

INFORMATION THEORY AND ELECTROMAGNETISM: ARE THEY RELATED?

Sergey Loyka¹, Juan Mosig²

¹School of Information Technology and Engineering (SITE), University of Ottawa,
161 Louis Pasteur, Ottawa, Ontario, Canada, K1N 6N5, email: sergey.loyka@ieee.org

²Swiss Federal Institute of Technology, LEMA-EPFL, Ecublens,
CH-1015 Lausanne, Switzerland, email: juan.mosig@epfl.ch

I. INTRODUCTION

Multi-antenna systems have recently emerged as a highly efficient strategy for wireless communications in rich multipath channels [1-4]. However, it is also well recognized that the wireless propagation channel has a profound impact on MIMO system performance [3-8]. In ideal conditions (uncorrelated high rank channel) the MIMO capacity scales roughly linearly as the number of Tx/Rx antennas. The effect of channel correlation is to decrease the capacity and, at some point, this is the dominant effect. This effect is highly dependent on the scenario considered. Many practically-important scenarios have been studied and some design guidelines have been proposed as well.

Here we analyze the effect of propagation channel from a completely different perspective [9-14]. Electromagnetic waves are used as the primary carrier of information. The basic electromagnetism laws, which control the electromagnetic field behaviour, are expressed as Maxwell equations [15,16]. Hence, we ask a question: What is, if any, the impact of Maxwell equations on the notion of information in general and on channel capacity in particular? In other words, do the laws of electromagnetism impose any limitations on the achievable channel capacity? Below, we concentrate on this last question and try to answer it. We are not targeting in particular scenarios, rather, we are going to look at fundamental limits that hold in any scenario. Analyzing MIMO channel capacity allows one, in our opinion, to come very close to answering this question.

Our approach is a three-fold one [13]. First, we employ the channel correlation argument and introduce the concept of an *ideal scattering* to demonstrate that the minimum antenna spacing is limited

to about half a wave length for *any* channel (i.e., locating antennas closer to each other will not result in a capacity increase because of correlation).

Secondly, we use the plane wave spectrum expansion of a generic electromagnetic wave and the Nyquist sampling theorem in the spatial domain to show that the laws of electromagnetism in its general form (Maxwell equations) limit the antenna spacing to half a wavelength, $d_{\min} = \lambda/2$, for linear antenna arrays, but only asymptotically, when the number of antennas $n \rightarrow \infty$. For a finite number of antennas, this limit is slightly less than $\lambda/2$ because a slight oversampling is required to reduce the truncation error when using the sampling series. In any case, the existence of the minimum spacing limits the number of antennas and the MIMO capacity for a given aperture size. It should be emphasized that this limitation is scenario-independent. It follows directly from Maxwell equations and is valid in any situation.

Thirdly, we consider the MIMO capacity of waveguide and cavity channels and demonstrate that there are finite number of degrees of freedom in that environment too, which is dictated directly by Maxwell equations, that can be exploited for MIMO communications. Electromagnetics and information theory can be nicely united in this case to produce insight that is not available by using either of these disciplines separately. In particular, it turns out that the traditional single-mode transmission, which is so popular in electromagnetics community, is optimal only at small signal-to-noise (SNR) ratio.

II. MIMO CHANNEL CAPACITY

We employ the celebrated Foschini-Telatar formula for the MIMO channel capacity [1,2], which is valid for a fixed linear $n \times n$ matrix channel with additive white Gaussian noise and when the transmitted signal vector is composed of statistically independent equal power components each with a gaussian distribution and the receiver knows the channel,

$$C = \log_2 \det \left(\mathbf{I} + \frac{\rho}{n} \mathbf{G} \cdot \mathbf{G}^+ \right) \text{ [bits/s/Hz]}, \quad (1)$$

Here n is the numbers of transmit/receive antennas, ρ is the average signal-to-noise ratio, \mathbf{I} is $n \times n$ identity matrix, \mathbf{G} is the normalized channel matrix (the entries are complex channel gains from each Tx to each

Rx antenna), $\text{tr}[\mathbf{G}\mathbf{G}^+] = n$, which is considered to be frequency independent over the signal bandwidth, and “+” denotes transpose conjugate. For simplicity, we consider a $n \times n$ channel, but the results also hold true, sometimes with minor modifications, for a $n_R \times n_T$ channel, where n_R and n_T are the number of Tx and Rx antennas respectively, $n_R \neq n_T$.

In an ideal case of orthogonal full-rank channel (1) reduces to

$$C = n \log_2(1 + \rho/n), \quad (2)$$

i.e. the capacity is maximum and scales roughly linearly with the number of antennas.

III. THE LAWS OF ELECTROMAGNETISM

It follows from (1) that the MIMO channel capacity crucially depends the propagation channel \mathbf{G} . Since electromagnetic waves are used as the carrier of information, the laws of electromagnetism must have an impact on the MIMO capacity. They ultimately determine the behaviour of \mathbf{G} in different scenarios. Hence, we outline the laws of electromagnetism from a MIMO system perspective. In their most general form, they are expressed as Maxwell equations with charge and current densities as the field sources [15,16]. Appropriate boundary conditions must be applied in order to solve them. We are interested in application of Maxwell equations to find the channel matrix \mathbf{G} in (1). Since the Rx antennas are physically separated from the Tx ones, we assume that the physical support of our channel is a source-free space, which includes scatterers, where EM waves do propagate. In this case, Maxwell equations simplifies to the system of two decoupled wave equations [15]:

$$\nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0 \quad \nabla^2 \mathbf{H} - \frac{1}{c^2} \frac{\partial^2 \mathbf{H}}{\partial t^2} = 0 \quad (3)$$

where \mathbf{E} and \mathbf{H} are electric and magnetic field vectors, and c is the speed of light. There are 6 independent field components (or “polarizational degrees of freedom”) associated with (3) (three for electric and three for magnetic fields), which can be used for communication in rich-scattering environment. Only two of them survive in free space at the far-field region (“poor scattering”). Hence, in

a generic scattering case the number of polarizational degrees of freedom varies between 2 and 6, and each of them can be used for communication. Using the Fourier transform in time domain,

$$\phi(\mathbf{r}, \omega) = \int \phi(\mathbf{r}, t) e^{-j\omega t} dt \quad (4)$$

(3) can be expressed as [15]

$$\nabla^2 \phi(\mathbf{r}, \omega) + (\omega/c)^2 \phi(\mathbf{r}, \omega) = 0 \quad (5)$$

where ϕ denotes any of the components of \mathbf{E} and \mathbf{H} , \mathbf{r} is a position vector and ω is the frequency. For a given frequency ω (i.e., narrowband assumption), (5) is a second-order partial differential equation in \mathbf{r} . It determines ϕ (for given boundary conditions, i.e. a Tx antenna configuration and scattering environment) and, ultimately, the channel matrix and the channel capacity. Note that (5) is general as it does not require any significantly-restrictive assumptions. The source-free region assumption seems to be quite natural (i.e., Tx and Rx antennas are separated) and the narrowband assumption is simplifying but not restrictive since (5) can be solved for *any* frequency and, further, the capacity of a frequency-selective channel can be evaluated using well-known techniques.

Unfortunately, the link between (5) and the channel matrix \mathbf{G} is not explicit at all. A convenient way to study this link is to use the spatial domain Fourier transform, i.e. the plane-wave spectrum expansion,

$$\begin{aligned} \phi(\mathbf{k}, \omega) &= \int \phi(\mathbf{r}, \omega) e^{j\mathbf{k} \cdot \mathbf{r}} d\mathbf{r} \\ \phi(\mathbf{r}, t) &= \frac{1}{(2\pi)^4} \iint \phi(\mathbf{k}, \omega) e^{j(\omega t - \mathbf{k} \cdot \mathbf{r})} d\mathbf{k} d\omega \end{aligned} \quad (6)$$

where \mathbf{k} is the wave vector. Using (6), (5) can be reduced to [15]

$$\left(|\mathbf{k}|^2 - (\omega/c)^2 \right) \phi(\mathbf{k}, \omega) = 0 \quad (7)$$

Hence, $|\mathbf{k}| = \omega/c$ and the electromagnetic field is represented in terms of its plane-wave spectrum $\phi(\mathbf{k}, \omega)$, which in turn is determined through given boundary conditions, i.e. scattering environment and

Tx antenna configuration. In the next sections, we discuss limitations imposed by (5)-(7) on the MIMO channel capacity.

IV. SPATIAL CAPACITY AND CORRELATION

The channel capacity is defined as the maximum mutual information [17],

$$C = \max_{p(\mathbf{x})} \{I(\mathbf{x}, \mathbf{y})\} \quad (8)$$

where \mathbf{x}, \mathbf{y} are Tx and Rx vectors, and the maximum is taken over all possible transmitted vectors subject to the total power constraint, $P_x = \langle \mathbf{x}\mathbf{x}^+ \rangle \leq P_T$. Under some conditions (quasi-static frequency-flat channel with additive white Gaussian noise (AWGN), with perfect channel state information (CSI) at the receiver), this results in (1). In order to study the impact of the electromagnetics laws on the channel capacity and following the approach of [1,2], we define the *spatial capacity* S as the maximum mutual information between the Tx vector on one side and the pair of the Rx vector \mathbf{y} and the channel \mathbf{G} (assuming perfect CSI at the Rx end) on the other, the maximum being taken over both the Tx vector and EM field distributions,

$$S = \max_{p(\mathbf{x}), \mathbf{E}} \{I(\mathbf{x}, \{\mathbf{y}, \mathbf{G}(\mathbf{E})\})\}, \quad (9)$$

$$\text{const.: } \langle \mathbf{x}^+ \mathbf{x} \rangle \leq P_T, \quad \nabla^2 \mathbf{E} - \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} = 0, \quad \mathbf{E} = \mathbf{E}_0 \forall \{\mathbf{r}, t\} \in B$$

where, to be specific, we assume that the electric field \mathbf{E} is used to transmit data (\mathbf{H} field can be used in the same way), and the last constraint is due to the boundary condition B associated with the scattering environment. The first constraint is the classical power constraint and the second one is due to the wave equation. The channel matrix \mathbf{G} is a function of \mathbf{E} since the electric field is used to send data. The spatial capacity S is difficult to find in general since the constraints include a partial differential equation with arbitrary boundary conditions.

