integral in (1) due to the wide scattering angle ranges, and the resulting values are about 18.8 and 30.1, respectively. The numbers of the transmit and receive antennas of the MIMO system are chosen to be 20 and 32, respectively ($\lambda/2$ element separation). The simulation results for this system are plotted in Figure 7 with thicker lines. It is seen from Figure 7(a) that the normalized average eigenvalues of the MIMO channels still possess a two-slope exponential decay property. However, the decay rate for the first stage is significantly reduced compared to the previous two cases. The capacity of the system is plotted in Figure 7(b) up to an SNR of 20 dB, which is the estimated upper limit of the power pre-saturation range. It can still be concluded that the ergodic capacity of this system grows faster than a squarelogarithmical rate and that the upper bound due to Jensen's inequality, which is immediately above the simulated capacity curve, is very tight. Special attention should be paid when comparing the capacity performance of the systems discussed so far. It is observed that an average received SNR of 20 dB is required for the ergodic capacity of System III to reach about 100 bps/Hz, while an SNR of 30 dB is needed in Figure 4 for System II. Clearly this 10 dB gain results from the increase in multipath richness when two scattering regions are used.

In a final scattering system (System IV), a very long rectangular slab of scattering medium is placed between the linear transmit volume and receive volume whose lengths are both 10λ . The two volumes are separated by 140λ , and the slab medium has a dimension of $140\lambda \times 300\lambda$, containing 120 cylinders [Figure 6(b)]. For such a geometry, the scattering angles on both the transmit side and the receive side are π in radian, and the WAP is calculated to be 20. The thinner lines in Figure 7 show the normalized average eigenvalues and the ergodic capacity of a MIMO system with 22 antennas $(0.45\lambda$ element separation) placed in either one of the transmit and receive volumes. Same conclusions about the MIMO capacity can be reached as before. The results for this system instead of the one where the communicating volumes are embedded in the middle of a large scattering medium is shown here because the WAP value as calculated from (1) and the simulated weak knee position disagree with each other in the latter case. The WAP for an infinite embedded system, where scatterers are present to an infinite distance in both the forward and backward half-spaces as seen from the locations of transmit and receive volumes, is 40. However the simulated eigenvalue weak knee position for a mediumtruncated (necessary for simulations) system is found to be greater than 20 but significantly less than 40. A few truncation sizes have been tried. The weak knee position moves to the right as the truncation size increases. But its rate of change is very slow, and the available simulation results are not conclusive enough to determine the specific relation between the truncation and the eigenvalue weak knee position. A possible explanation for this disagreement is that the strong forward scattering property of the electrically large cylinders considered in the simulations makes the wave interactions in the area between the antenna arrays far more important than those in the rest of the scattering medium. The overall back scattering is further reduced by the medium truncation because fewer reflectors will be present. The WAP expression of (1) does not reflect this physical characteristic, and thus overestimates the true multipath richness.

A comparison of the results for the two systems in Figure 6 also reveals that System III performs slightly better than System IV in terms of both EM DOF and MIMO capacity, even though its WAP value is smaller. To understand this seemingly contradiction, one should recall that the (WAP_T,WAP_R) pairs for the two systems are (18.8,30.1) and (20.0,20.0), respectively. The large WAP_R of System III does not move the position of the eigenvalue weak knee significantly, but it helps reshape the eigenvalue curve by reducing its primary slope. The resulting improvement in the equivalent SISO channel quality in turn increases the MIMO capacity of the system. This again proves the authors' assertion in [7] that the performance of a system is not completely determined by the smaller WAP between the transmit side and the receive side. In terms of MIMO design, one would expect the capacity



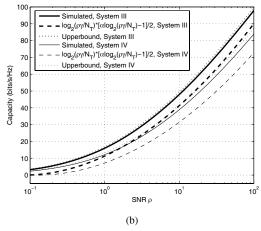


Fig. 7. Simulation results for systems in Figure 6. (a): Normalized average eigenvalues; (b): Ergodic capacity. $\alpha=1.70,\ \gamma=310.68$ for System III; $\alpha=1.54,\ \gamma=230.95$ for System IV.

performance to worsen when the numbers of transmit and receive antennas are both chosen following the smaller WAP instead of the WAPs on their respective sides.