One may consider a reduced version of this problem by defining a spatial MIMO capacity as a maximum of the conventional MIMO channel capacity (per unit bandwidth, i.e. in bits/s/Hz) over

possible propagation channels (including Tx & Rx antenna locations and scatterers' distribution), subject to some possible constraints. In this case, we replace (9) by

$$S = \max_{\mathbf{G}} \{C(\mathbf{G})\}, \text{ const.: } \mathbf{G} \in \mathcal{M}(\text{Maxwell}) \quad (10)$$

where the constraint $\mathcal{M}(\text{Maxwell})$ is due to the Maxwell (wave) equations and the capacity is maximized by changing \mathbf{G} (within some limits), for example, by appropriate positioning of antennas. Unfortunately, the explicit form of the constraint \mathcal{M} is not known. Additional constraints (due to a limited aperture, for example) may be also included. The aperture constraint was discussed in [25] by introducing the concept of intrinsic capacity, which is somewhat similar to our concept of spatial capacity. Note that the second definition (10) will give a spatial capacity, which is, in general, less than that in the first definition (9).

We have termed the maxima in (9) and (10) “spatial capacity” or “capacity of a given space”. Since we have to vary the channel during this maximization the name “channel capacity” seems to be inappropriate simply because the channel is not fixed. On the other hand, we vary channel within some limits, i.e. within given space. Thus, the term “capacity of a given space”, or “spatial capacity”, seems to be appropriate.

The question arises: what is this maximum and what are the main factors that have an impact on it? Using the ray tracing (geometrical optics) arguments and the recent result on the MIMO capacity, we further demonstrate that there exists an optimal distribution of scatterers and of Tx/Rx antennas that provides the maximum possible capacity in a given region of space. Hence, the MIMO capacity per unit spatial volume can be defined in a fashion similar to the traditional definition of the channel capacity per unit bandwidth. This allows the temporal and spatial domains to enter into the analysis on equal footing and, hence, demonstrates explicitly the space-time symmetry of the capacity problem in the spirit of special relativity in physics.

In order to proceed further, we need some additional assumptions. Considering a specific scenario would not allow us to find a fundamental limit simply because the channel capacity would depend on too many specific parameters. For example, in outdoor environments the Tx and Rx ends of the system are

usually located far away from each other. Hence, any MIMO capacity analysis (and optimization) must be carried out under the constrain that the Tx and Rx antennas cannot be located close to each other. However, there exists no fundamental limitation on the minimum distance between the Tx and Rx ends. Thus, this maximum capacity would not be a fundamental limit. In a similar way, a particular antenna design may limit the minimum distance between the antenna elements but it is just a design constrain rather than a fundamental limit. Similarly, the antenna design has an effect on the signal correlation (due to the coupling effect, for example), but this effect is very design-specific and, hence, is not of fundamental nature. In other words, the link between the wave equations (3) or (7) and the channel matrix \mathbf{G} is far from explicit since too many facts depend on Tx and Rx antenna designs and on many other details.

We will rather consider a reduced version of this problem. In particular, we investigate the case when the Tx and Rx antenna elements are constrained to be located within given Tx and Rx antenna apertures. We are looking for such location of antenna elements (within the given apertures) and such distribution of scatterers that the MIMO capacity (“spatial capacity”) is maximum. While this maximum may not be achievable in practice, it gives a good indication as to what the potential limits of MIMO technology are.

In order to avoid the effect of design-specific details, we adopt the following assumptions. Firstly, we consider a limited antenna aperture size (1-D, 2-D or 3-D) for both the Tx and Rx antennas. All the Tx (Rx) antenna elements must be located within the Tx (Rx) aperture. As it is well-known, a rich scattering environment is required to order to achieve high MIMO capacity. Thus, secondly, the rich (“ideal”) scattering assumption is adopted in its most abstract form. Specifically, it is assumed that there is infinite number of randomly and uniformly-located ideal scatterers (the scattering coefficient equals to unity), which form a uniform scattering medium in the entire space (including the space region considered) and which do not absorb EM field. This is the concept of “ideal scattering” (which cannot be better than that). Thirdly, antenna array elements are considered to be ideal field sensors with no size and no coupling between the elements in the Rx (Tx) antenna array. Our goal is to find the maximum MIMO channel

capacity in such a scenario (which posses no design-specific details) and the limits imposed by the electromagnetism laws. It should be emphasized that the effect of electromagnetism laws is already implicitly included in some of the assumptions above. In order to simplify analysis further, we use the ray (geometrical) optics approximation (which justifies the ideal scattering assumption above).

Knowing that the capacity increases with the number of antennas, we try to use as many antennas as possible. Is there any limit to it? Since antennas have no size (by the assumption above), the given apertures can accommodate the infinite number of antennas. However, if antennas are located close to each other the channel correlation increases and, consequently, the capacity decreases. A certain minimum distance between antennas must be respected in order to avoid capacity decrease, even in ideal rich scattering. Fig. 2 demonstrates this effect for uniformly-spaced linear array antennas for the scattering scenario depicted in Fig. 1: if $d < d_{\min}$, the effect of correlation is significant and the capacity is less than the maximum one [8]. While d_{\min} depends on scattering environment, i.e. the angular spread Δ of incoming multipath,

$$\frac{d_{\min}}{\lambda} \approx \max \left\{ \frac{1}{2\Delta}, 0.5 \right\},$$

even in rich scattering (i.e. $\Delta = 360^\circ$) $d_{\min} \approx \lambda/2$, which is consistent with the Jakes model [22]. While the model above is a two-dimensional (2D) one, it can be extended to 3D applying it to both orthogonal planes and, due to the symmetry of the problem (no preferred direction), similar result should hold in 3D as well. Rigorous analysis shows that the correlation between adjacent elements in that case is $\sin(2\pi d/\lambda)/(2\pi d/\lambda)$ (with the first zero at $d_0 = \lambda/2$) and the same minimum spacing requirement holds true, $d_{\min} \approx \lambda/2$.

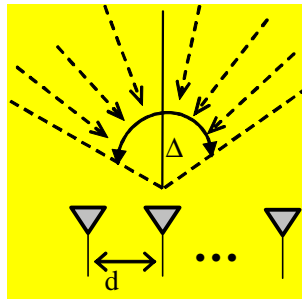


Fig.1. Incoming multipath signals arrive to a uniform linear antenna array of isotropic elements within $\pm\Delta/2$ of the broadside direction.

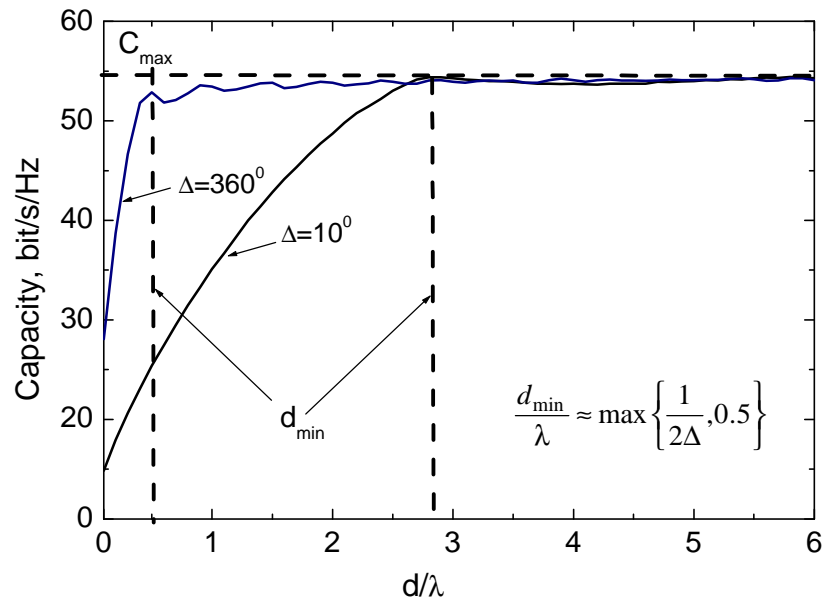


Fig. 2. The average capacity versus antenna spacing for the uniform linear array and the single-cluster multipath channel of Fig. 1 with angular spread Δ at the Rx end (the Tx end is assumed to be uncorrelated); $n=10$, SNR=30 dB . When $\Delta = 360^\circ$, $d_{\min} \approx \lambda/2$.

We summarize the effect of d_{\min} as follows. When we increase the number of antenna elements over a fixed aperture, the capacity at first increases. But at some point, due to aperture limitation, we have to decrease the distance between adjacent elements to accommodate new ones within the fixed aperture. When the element spacing decreases, the capacity increase slows down and finally, when the element

spacing is less than the minimum distance, $d_{\min} \approx \lambda/2$, the capacity saturates. Hence, there is an optimal number of antennas, for which the full capacity is achieved with the minimum number of antenna elements (i.e., the minimum complexity). Fig. 3 demonstrates this capacity saturation effect for a fixed-aperture antenna array using the model in [8]. Note the dual nature of capacity saturation: the capacity saturates with increasing the element spacing over d_{\min} for given number of elements n (1st type saturation - Fig. 2) and also with increasing n over number N_{opt} for given aperture length L (2nd type saturation - Fig. 3).

A similar capacity saturation argument has already been presented earlier in [9]. However, no appropriate model has been developed and also the optimal number of antennas has not been evaluated. Using the model in [8] or the equivalent 3D model results in $d_{\min} \approx \lambda/2$ for $\Delta = 360^\circ$ (uniform scattering), and the optimal number N_{opt} of antennas for a given aperture size L is straightforward to evaluate (1-D aperture, i.e. linear antenna array):

$$N_{opt} \approx 2L/\lambda + 1 \quad (11)$$

For an angular spread smaller than 360° , the optimal number of antennas decreases correspondingly,

$$N_{opt} \approx \frac{2L}{\lambda} \min\{\Delta, 1\} + 1$$

It should be emphasized that the 2nd type capacity saturation has been observed under the assumption of fixed average SNR at the receiver, which is equivalent to $tr\{\mathbf{G}\mathbf{G}^+\} = n_T$, where n_T is the number of Tx antennas, which is fixed. Since increasing n results in more power collected by the Rx antenna elements, which is equivalent to increasing the average SNR, it was speculated that the 2nd type saturation may not exist if the SNR increase is taken into account [26]. As we show later, this is not so (since the total power collected by the Rx array is limited for fixed L).

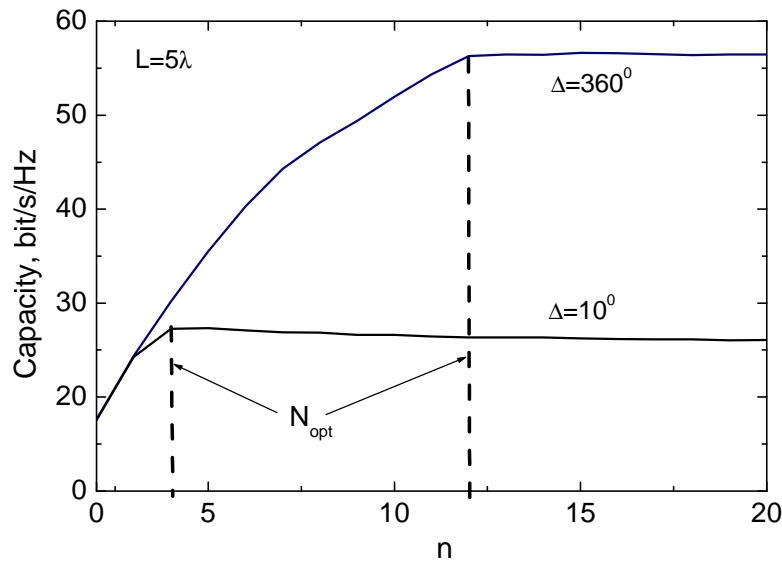


Fig. 3. The average capacity versus the number of elements n of a uniform linear array and a single-cluster multipath channel with the angular spread Δ at the Rx end. The Tx end is assumed to be uncorrelated, the number of Tx antenna elements is $n_T = 10$; the aperture length $L = 5\lambda$, SNR=30 dB.

While our analysis above was based on 1-D antenna arrays, the similar saturation effects can be observed for 2-D and 3-D arrays as well. Additionally, the capacity saturation effect has been also noted for circular arrays [24]. Hence, this effect is not a consequence of a specific array geometry but rather a generic property of any array: capacity saturates as long as adjacent element spacing is about $\lambda/2$, regardless of the geometry¹. Finally, we note that the results above are consistent with the diversity combining analysis, where the minimum spacing is about half a wavelength (for $\Delta = 360^\circ$) as well [22], and with the earlier speculation in [1].

¹ We should note that mutual coupling between antenna elements is not taken into account in the present study. Based on the results in [30,31], if this effect is accounted for, the capacity may actually decrease for $n > N_{opt}$ since $d < \lambda/2$ in that region.

V. SPATIAL SAMPLING AND MIMO CAPACITY

In the previous section, we argued that the channel correlation limits the minimum antenna spacing to half a wavelength (even in the case of “ideal” scattering). In this section, we demonstrate that the same limit can be obtained directly from the wave equations (3) or (5), without reference to the channel correlation.

Let us start with the wave equation (5). The field spectrum $\phi(\mathbf{k}, \omega)$ can be computed in a general case provided there is a sufficient knowledge of the propagation channel and of the Tx antennas (note that we have not made so far any simplifying assumptions regarding the propagation channel). Knowing the field, which is given by the inverse Fourier transform in (6), and Rx antenna properties, one may further compute the signal at the antenna output and, hence, the channel matrix \mathbf{G} . The result will, of course, depend on the Rx antenna design details. In order to find a fundamental limit, imposed by the wave equations (5) on the channel capacity (1), we have to avoid any design-specific details. Thus, as we did earlier on, we assume that the Rx antennas are ideal field sensors (with no size, no coupling between them etc.) and, consequently, the signal at the antenna output is proportional to the field (any of the 6 field components may be used). Hence, the channel matrix entries g_{ij} must satisfy the same wave equation as the field itself. In general, different Tx antennas will produce different plane-wave spectra around the Rx antennas and, hence, the wave equation is:

$$\left(|\mathbf{k}|^2 - (\omega/c)^2 \right) \mathbf{g}_j(\mathbf{k}, \omega) = 0 \quad (12)$$

where $\mathbf{g}_j(\mathbf{k}, \omega)$ is the plane-wave spectrum produced by j-th Tx antenna. To simplify things further, we employ the narrowband assumption: $\omega = \text{const}$, and, hence, $|\mathbf{k}| = \omega/c$ is constant (the case of a frequency-selective channel can be analyzed in a similar way – see below). The channel matrix entries for given locations of the Rx antennas can be found using the inverse Fourier transform in the wave vector domain:

$$\mathbf{g}_j(\mathbf{r}, \omega) = \frac{1}{(2\pi)^3} \int \mathbf{g}_j(\mathbf{k}, \omega) e^{-j\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}, \quad g_{ij} = \mathbf{g}_j(\mathbf{r}_i, \omega) \quad (13)$$

where \mathbf{r}_i is the position vector of i -th Rx antenna, and $\mathbf{g}_j(\mathbf{r}, \omega)$ is the channel “vector”, i.e. propagation factor from j -th Tx antenna to an Rx antenna located at position \mathbf{r} . The integration in (13) is performed on a hypersurface $|\mathbf{k}| = \omega/c$. As we show below, this results in a very important consequence. Consider, for simplicity, 2-D case (3-D case can be considered in a similar way). In this case, the integration in (13) is performed along the line given by

$$k_x^2 + k_y^2 = (\omega/c)^2 \rightarrow k_x = \pm\sqrt{(\omega/c)^2 - k_y^2} \quad (14)$$

Assume that the Rx antenna is a linear array of elements located on the OX axis, i.e. $r_y = 0$. In this case, (13) reduces to

$$\mathbf{g}_j(x, \omega) = \frac{1}{(2\pi)^2} \int_{-k_{\max}}^{k_{\max}} \mathbf{g}_j(k_x, \omega) e^{-jk_x r_x} dk_x, \quad (15)$$

$$g_{ij} = \mathbf{g}_j(x_i, \omega)$$

where $k_{\max} = \omega/c$ due to (14). At this point, we ignored the evanescent waves with $|k| > k_{\max}$ because they decay exponentially with distance and can be ignored at distances more than few λ from the source [15,16]. Note that computing g_{ij} corresponds to sampling $\mathbf{g}_j(x, \omega)$ with sampling points being x_i . Let us now apply the Nyquist sampling theorem to (15). According to it, a band-limited signal, $\mathbf{g}_j(k_x, \omega)$ in our case (it is band-limited in k_x -domain), can be exactly recovered from its samples taken at a rate equal at least to twice the maximum signal frequency (Nyquist rate). In our case, the Nyquist rate is $2k_{\max}$ and the sampling interval is

$$\Delta x_{\min} = 2\pi/(2k_{\max}) = \lambda/2 \quad (16)$$

where $\lambda = 2\pi c/\omega$ is the wavelength. There is no any loss of information associated with the sampling since the original channel “vector” $\mathbf{g}_j(\mathbf{r}, \omega)$ (as well as the field itself) can be recovered exactly from its samples at $x = 0, \pm\Delta x_{\min}, \pm 2\Delta x_{\min}, \dots$. This means that by locating the field sensors at sampling points, which are separated by Δx_{\min} , we are able to recover all the information transmitted by electromagnetic

waves to the receiver. Hence, the channel capacity is not reduced. This implies, in turn, that the minimum spacing between antennas is half a wavelength:

$$d_{\min} = \Delta x_{\min} = \lambda/2 \quad (17)$$

Locating antennas more close to each other does not provide any additional information and, hence, does not increase the channel capacity. It should be noted that the same half-wavelength limit was established in Sec. IV using the channel correlation argument, i.e. locating antennas closer will increase correlation and, hence, capacity will decrease. However, while the channel correlation argument may produce some doubts as whether the limit is of fundamental nature or not (correlation depends on a scenario considered), the spatial sampling argument demonstrates explicitly that the limit is of fundamental nature because it follows directly from Maxwell equations (i.e., the wave equation), without any simplifying assumptions as, for example, the geometrical optics approximation [18] (when evaluating correlation, we have to use it to make the ray tracing valid). Note that the spatial sampling arguments holds also for a broadband channel (the smallest wavelength, corresponding to the highest frequency, should be used in this case to find Δx_{\min}) and for the case of 2-D and 3-D antenna apertures. However, in the latter two cases the minimum distance (i.e., the sampling interval) is different [21]. If one uses a 2-D antenna aperture (i.e. 2-D sampling), the sampling interval is

$$\Delta x_{\min,2} = \lambda/\sqrt{3}, \quad (18)$$

and in the case of 3-D aperture,

$$\Delta x_{\min,3} = \lambda/\sqrt{2}. \quad (19)$$

While the minimum distance in these two cases is different from the 2-D case, $\Delta x_{\min} < \Delta x_{\min,2} < \Delta x_{\min,3}$ (i.e., each additional dimension possesses less degrees of freedom than the previous one), the numerical values are quite close to each other.

Another interpretation of the minimum distance effect can be made through a concept of the number of degrees of freedom. As the sampling argument shows, for any limited region of space (1-D, 2-D or 3-D), there is a limited number of degrees of freedom possessed by the EM field itself. No any

antenna design or their specific location can provide more. This is a fundamental limitation imposed by the laws of electromagnetism (Maxwell equations) on the MIMO channel capacity.

An important remark, often overlooked, on using the sampling theorem to find the minimum antenna spacing is worth mentioning. The sampling theorem guarantees that the original band-limited signal can be recovered from its samples provided that the infinite number of samples are used (band-limited signal cannot be time limited!). Hence, the half wavelength limit, as derived using the sampling theorem, holds true only asymptotically, when $n \rightarrow \infty$. For finite n , the sampling series does not represent exactly anymore the continuous signal (field) due to the truncation error [19]. This is often overlooked in the array processing area [28] when the minimum antenna (sensor) spacing is derived for $n = \infty$ while the number of antennas is actually finite. In the latter case, the optimal number of antennas may be larger than that given by (11), i.e. the minimum spacing may be less than $\lambda/2$ because a slight oversampling is required to reduce the truncation error. The maximum truncation error of the sampling series for a given limited space region (i.e., the antenna aperture in our case) decreases to zero as the number of terms in the sampling series (i.e., the number of antennas in our case) increases and provided that there is a small oversampling [19, 20]. Below we present some truncation error bounds and discuss them in the context of spatial sampling for MIMO system.