The run time for the full wave simulations described in this section varied with the numbers of transmit antennas and scattering cylinders used in each system. A scattering problem must be solved for each scatterer distribution and each individually excited transmit antenna, which gives rise to one column of the MIMO channel matrix for one instance of the channel ensembles. The discretization adopted by this study introduces 240 unknown for each cylinder. Therefore, when the numbers of cylinders, antennas, and realizations are large, the Monte Carlo simulation could be very computationally intensive, even with the FMM acceleration. For the largest system studied (System IV), every hundred realizations took approximately 18 hours on a DELL desktop computer with a 2.8GHz CPU and 2GB RAM to finish when the residue threshold of the iterative solver was set to 5×10^{-7} .

Two dimensional simulations greatly reduce the computation burden compared to more realistic 3-D models, so that the study of sufficiently large systems is allowed. Despite this seemingly departure from reality, 2-D models are still believed to be able to closely approximate many practical propagation scenarios and are frequently used in the research community. The Jakes one-ring geometric model is one of the most well-known 2-D models in the literature and has been proved successful in many studies of mobile communications [20]. Other similar models include the discrete uniform model [21], the circular scattering model [22], the elliptical scattering model [23], the Gaussian angle-of-arrival model [24], and etc. Most of them employ the simple single bounce scattering from point scatterers (isotropic scattering), and they are widely used to characterize certain 3-D systems whose propagation primarily takes place in the transmitter/receiver plane. However, for the study of spatial diversity and multipath richness, which are extremely important for MIMO systems, singleorder scattering and point scatterers will cause severe oversimplification that fails to capture the true physics of the wave propagation. The importance of multiple scattering was exemplified by the work in [7], and the anisotropy of the angular power spectra of waves through media consisting of practically shaped obstacles was demonstrated in a recent study of the authors [25]. The use of cylinders in full wave electromagnetic simulations includes the influence of the above two factors in a self-consistent way, so that a realistic environment is more accurately characterized. For example, the systems considered in this study all have dimensions around 200λ . Therefore the simulated ensembles can be used to model various indoor environments with an operating frequency 3 GHz, at which the wavelength of the EM fields is $\lambda = 10$ cm.

V. CONCLUSIONS AND DISCUSSIONS

In this paper, the dependence of the MIMO capacity on received SNR is investigated based on a previous study of electromagnetic degrees of freedom between two volumes communicating with each other. A general two-slope exponentially decaying eigenvalue set (squared singular values) of the MIMO channel matrix is found for various types of systems including systems with a single scattering cluster, multiple scattering clusters, a slab scattering medium, and truncated surrounding scattering medium, for the last of which the results are not included because of the disagreement between the WAP and the eigenvalue weak knee position. With this two-slope exponential decay eigenvalue behavior, a lower bound for the MIMO capacity is derived for the power pre-saturation range, which is defined to be the power range where no equivalent SISO channels corresponding to the secondary slope eigenvalues are effective. The obtained lower bound is in the form of the product of two logarithmically increasing functions of SNR. When plotted against SNR in dB, the capacity curve demonstrates a quadratic shape. This square-logarithmical rate may be more accurate to interpret the capacity-SNR behavior in a significant portion of the curve than does the concavity of a simple logarithmic function. It also clarifies that, instead of the commonly accepted linear capacity-SNR dependence, the true capacity increase rate could be quadratic for many MIMO systems. It may be beneficial to increase the transmit power as much as possible to take advantage of this second order return.