A. Bounds on Truncation Error in Sampling Series

Consider reconstruction of a band-limited signal $x(t)$ from its samples $x(n\Delta)$:

$$x(t) = \sum_{n=-\infty}^{\infty} x(n\Delta) \text{sinc}(f_s t - n), \quad (20)$$

where $\text{sinc}(t) = \sin(\pi t) / (\pi t)$, $\Delta = 1/f_s$ and $f_s \geq 2f_{\max}$ are the sampling interval and frequency respectively, f_{\max} is the maximum frequency in the spectrum of $x(t)$,

$$x(t) = \int_{-f_{\max}}^{f_{\max}} S_x(f) e^{j2\pi f t} df, \quad (21)$$

where $S_x(f)$ is the spectrum of $x(t)$. When the series in (20) is truncated to $|n| \leq N$,

$$x_N(t) = \sum_{n=-N}^N x(n\Delta) \text{sinc}(f_s t - n), \quad (22)$$

the truncation error is

$$\varepsilon(t) = x(t) - x_N(t) = \sum_{|n|>N} x(n\Delta) \text{sinc}(f_s t - n), \quad (23)$$

Several bounds to $|\varepsilon(t)|$ are known [19, 20], depending on the nature of the signal and the interval of interest. When the recovered signal $x_N(t)$ is considered over a finite interval only (i.e. limited antenna aperture), $|t| \leq T = N\Delta$, $|\varepsilon(t)|$ can be bounded as [20]:

$$\frac{|\varepsilon(t)|}{\sqrt{E}} \leq \frac{\sqrt{2}}{\pi} \left| \sin \frac{\pi t}{\Delta} \right| \sqrt{\frac{T\Delta}{T^2 - t^2}}, \quad |t| \leq T, \quad (24)$$

where E is the signal's energy,

$$E = \int_{-f_{\max}}^{f_{\max}} |S_x(f)|^2 df, \quad (25)$$

As (24) indicates, when $\Delta \rightarrow 0$ (i.e. increasing oversampling) for fixed T (i.e. more antennas for fixed antenna aperture) or when $T \rightarrow \infty$ for fixed Δ , we obtain $|\varepsilon(t)| \rightarrow 0$. In practical terms, as the mean squared error (MSE) $\overline{\varepsilon^2} = T^{-1} \int_0^T |\varepsilon(t)|^2 dt$ becomes smaller than the noise power, $\overline{\varepsilon^2} < \sigma_0^2$, its impact on the capacity is small and, hence, it can be neglected. A tighter bound can be obtained from (24) by using the energy carried out by the truncated samples instead of the total energy E [20]. We also note that (24) does not necessarily require oversampling.

Another bound to $|\varepsilon(t)|$, which does involve oversampling, is of the following form [19]:

$$\frac{|\varepsilon(t)|}{\max\{|x(t)|\}} \leq \frac{4}{\pi^2 N(1-\alpha)}, \quad -\infty < t < \infty, \quad (26)$$

where $\alpha = f_{\max} / f_s$ is the oversampling ratio. Note that this bound limits the error for the entire range of t . Clearly, as $N \rightarrow \infty$, the truncation error $|\varepsilon(t)| \rightarrow 0$ for any $\alpha < 1$. (We note that (24) and (26) also

justify the use of truncated series in the time-domain as any transmission spans a finite number of symbols. The difference between time-domain and spatial domain cases is that in the former case the number of samples is much larger and, hence, the truncation error is much smaller. On the contrary, since the number of antennas in many practical systems is small, the truncation error may be, generally speaking, significant). The smaller α (i.e. larger oversampling), the smaller N is required for the same bound. While the convergence of $|\epsilon(t)|$ to 0 in (24) and (26) is slow in N , these bounds, in many cases, overestimate the error, which converges to 0 much faster with N . Additionally, it should be noted that the bound in (26) applies to the entire range, $-\infty < t < \infty$ while the function is recovered from the samples in $-T < t < T$ and, hence, significant contribution to the error may come from the interval $|t| > T$, i.e. outside of the antenna aperture. This interval does not contribute anything to the capacity for aperture-limited system and, hence, this part of the error is irrelevant. **Needless to say, more accurate bounds can be obtained if more details are known about the signal [35].**

Fig. 4 illustrates the normalized MSE $\overline{\epsilon^2}/c$ using the bounds in (24) and (26), where $c = E$ and $c = \max |x(t)|^2$ respectively, versus the number of samples. Clearly, for more than 10 samples the error is already small.

B. Impact of Truncation Error on the Capacity

As the discussion above demonstrates, even for small oversampling the truncation error goes to zero as the number of antennas (samples) increases. While the truncation error expressions above are useful on their own (in particular, because they provide a confidence that indeed a truncated sampling series can be good enough), they not only overestimate the error in many cases, but also do not indicate explicitly the effect of the truncation on the capacity.

A way to overcome this difficulty is consider the true mean squared error and to compare it with the noise power. When the squared truncation error averaged over the antenna aperture is less than the noise power, $MSE < 1/SNR$, it is negligible as one is able to recover almost all the information conveyed

by the EM field to the antenna aperture (but, possibly, not outside of the aperture) in given noise. For example, using Fig. 4, $SNR = 20dB$ corresponds to $MSE < 0.01$ and $N > 20$ or $N > 35$ using (24) or (26) respectively. It should not be surprising that these bounds are different as different normalizations are used in (24) and (26); also the nature of the bounds themselves is different, i.e. (26) implies oversampling but (24) does not (it is clear from Fig. 4 that oversampling results in much smaller truncation error when N is not too small). Note also that larger SNR requires larger number of samples to make the truncation error small (less than the noise). Using (26), the required number of samples, which provides negligible truncation error for given SNR ρ , can be estimated as $N > \frac{4\sqrt{\rho}}{\pi^2(1-\alpha)}$.

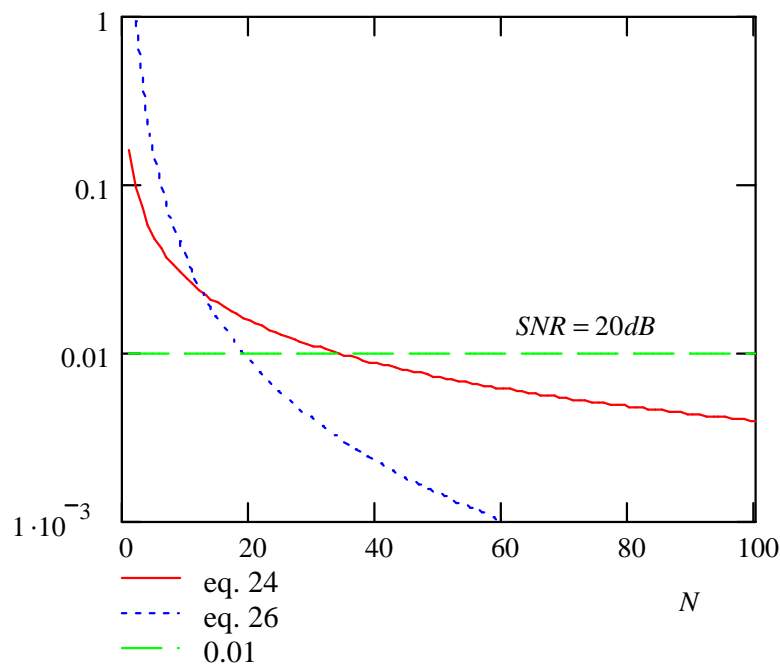


Fig. 4. Normalized mean square truncation error bounds versus the number of samples; $\alpha = 0.8$.

Since the truncation error is zero for infinite number of samples and the required spacing is $d_{\min} = \lambda/2$ in this case, one may expect that the actual minimum antenna spacing is quite close to half a

wavelength for a finite but large number of antennas. The channel correlation argument, which roughly does not depend on n , also confirms this. Detailed analysis shows that the truncation error effect can be eliminated by approximately 10% increase in the number of antennas for many practical cases. Fig. 5 illustrates the effect of oversampling by considering the MIMO capacity versus the number of antennas for given (fixed) aperture length (linear antenna) $L = 5\lambda$ for different realizations of an independent identically-distributed (i.i.d.) Rayleigh fading channel. Clearly, there exists a maximum number of antennas n_{\max} ; using more antennas does not result in higher capacity for any channel realization. Remarkably, that this maximum is slightly larger than that in (11), i.e. spatial sampling and correlation arguments agree well. There is, however, one significant difference between these two arguments: while the latter is valid “on average” (i.e. for the mean capacity), the former is valid for each channel realization and not only on average. Clearly, the sampling argument is more powerful in this respect.

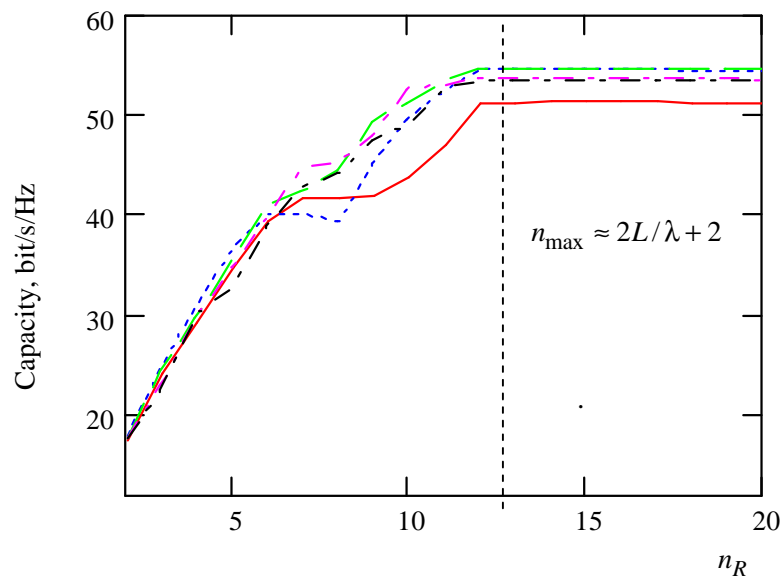


Fig. 5 MIMO channel capacity versus the number of Rx antennas for $L = 5\lambda$, $n_T = 10$. Capacity of 5 different realizations of a Rayleigh fading channel are shown. Capacity saturation for each of them is clear.

Keeping this in mind, one may say, based on the sampling theorem, that the optimal number of antennas for a given aperture size is given approximately by (11). Due to the reciprocity of (1), the same argument holds true for the transmit antennas as well. Hence, using (2) and (11) the maximum MIMO capacity can be found for a given aperture size.

It should be noted that, in some practical cases, increasing n over N_{opt} in (11) may result in higher SNR due to antenna gain increase (i.e. more power collected by the Rx antenna elements) and, consequently, in logarithmic increase in capacity². However, if this increase does take place, it is very slow (logarithmic) and it does not occur if the SNR is fixed, i.e. when one factors out the effect of the antenna gain. From physical perspective, the total power collected by the antenna array cannot exceed the power collected by the ideal continuous aperture of the same size, which equals to the total power delivered by the electromagnetic wave. Consequently, the array antenna gain versus the number of elements for a fixed aperture is limited by the gain of a continuous antenna (with the same aperture). As an example, Fig. 6 shows the gain of a uniform linear array of isotropic elements versus the number of elements, computed using the well-know model [26], and compares it to the gain of continuous linear antenna (aperture) of the same size $L = 5\lambda$. Clearly, the array gain saturates at about the same point as the capacity, $n \approx 2L/\lambda$, which corresponds to $d \approx \lambda/2$, and equals to that of the continuous aperture at that point (this is explained, of course, by the convergence of the array antenna pattern to that of the linear aperture as the number of array elements increases within the fixed aperture). Hence, there is no increase in SNR beyond that point and, consequently, no increase in capacity can be expected due to higher SNR. As a side remark, we note that the similar gain saturation effect can be observed for an array of collinear short dipoles and other elements.

In many practical cases, the minimum spacing can be substantially larger than that in (17). For example, when all the multipath components arrive within a narrow angle spread $\Delta \ll 1$, $d_{min} \approx \lambda/(2\Delta) \gg \lambda/2$ [8] (remarkably, the same limit follows from the spatial sampling argument).

² A. Molisch, private communication.

Hence, less antennas can be accommodated within given aperture,

$$N_{opt} \approx 2L\Delta/\lambda + 1,$$

and, consequently, the MIMO capacity is smaller for a given aperture size.

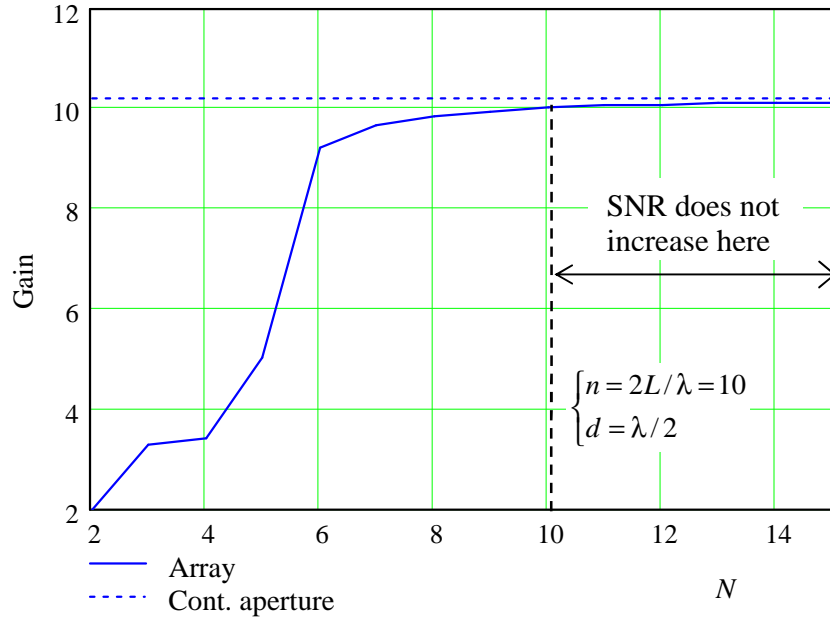


Fig. 6. Linear uniform array gain versus the number of (isotropic) elements for $L = 5\lambda$. Continuous aperture (1-D) gain is also indicated.

A note is in order on practical value of the minimum spacing results above. Since these results were derived under the assumption of ideal antennas (isotropic field sensors with no mutual coupling), it is expected that practical implementation of real antenna arrays operating in real scattering environment may result in some deviations from the results above. For example, it was observed in the literature that mutual coupling (neglected in the present study) may have a significant effect on the capacity [28-31]. Since both positive (i.e. higher capacity) [28,29] and negative [30,31] effects of the mutual coupling have been reported, one has to conclude that this effect depends significantly on the environment, antenna design and also on the assumptions, i.e. whether the SNR is assumed to be fixed or affected by the

coupling, whether the matching network takes the coupling into account etc. [32]. This demonstrates once more that numerous details tend to hide fundamental principles behind them. A side remark is that when this effect is positive, smaller element spacing becomes feasible without significant capacity loss [29]. This clearly shows that practical implementation may somewhat deviate from our idealistic theoretical analysis³. There exists, however, a final number of degrees of freedom possessed by the electromagnetic field itself, which ultimately limits any practical system (with or without mutual coupling etc.). The way these degrees of freedom are used in practice by realistic antennas may of course differ from what the idealistic theory above suggests.

We should note that applications of the sampling theorem to electromagnetic problems have a long history. Among others, these applications include spatial sampling techniques in optics [33], which is electromagnetics at very high, i.e. optical, frequencies, and also in near-field measurements and numerical analysis of electromagnetic fields [34]. A significant difference, however, with the present analysis is that while the former deal mostly with free-space propagation, the latter considers a (rich) scattering environment, where the advantages of MIMO systems are most pronounced (as free-space propagation results in rank-deficient channel and, hence, low capacity, unless the Tx-Rx antenna spacing is very small). Additionally, while the earlier applications have used the sampling theorem just as a convenient analysis tool, when the sampling theorem is considered in the context of MIMO systems, a fundamental link with information theory becomes clear.

With respect to the minimum antenna spacing, it is interesting to note that the MIMO capacity analysis of waveguide channels, which is based on a rigorous electromagnetic approach and does not involve the usage of the sampling theorem, indicates that the minimum antenna spacing is about $\lambda/2$ as well [12,14]. This is discussed in details in the next section.

³ While mutual coupling can have a significant effect on antenna array pattern (especially in the sidelobe region) even for $d > \lambda/2$, the MIMO capacity is not significantly affected by it in that case [28,30,31]. A possible explanation for this is that the channel matrix, which includes the effect of mutual coupling, is known to the receiver and taken into account in the processing, and, hence, the mutual coupling is implicitly compensated for. Consequently, most of our results for $d > \lambda/2$ will hold true even if the effect of mutual coupling is taken into account.

VI. MIMO CAPACITY OF WAVEGUIDE CHANNELS⁴

The case of an ideal waveguide MIMO channel (i.e. lossless uniform waveguide) is especially interesting because the relationship between information theory and electromagnetics manifests itself in the clearest form in this scenario. We further consider such a waveguide unless otherwise indicated.

Arbitrary electromagnetic field inside of a waveguide can be presented as a linear combination of the modes [15,16],

$$\begin{aligned} \mathbf{e}(x, y, z) &= \sum_n \alpha_n \mathbf{E}_n(x, y) e^{-jk_{zn}z} \\ \mathbf{h}(x, y, z) &= \sum_n \beta_n \mathbf{H}_n(x, y) e^{-jk_{zn}z} \end{aligned} \quad (27)$$

where $\mathbf{E}_n(x, y)$ and $\mathbf{H}_n(x, y)$ are the normalized modal functions of the electric and magnetic fields, α_n and β_n are the expansion coefficients (mode amplitudes), k_{zn} is the axial component of the wave vector, and n is the (composite) mode index. The modal functions $\mathbf{E}_n(x, y)$ and $\mathbf{H}_n(x, y)$ give the field variation in the transverse directions (x, y) and the variation along the axial direction (z) is given by $e^{-jk_{zn}z}$. While particular form of the modal functions depend on the guide cross-section and may be difficult to find in explicit form (unless some symmetry is present), an important general property of the modal functions of a lossless cylindrical waveguide is their orthogonality in the following sense [16],

$$\begin{aligned} \iint_S \mathbf{E}_n \mathbf{E}_m dx dy &= \delta_{mn} \\ \iint_S \mathbf{H}_n \mathbf{H}_m dx dy &= \delta_{mn} \\ \iint_S \mathbf{E}_n \mathbf{H}_m dx dy &= 0 \end{aligned} \quad (28)$$

where the integrals are over the guide cross-sectional area S , $\delta_{mn} = 1$ if $m = n$ and 0 otherwise. For given frequency, there exist a finite number of propagating modes and all the other modes are evanescent, i.e. they decay exponentially with z .

⁴ This section is based on [12],[14].