The ergodic capacity of a number of MIMO systems is calculated through rigorous full wave electromagnetic simulations that include the complete effect of the multiple scattering of waves from the propagation environments. The results confirmed the square-logarithmical increase rate of capacity with respect to the SNR. The simulated ergodic capacity is also compared with an upper bound due to the concavity of the logarithmic functions and Jensen's inequality. This upper bound is evidenced to be very tight for all the systems considered in this study. However, it needs to be pointed out that the tightness of the bound does not necessarily always hold, which is exemplified by the study of [16]. Furthermore, the tight behavior of the upper bound is also based on the condition that the antenna array is slightly saturated, i.e., the numbers of antennas are slightly larger than the WAP of the underlying EM system. When the numbers of antennas are increased and their separation decreased to arrive at an oversaturated system, the resulting MIMO system will start to pick up some of the less effective second-slope eigenvalues of the propagation operator. It is understandable that the tightness of the upper bound will degrade accordingly. It is also found that two systems that differs only in the area density of the scatterers result in very similar performance as far as the capacity-SNR relation is concerned. Higher scatterer density does not increase the multipath richness of the system indefinitely. Eventually its effect is completely transferred to the overall path loss of the propagation.

The presented study also revealed some implications for MIMO design. The WAP manifests itself to be a very important parameter in determining the number of antennas used in a system. An antenna number slightly larger than the WAP will make all the first slope eigenvalues of the propagation operator potentially usable, while the secondary slope ones, which require much more power to be effective, are ignored. It is also reasonable to reduce the antenna number if only the first few significant eigenvalues are intended for use and less system complexity is desired. The practical significance of the uniform linear array structure was shown in the sense that a uniform array can generally capture most of the spatial domain benefit provided by the linear volume it occupies, and that higher dimensional arrays such as circular ones do not greatly enhance the system performance when wave scattering is concentrated in a limited 2-D (e.g., azimuthal) angular range. A conjecture can be made that a uniform grid planar array geometry normal to the direction of communications may be appropriate when wave scattering happens primarily in a limited 3-D solid angular region between the transmitter and the receiver.

As a last comment, it is the authors' belief that the failure of having a step-like eigenvalue set for the MIMO systems considered in this study is because of the asymmetry of the angular wave propagation between the transmit and receive volumes. It is well known that when the objects in a discrete random medium are electrically large, the scattering of waves therein is mostly forward directing. In other words, the wave interaction between the two volumes is mostly concentrated in the region between them. The study of [7] gave an example where the transmit volume is a circular disk and the receive volume is a circle concentric to and enclosing the disk. The wave interaction in such a system is angularly symmetric, and its eigenvalues were analytically shown to possess an ideal step-like behavior.

ACKNOWLEDGMENT

The research presented here was supported by the National Science Foundation under grant ECS 0300130.

REFERENCES

- [1] E. Telatar, "Capacity of multi-antenna gaussian channels," Eur. Trans. Telecomm., vol. 10, pp. 585-595, 1999.
- [2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," Wireless Personal Commun., vol. 6, no. 3, pp. 311-335, Mar. 1998.
- [3] C. N. Chuah, D. N. Tse, J. M. Kahn, and R. A. Valenzuela, "Capacity scaling in MIMO wireless systems under correlated fading," IEEE Trans. Inform. Theory, vol. 48, no. 3, pp. 637-650, Mar. 2002.
- [4] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multiple-antenna channel hardening and its implications for rate feedback and scheduling, IEEE Trans. Inform. Theory, vol. 50, no. 9, pp. 1893-1909, Sept. 2004.
- [5] S. Wei, D. Goeckel, and R. Janaswamy, "On the asymptotic capacity of MIMO systems with antenna arrays of fixed length," IEEE Commun. Lett., vol. 7, no. 11, pp. 1608-1621, Sept. 2005.
- [6] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple-antenna channels: a signal space approach," IEEE Trans. Inform. Theory, vol. 51, no. 2, pp. 523-536, Feb. 2005.
- [7] J. Xu and R. Janaswamy, "Electromagnetic degrees of freedom in 2-D scattering environments," *IEEE Trans. Antennas Propag.*, vol. 54, no. 12, pp. 3882-3894, Dec. 2006.
- [8] E. Khan and C. Heneghan, "A closed form expression for the ergodic capacity of MIMO systems," Inform. Processes J., vol. 1, pp. 47-57, Feb. 2005.
- [9] J. Salo, P. Suvikunnas, H. M. El-Sallabi, and P. Vainikainen, "Some insights into MIMO mutual information: the high SNR case," IEEE Trans. Wireless Commun., vol. 5, no. 11, pp. 2997-3001, Nov. 2006.
- [10] J. W. Wallace and M. A. Jensen, "MIMO capacity variation with SNR and multipath richness from full-wave indoor FDTD simulations," in Proc. (IEEE) International Symp. Antennas Propagation (AP-S'03), vol. 2, Columbus, OH, June 2003, pp. 523-526.
- [11] P. Almers, F. Tufvesson, P. Karlsson, and A. F. Molisch, "The effect of horizontal array orientation on MIMO channel capacity," in Proc. (IEEE) Semiannual Veh. Technol. Conf. (VTC'03), vol. 1, Jeju, Korea, Apr. 2003, pp. 34-38.
- [12] D. P. McNamara, M. A. Beach, P. N. Fletcher, and P. Karlsson, "Temporal variation of multiple-input multiple-output (MIMO) channels in indoor environments," in Proc. 11th IEE International Conference on Antennas and Propagation (ICAP'01), vol. 2, Manchester, UK, Apr. 2001, pp. 578-582.
- M. A. Jensen and J. W. Wallace, "RF and algorithmic considerations for practical MIMO wireless implementation," in Proc. (IEEE) Radio Wireless Conf. (RAWCON'04), Atlanta, GA, Sept. 2004, pp. 147-150.
- [14] T. S. Pollock, T. D. Abhayapala, and R. A. Kennedy, "Antenna saturation effects on dense array MIMO capacity," in Proc. (IEEE) International Conf. Acoustics, Speech Signal Processing (ICASSP'03), vol. 4, Hong Kong, Apr. 2003, pp. 361-364.
- S. Wei and D. Goeckel, "On minimax robust transmission strategies for MIMO systems," IEEE Trans. Wireless Commun., vol. 4, no. 4, pp. 523-524, Nov. 2003.
- [16] S. Loyka and A. Kouki, "On the use of Jensen's inequality for MIMO channel capacity estimation," in Proc. Canadian Conf. Electrical Computer Engineering (CCECE'01), Toronto, Canada, May 2001, pp. 11-16.
- [17] L. Vuokko, P. Suvikunnas, J. Salo, J. Kivinen, and P. Vainikainen, "Comparison of measured MIMO capacities at 2 and 5 GHz," in Proc. XXVIIIth General Assembly International Union Radio Science (URSI'05), New Delhi, India, Oct. 2005.
- [18] M. Payaró, A. Pascual-Iserte, and M. Lagunas, "Mutual information optimization and capacity evaluation in MIMO systems with magnitude knowledge and phase uncertainty," Circuits, Systems, Signal Processing (special issue: Signal Processing Uncertain Systems), vol. 26, no. 4, pp. 527-549, Aug. 2007.
- [19] M. D. Migliore, "On the role of the number of degrees of freedom of the field in mimo channels," IEEE Trans. Antennas Propag., vol. 54, p. 620°C628, Feb. 2006.
- [20] W. C. Jakes, Microwave Mobile Communications. Piscataway, NJ: IEEE Press, 1974.
- [21] H. Suzuki, "A statistical model for urban radio propagation," IEEE
- Trans. Commun., vol. 25, no. 7, pp. 673-680, July 1977.
 [22] R. B. Ertel and J. H. Reed, "Angle and time of arrival statistics for circular and elliptical scattering models," IEEE J. Select. Areas Commun., vol. 17, no. 11, pp. 1829-1840, Nov. 1999.