Eq. 27 and 28 immediately suggest the transmission strategy for a waveguide channel, which is to use all the eigenmodes (or simply modes) as independent sub-channels since they are orthogonal and it is well-known that the MIMO capacity is maximum for independent sub-channels. In this case, the maximum number of independent sub-channels equals to the number of modes and there is no loss in capacity if *all* the modes are used. For lossy and/or non-uniform waveguide, there exist some coupling between the modes [16] and, hence, the capacity is smaller (due to the power loss as well as to the mode coupling). Thus, the capacity of a lossless waveguide will provide an upper bound for a true capacity since some loss and non-uniformity is always inevitable. It should be noted that if the coupling results in the normalized mode correlation being less than approximately 0.5, the capacity decrease is not significant. We further assume that the waveguide is lossless and is matched at both ends. Fig. 7 shows the system block diagram. At the Tx end, all the possible modes are excited using any of the well-known techniques (i.e., eigenmode modulator) and at the Rx end the transverse electric field is measured on the waveguide cross-sectional area (proper spatial sampling may be used to reduce the number of field sensors) and is further correlated with the distribution functions of each mode (i.e., spatial correlation receiver). The signals at the correlator outputs are proportional to the corresponding transmitted signals since the modes are orthogonal and, hence, there is no cross-coupling between different Tx signals. Thus, the channel matrix (i.e., Tx end-Rx end-correlator outputs) for this system can be expressed using the modal functions (for simplicity, we use only the E-field – the H-field can be used in the same way) as follows [14]

$$G_{ij} = \iint_S \mathbf{E}_i(x, y) \mathbf{E}_j(x, y) dx dy \quad (29)$$

and, for a uniform lossless waveguide, $\mathbf{G} = \mathbf{I}_N$, where \mathbf{I}_N is $N \times N$ identity matrix, and N is the number of modes. Clearly, the capacity achieves its maximum (2) in this case. Knowing the number of modes N , the maximum MIMO capacity can easily be evaluated. The maximum capacity (we call it further simply “capacity”) of the present MIMO architecture described above does not vary along the waveguide length

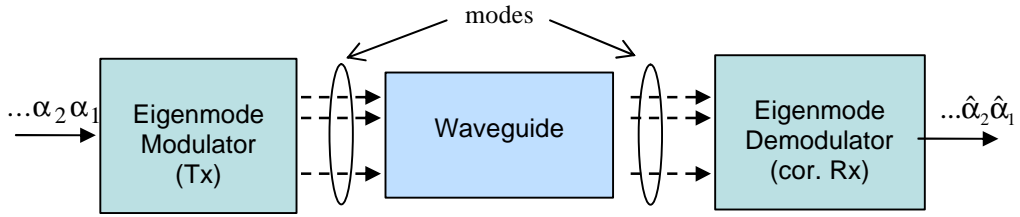


Fig. 7 MIMO system architecture for a waveguide channel. (© 2005 IEEE)

and it increases with the number of modes, as one would intuitively expect. If not all the available modes are used, the capacity decreases accordingly. The capacity may also decrease if the Rx antennas measure the field at some specific points rather than the field distribution along the cross-sectional area (since the mode orthogonality cannot be efficiently used in this case). In order to evaluate the maximum capacity, we further evaluate the number of modes.

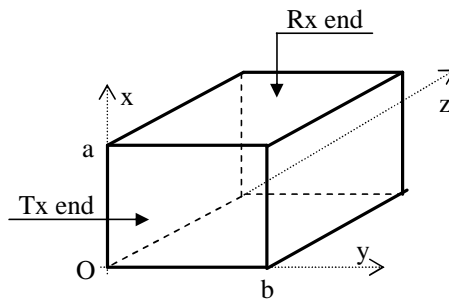


Fig. 8 Rectangular waveguide geometry. (© 2005 IEEE)

A. Rectangular Waveguide Capacity

Let us consider first a rectangular waveguide located along OZ axis (see Fig. 8). The field distribution at the XY plane (cross-section of the waveguide) for *E* and *H* modes is given by well-known expressions [16] and the variation along the OZ axis is given by $e^{-jk_z z}$, where *j* is imaginary unit, and k_z is the longitudinal component of the wavenumber:

$$k_z = \sqrt{\left(\frac{\omega}{c_0}\right)^2 - \gamma_{mn}^2}, \quad \gamma_{mn}^2 = \left(\frac{\pi m}{a}\right)^2 + \left(\frac{\pi n}{b}\right)^2, \tag{30}$$

where ω is the radial frequency, c_0 is the speed of light, and m and n designate the mode (note that E and H modes with the same (m,n) pair have the same γ_{mn}). The sign of k_z is chosen in such a way that the field propagates along OZ axis (i.e., from the Tx end to the Rx end). The case of $\gamma_{mn} > \omega/c$ corresponds to the evanescent field, which decays exponentially with z and is negligible at few wavelength from the source [16]. Assuming that the Rx end is located is far enough from the Tx end (i.e., at least few wavelengths), we neglect the evanescent field. Hence, the maximum value of γ_{mn} is $\gamma_{mn,\max} = \omega/c$. This limits the number of modes that exist in the waveguide at given frequency ω . All the modes must satisfy the following inequality, which follows from (30):

$$\left(\frac{m}{a'}\right)^2 + \left(\frac{n}{b'}\right)^2 \leq 4, \quad (31)$$

where $a' = a/\lambda$, $b' = b/\lambda$ and λ is the free-space wavelength; and $m,n=1,2,\dots$ for E mode and $m,n=0,1,\dots$, $m+n \neq 0$ for H mode. Using a numerical procedure and (31), the number of modes N can be easily evaluated. A closed-form approximate expression can be obtained for large a' and b' by observing that (31) is, in fact, an equation of ellipse in terms of (m,n) and all the allowed (m,n) pairs are located within the ellipse. Hence, the number of modes is given approximately by the ratio of areas:

$$N \approx 2 \frac{S_e/4}{S_0} = \frac{2\pi ab}{\lambda^2} = \frac{2\pi S_w}{\lambda^2}, \quad (32)$$

where $S_e = 4\pi a'b'$ is the ellipse area, $S_0 = 1$ is the area around each (m,n) pair, $S_w = ab$ is the waveguide cross-sectional area, the factor $1/4$ is due to the fact that only nonnegative m and n are considered, and the factor 2 is due to the contributions of both E and H modes. As (32) demonstrates, the number of modes is determined by the ratio of the waveguide cross-section area ab to the wavelength squared. As we will see later on, this is true for a circular waveguide as well. Hence, one may conjecture that this is true for a waveguide of arbitrary cross-section as well. This conjecture seems to be consistent with the spatial sampling argument (2-D sampling must be considered in this case). In fact, (32) gives the number of degrees of freedom the rectangular waveguide is able to support and which can be used for MIMO

communication. Fig. 9 compares the exact number of modes computed numerically using (31) and the approximate number (32). As one may see, (32) is quite accurate when a and b are greater than approximately a wavelength. Note that the number of modes has a step-like behavior with a/λ , which is consistent with (31). Using (2) and (32), the maximum capacity of the rectangular waveguide channel can be easily evaluated.

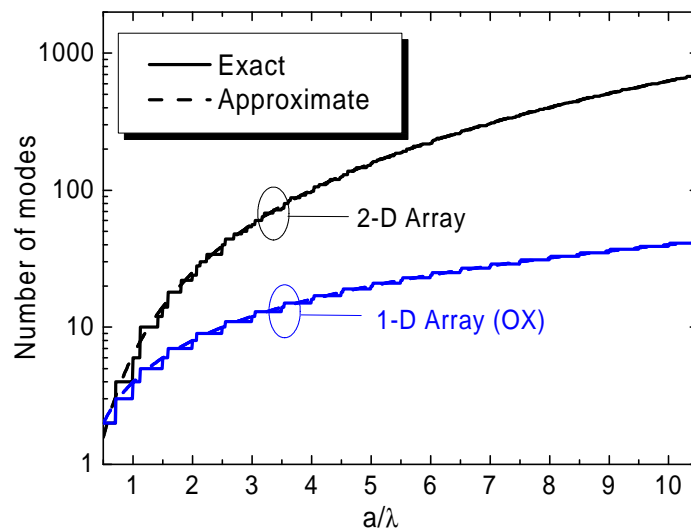


Fig. 9. Number of modes in a rectangular waveguide for $a=b$. (© 2005 IEEE)

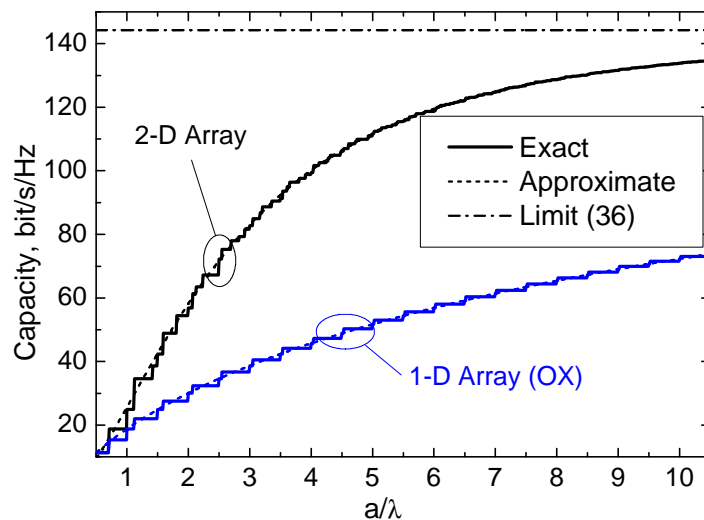


Fig. 10. MIMO capacity in a rectangular waveguide for $a=b$ and SNR=20 dB. (© 2005 IEEE)

The analysis above assumes that the vector E -field (including both E_x and E_y components) is measured on the entire cross-sectional area (or at a sufficient number of points to recover it using the sampling expansion). However, it may happen in practice that only one of the components is measured, or that the field is measured only along OX (or OY) axis. Apparently, it should lead to the decrease of the available modes.

To analyse this in details, let us assume that the E -field (both components) is measured along the OX axis only, which corresponds to a 1-D antenna array located along the OX axis. Due to this limitation, one can compute the correlations at the Rx using the integration over the OX axis only since the field distribution along OY axis is not known. Hence, we need to find the modes that are orthogonal in the following sense:

$$I = \int_0^a \mathbf{E}_\mu \mathbf{E}_\nu dx = c\delta_{\mu\nu}, \quad (33)$$

where μ and ν are composite mode indices. In this case, one finds that two different E -modes $E_{m_1n_1}$ and $E_{m_2n_2}$ are orthogonal provided that $m_1 \neq m_2$; if these modes have the same m index, they are not orthogonal. The same is true about two H -modes and about one E -mode and one H -mode. This results in a substantial reduction of the number of orthogonal modes since, in the general case, two E -modes are orthogonal if at least one of the indices is different, i.e. if $m_1 \neq m_2$ or $n_1 \neq n_2$. Surprisingly, if one measures only E_x component in this case, the modes are still orthogonal provided that $m_1 \neq m_2$. Hence, if the receive antenna array is located along OX axis, there is no need to measure E_y component – it does not provide any additional degrees of freedom, which can be used for MIMO communications (recall that only orthogonal modes can be used). The number of orthogonal modes can be evaluated using (31):

$$N_x \approx 4a/\lambda, \quad (34)$$

This corresponds to $2a/\lambda$ degrees of freedom for each (E and H) field. Note that this result is similar to that obtained using the spatial sampling argument, i.e., independent field samples (which are, in fact, the degrees of freedom) are located at $\lambda/2$.