- [23] P. Petrus, J. H. Reed, and T. S. Rapport, "Geometrically based statistical channel model for macro cellular mobile environments," in *Proc. (IEEE) Global Telecommun. Conf. (Globecom'96)*, vol. 2, London, UK, Nov. 1996, pp. 1197-1201.
- [24] B. Ottersten, "Spatial division multiple access in wireless communications," in *Proc. Nordic Radio Symposium*, 1995.
- [25] J. Xu and R. Janaswamy, "A double-angular propagation model with cluster scattering," *IEEE Trans. Antennas Propag.*, vol. 57, no. 4, pp. 1228-1240, Apr. 2009.



Jie Xu Jie Xu (S'05-M'08) was born in Wuhan, China, on October 9, 1977. He received his Ph.D. degree in electrical engineering in 2008 from the University of Massachusetts Amherst and his M.S. and B.S. degrees, both in electrical engineering, from Tianjin University, China, in 2003 and 2000, respectively. Before joining University of Massachusetts in fall of 2003, he worked as a software engineer in Zhongxing Telecommunications Equipment (ZTE) Corp. in Shenzhen, China for six months, where he took part in the development

of the system software for IS-2000 1X CDMA cellular systems and was responsible for the radio frequency control part on the BS side. He is currently working as a post-doc research fellow at the Biomedical Sensing and Signal Processing (BioSSP) Center of UMass Amherst and conducting research on cancer detections using microwave frequencies.



Dennsi Goeckel Dennis Goeckel (S'89-M'92-SM'04) split time between Purdue University and Sundstrand Corporation from 1987-1992, receiving his BSEE from Purdue in 1992. From 1992-1996, he was a National Science Foundation Graduate Fellow and then Rackham Pre-Doctoral Fellow at the University of Michigan, where he received his MSEE in 1993 and his Ph.D. in 1996, both in Electrical Engineering with a specialty in Communication Systems. In September 1996, he joined the Electrical and Computer Engineering department at

the University of Massachusetts, where he is currently a Professor. His current research interests are in the areas of communication systems and wireless network theory.

Dr. Goeckel was the recipient of a 1999 CAREER Award from the National Science Foundation for "Coded Modulation for High-Speed Wireless Communications." He was a Lilly Teaching Fellow at UMass-Amherst for the 2000-2001 academic year and received the University of Massachusetts Distinguished Teaching Award in 2007. He served as an Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS: Wireless Communication Series during its transition to the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 1999-2002, and as a Technical Program Committee Co-Chair for the Communication Theory Symposium at IEEE Globecom 2004. He is currently an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and a Co-Chair for the Wireless Communications Symposium at IEEE Globecom 2008.



Ramakrishna Janaswamy Ramakrishna Janaswamy (S'82-M'83-SM'93-F'03) received his Ph.D. degree in electrical engineering in 1986 from the University of Massachusetts, Amherst. From August 1986 to May 1987, he was an Assistant Professor of electrical engineering at Wilkes University, Wilkes Barre, PA. From August 1987-August 2001 he was on the faculty of the Department of Electrical and Computer Engineering, Naval Postgraduate School, Monterey, CA. In September 2001 he joined the Department

of Electrical & Computer Engineering, University of Massachusetts, Amherst, where he is a currently a Professor. Rama Janaswamy is a Fellow of IEEE. He was the recipient of the R. W. P. King Prize Paper Award of the IEEE Transactions on Antennas and Propagation in 1995. For his services to the IEEE Monterey Bay Subsection, he received the IEEE 3rd Millennium Medal from the Santa Clara Valley Section in 2000. He is an elected member of U.S. National Committee of International Union of Radio Science, Commissions B and F. He served as an Associate Editor of Radio Science from January 1999-January 2004 and as an Associate Editor of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2003-2006. He is the author of the book Radiowave Propagation and Smart Antennas for Wireless Communications (Kluwer Academic Publishers, November 2000), and a contributing author in Handbook of Antennas in Wireless Communications (L. Godara (Ed.), CRC Press, August 2001), and the Encyclopedia of RF and Microwave Engineering (K. Chang (Ed.), John Wiley & Sons, 2005). His research interests include deterministic and stochastic radio wave propagation modeling, analytical and computational electromagnetics, antenna theory and design, and wireless communications.