A similar argument holds true when the receive array is located along OY axis. In this case two modes are orthogonal provided that $n_1 \neq n_2$ and there is also no need to measure the E_x component. The number of orthogonal modes is approximately

$$N_y \approx 4b/\lambda, \quad (35)$$

Fig. 10 shows the MIMO capacity of a rectangular waveguide of the same geometry as in Fig. 8 for SNR $\rho = 20$ dB. Note that the capacity saturates as a/λ increases. This is because (2) saturates as well as N increases:

$$\lim_{N \rightarrow \infty} C = \rho / \ln 2 \quad (36)$$

C in (2) can be expanded as

$$C = \frac{\rho}{\ln 2} \sum_{i=0}^{\infty} \frac{(-1)^i}{i+1} \left(\frac{\rho}{N} \right)^i \quad (37)$$

For large N, i.e. for small ρ/N , this series converges very fast and it can be approximated by first two terms:

$$C \approx \frac{\rho}{\ln 2} \left(1 - \frac{\rho}{2N} \right) \quad (38)$$

The capacity does not change substantially when the contribution of the 2nd term is small:

$$\frac{\rho}{2N} \ll 1 \Rightarrow N > N_{\max} \approx \rho \quad (39)$$

N_{\max} is the maximum “reasonable” number of antennas (modes) for given SNR (or vice versa): if N increases above this number, the capacity does not increase significantly. It may be considered as a practical limit (since further increase in capacity is very small and it requires for very large increase in complexity). Using (32) and (34), one finds the maximum “reasonable” size of the waveguide for the case of 2-D and 1-D arrays correspondingly:

$$\frac{a_{\max}}{\lambda} \approx \sqrt{\frac{\rho}{2\pi}} \text{ (2-D array), } \frac{a_{\max}}{\lambda} \approx \frac{\rho}{4} \text{ (1-D OX array),} \quad (40)$$

Note that Fig. 10 shows, in fact, the fundamental limit of the waveguide capacity, which is imposed jointly by the laws of information theory and electromagnetics.

B. Rectangular Cavity Capacity

The analysis of MIMO capacity in cavities is different from that in waveguides in one important aspect. Namely, the modes of a cavity exist only for some finite discrete set of frequencies (recall that, as in the case of waveguide, we consider a lossless cavity). Hence, there may be no modes for an arbitrary frequency. To avoid this problem, we evaluate the number of modes for a given bandwidth, $f \in [f_0, f_0 + \Delta f]$, starting at f_0 . For a rectangular cavity, the wave vector must satisfy [16]:

$$k^2 = \left(\frac{\pi m}{a}\right)^2 + \left(\frac{\pi n}{b}\right)^2 + \left(\frac{\pi p}{c}\right)^2 = \left(\frac{\omega}{c_0}\right)^2, \quad (41)$$

where c is the waveguide length (along OZ axis in Fig. 8), and p is a non-negative integer; $m, n = 1, 2, 3, \dots$, $p = 0, 1, 2, \dots$ for E-modes, and $m, n = 0, 2, 3, \dots$, $p = 1, 2, \dots$ for H-modes ($m = n = 0$ is not allowed). Noting that (41) is a equation of a sphere in terms of $(\pi m/a, \pi n/b, \pi p/c)$, the number of modes with $k \in [k_0, k_0 + \Delta k]$ can be found as the number of (m, n, p) points between two spheres with radiuses of k_0 and $k_0 + \Delta k$ correspondingly. Using the ratio of areas approach described above, the number of modes is approximately:

$$N_c \approx 2 \frac{V_e/8}{V_0} = \frac{8\pi V_c \Delta f}{\lambda^3 f_0}, \quad (42)$$

where $V_e = 4\pi k^2 \Delta k$ is the volume between the two spheres, $V_0 = \pi^3/V_c$ is the volume around each (m, n, p) point, $V_c = abc$ is the cavity volume; factor 2 is due to two types of modes, and factor 1/8 is due to the fact that only nonnegative values of (m, n, p) are allowed. An important conclusion from (42) is that the number of modes is determined by the cavity volume expressed in terms of wavelength and by the normalized bandwidth. Detailed analysis shows that (42) is accurate for large a , b , and c , and if $c/\lambda < f_0/4\Delta f$.

It should be noted that the mode orthogonality for cavities is expressed through the volume integral (over the entire waveguide volume),

$$\iiint_{V_c} \mathbf{E}_\mu \mathbf{E}_\nu dV = c \delta_{\mu\nu}, \quad (43)$$

and, hence, all the modes are orthogonal provided that the field is measured along all 3 dimensions, which, in turn, means that a 3-D arrays must be used, which may not be feasible in practice. If only 2-D arrays are used, then the mode orthogonality is expressed as for a waveguide, i.e. (28), and, consequently, only those modes are orthogonal that have different (m,n) indices. The use of a 2-D array results in significant reductions of the number of modes for large c , as Fig. 11 demonstrates. Note that for small c , there is no loss in the number of orthogonal modes. This is because different p correspond in this case to different (m,n) pairs (this can also be seen from (41)). However, as c increases, different p may include the same (m,n) pairs, which results in the number loss if a 2-D array is used. In fact, the 2-D case with large c is the same as the waveguide case (with the same cross-sectional area), as it should be. The value of c for which the cavity has the same number of orthogonal modes as the corresponding waveguide can be found from the following equality:

$$N_c \approx N_w \Rightarrow \frac{c_t}{\lambda} = \frac{f_0}{4\Delta f}, \quad (44)$$

Hence, if 2-D antenna arrays are used and $c \geq c_t$, the waveguide model provides approximately the same results as the cavity model does, i.e. the cross-section has the major impact on the capacity, while the effect of cavity length is negligible. The waveguide model should be used to evaluate the number of orthogonal modes (and capacity) in this case because it is more simple to deal with. For example, a long corridor can be modelled as a waveguide rather than cavity (despite of the fact that it is closed and looks like a cavity). Fig. 12 shows the MIMO capacity in the cavity. While the capacity of a 2-D array system saturates like the waveguide capacity, which is limited by a and b , the capacity of a 3-D system is larger and saturates at the value given by (36). It should be noted that (36) is the capacity limit due to the

information theory laws, and (32), (34), (35), and (44) are the capacity limits due to the laws of electromagnetism (i.e., limited due to the number of degrees of freedom of the EM field).

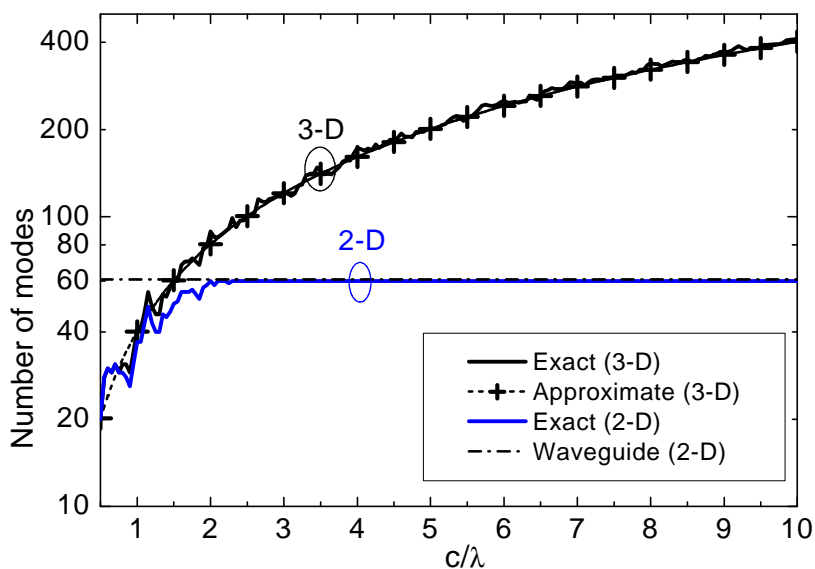


Fig. 11. Number of orthogonal modes in a rectangular cavity for $a = 4\lambda$, $b = 2\lambda$ and $\Delta f / f_0 = 0.2$.

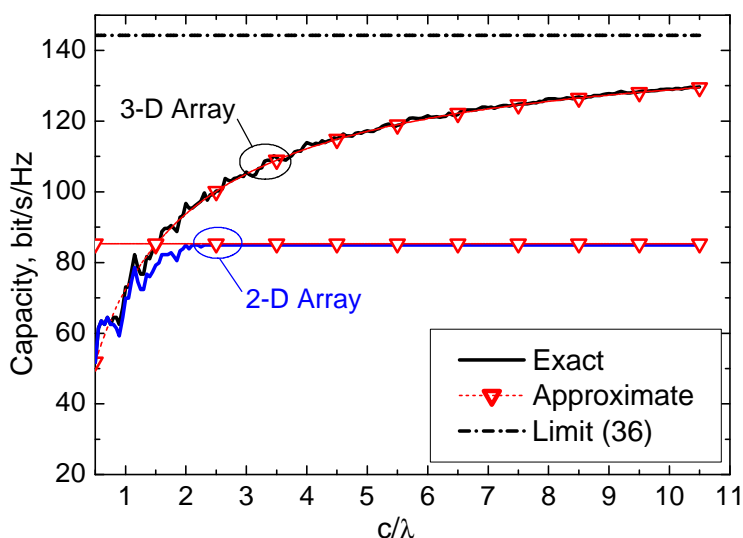


Fig. 12. Capacity in a rectangular cavity for $a = 4\lambda$, $b = 2\lambda$ and $\Delta f / f_0 = 0.2$.

VII. SPATIAL CAPACITY OF WAVEGUIDE CHANNELS

While the maximum in (9) is difficult to find in general since one of the constraints is a partial differential equation with an arbitrary boundary condition, it can be found in an explicit closed form for some waveguide channels. Consider, for example, a lossless uniform waveguide. Using (27), we conclude that (i) the optimizations over $p(\mathbf{x})$ and \mathbf{E} can be carried out separately (since they are independent of each other) and (ii) the optimization over \mathbf{E} is equivalent to optimization over α_n (since the expansion coefficients determine the field uniquely). When the Tx does not know the channel, \mathbf{x} is i.i.d. complex Gaussian because (i) the channel is AWGN and Gaussian distribution maximizes the entropy, and (ii) the lack of the channel knowledge at the Tx forces the covariance of \mathbf{x} to be the identity matrix, i.e. no “preferred direction” in the eigenspace (see [2] for more details), $p(\mathbf{x}) = CN(\mathbf{0}, P_T / n_T \mathbf{I})$, and the capacity is given by (1) [2]. Further optimization of (9) over α_n subject to $\text{tr}(\mathbf{G}\mathbf{G}^+) = n_T$ results in $\alpha_n = 1$ and the channel correlation matrix $\mathbf{R} = \mathbf{G}\mathbf{G}^+ = \mathbf{I}$, i.e. all the modes are independent and carry the same power. Hence, all the capacity results above give, in fact, the spatial capacity, i.e. the maximum capacity limited by the laws of electromagnetics. One may say that the spatial capacity concept encompasses the limits in information transmission due to both the information theory and electromagnetics.

When there is mode coupling and, additionally, different modes experience different attenuation (e.g. lossy waveguide with different γ_i), one has to consider generic correlation matrix $\mathbf{R} \neq \mathbf{I}$. The optimum power allocation in the mode eigenspace can be found by applying the water-filling solution [14] to the eigenvalues λ_i of \mathbf{R} ,

$$\alpha_i = \sqrt{\left[v - \frac{\lambda_i^{-1} n_T}{\rho} \right]_+} \quad (45)$$

where $(x)_+ = x$ if $x \geq 0$ and 0 otherwise, and v is chosen to satisfy the power constraint,

$$\sum_i \left[\nu - \frac{\lambda_i^{-1} n_T}{\rho} \right]_+ = n_T \quad (46)$$

α_i is large for large eigenvalues, and small or even zero (i.e. no transmission on the eigenmode) for small λ_i . Without loss of generality, we further assume that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{n_T}$. Two important conclusions follow from (45), (46) [14]. In the large SNR mode, $\rho \gg n_T / \lambda_{n_T}$, all the eigenmodes carry approximately the same power ($\alpha_i \approx 1$). We also note that the same solution applies when all the eigenvalues are equal, $\lambda_i = \lambda \rightarrow \alpha_i = 1 \forall i$, regardless of the SNR. In the small SNR regime, $\rho \leq 1/\lambda_2 - 1/\lambda_1$, all the power is allocated to the largest eigenmode, $\alpha_1 = n_T$, $\alpha_i = 0 \forall i \neq 1$. From this, we conclude that the traditional transmission strategy (i.e. using a single dominant mode only) is optimum in the small SNR regime only. For large SNR, the best strategy is to use all the modes. In all the other cases, between these two extremes, several dominant modes should be used, the exact number being determined by the available SNR and by the eigenvalues λ_i .

Remarkably, this insight asks a joint application of electromagnetic and information-theoretic techniques. It is expected that an efficient design of eigenmode modulators and demodulators (see Fig. 7) will also call for significant unification of this two presently disconnected areas of research.

VIII. ACKNOWLEDGEMENT

The authors would like to acknowledge numerous insightful discussions with C. Charalambous, A. Kouki, A. Molisch, I. E. Telatar, D. McNamara and also the colleagues within the European projects COST 273 and 284.

IX. REFERENCES

- [1] G.J. Foschini, M.J Gans: 'On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas', *Wireless Personal Communications*, vol. 6, No. 3, pp. 311-335, March 1998.
- [2] I.E. Telatar, "Capacity of Multi-Antenna Gaussian Channels," AT&T Bell Lab. Internal Tech. Memo., June 1995 (European Trans. Telecom., v.10, N.6, Dec.1999).
- [3] D. Gesbert et al, From Theory to Practice: An Overview of MIMO Space-Time Coded Wireless Systems, *IEEE Journal on Selected Areas in Communications*, v. 21, N. 3, pp. 281-302, Apr. 2003.
- [4] A. Paulraj, R. Nabar, D. Gore, *Introduction to Space-Time Wireless Communications*, Cambridge University Press, 2003.
- [5] A.M. Sayeed, Deconstructing Multiantenna Fading Channels, *IEEE Trans. Signal Processing*, v. 50, N. 10, pp. 2563-2579, Oct. 2002.
- [6] H. Boleskei et al, On the Capacity of OFDM-Based Spatial Multiplexing Systems, *IEEE Trans. Communications*, v. 50, N. 2, pp. 225-234, Feb. 2002.
- [7] D. Chizhik, G.J. Foschini, R.A. Valenzuela, 'Capacities of multi-element transmit and receive antennas: Correlations and keyholes', *Electronics Letters*, vol. 36, No. 13, pp.1099-1100, 22nd June 2000.
- [8] S. Loyka, G. Tsoulos, Estimating MIMO System Performance Using the Correlation Matrix Approach, *IEEE Communication Letters*, v. 6, N. 1, pp. 19-21, Jan. 2002.
- [9] S.L. Loyka, J.R. Mosig, Spatial Channel Properties and Spectral Efficiency of BLAST Architecture, AP2000 Millennium Conference on Antennas & Propagation, Davos, Switzerland, 9-14 April, 2000.
- [10] S. Loyka, MIMO Channel Capacity: Electromagnetic Wave Perspective, 27th General Assembly of the International Union of Radio Science, Maastricht, The Netherlands, Aug. 17-24, 2002.

- [11] S.L. Loyka, On MIMO Channel Capacity, Spatial Sampling and the Laws of Electromagnetism, the 3rd IASTED International Conference on Wireless and Optical Communications (WOC 2003), July 14-16, 2003, Banff, Alberta, Canada, pp. 132-137.
- [12] S.L. Loyka, Multi-Antenna Capacities of Waveguide and Cavity Channels, IEEE CCECE'03, Montreal, May 2003.
- [13] S. Loyka, Information Theory and Electromagnetism: Are They Related? (Invited), the Joint COST 273/284 Workshop on Antennas and Related System Aspects in Wireless Communications, Chalmers University of Technology, Gothenburg, Sweden, June 2004.
- [14] S. L. Loyka, Multi-Antenna Capacities of Waveguide and Cavity Channels, IEEE Trans. Vehicular Technology, v. 54, N. 3, May 2005.
- [15] E.D. Rothwell, M.J. Cloud, Electromagnetics, CRC Press, Boca Raton, 2001.
- [16] R.E. Collin, Field Theory of Guided Waves, IEEE Press, New York, 1991.
- [17] J.D. Gibson (Ed.), The Communications Handbook, CRC Press, Boca Raton, 2002.
- [18] S.R. Saunders, Antennas and Propagation for Wireless Communication Systems, Wiley, Chichester, 1999.
- [19] A.J. Jerry, The Shannon Sampling Theorem – Its Various Extensions and Applications: A Tutorial Review, Proc. of IEEE, v. 65, N. 11, pp. 1565-1596, Nov. 1977.
- [20] Y.I. Khurgin, V.P. Yakovlev, Progress in the Soviet Union on the Theory and Applications of Bandlimited Functions, Proceedings of the IEEE, v. 65, N.7, pp. 1005-1029, July 1977.
- [21] D.P. Petersen, D. Middleton, Sampling and Reconstruction of Wave-Number-Limited Functions in N-Dimensional Euclidean Spaces, Information and Control, v. 5, pp. 279-323, 1962
- [22] Jakes, W.C. Jr.: 'Microwave Mobile Communications', John Wiley and Sons, New York, 1974.
- [23] S.L. Loyka, Channel Capacity of MIMO Architecture Using the Exponential Correlation Matrix, IEEE Communication Letters, v.5, N. 9, pp. 369 –371, Sep 2001.
- [24] T.S. Pollock et al, Antenna Saturation Effects on Dense Array MIMO Capacity, IEEE ICASSP 2003.

- [25] J.W. Wallace, M.A. Jensen, Intrinsic Capacity of the MIMO Wireless Channel, IEEE VTC (Fall), 24-28 Sept. 2002.
- [26] R.C. Johnson, Antenna Engineering Handbook, McGraw Hill, New York, 1993.
- [27] D.H. Johnson, D.E. Dudgeon, Array Signal Processing, Prentice Hall, Upper Saddle River, 1993.
- [28] J.W. Wallace, M.A. Jensen, Mutual Coupling in MIMO Wireless Systems: A Rigorous Network Theory Analysis, IEEE Trans. Wireless Commun., v. 3, N. 4, pp. 1317-1325, July 2004.
- [29] V. Jungnickel, V. Pohl, C. von Helmolt, Capacity of MIMO Systems With Closely Spaced Antennas, IEEE Communications Letters, v.7, N. 8, pp. 361-363, Aug. 2003.
- [30] R. Janaswamy, Effect of Element Mutual Coupling on the Capacity of Fixed Length Linear Arrays, IEEE Antennas and Wireless Propagation Letters, v. 1, pp. 157-160, 2002.
- [31] P.S. Kildal, K. Rosengren, Correlation and Capacity of MIMO Systems and Mutual Coupling, Radiation Efficiency, and Diversity Gain of Their Antennas: Simulations and Measurements in a Reverberation Chamber, IEEE Communications Magazine, v. 42, N.12, pp. 104-112, Dec. 2004.
- [32] M.K. Ozdemir, H. Arslan, E. Arvas, Mutual Coupling Effect in Multi-Antenna Wireless Communication Systems, IEEE Globecom, pp. 829-833, 2003.
- [33] F. Gori, Sampling in Optics, in R.J. Marks II (Ed.), Advanced Topics in Shannon Sampling and Interpolation Theory, Springer-Verlag, New York, 1993.
- [34] T.B. Hansen, A.D. Yaghjian, Plane-Wave Theory of Time-Domain Fields, IEEE Press, New York, 1999.
- [35] A.I. Zayed, Advances in Shannon's Sampling Theory, CRC Press, Boca Raton, 1993.