Joint COST 273/284 Workshop on Antennas and Related System Aspects in Wireless Communications, Chalmers University of Technology, Gothenburg, Sweden, June 2004

INFORMATION THEORY AND ELECTROMAGNETISM: ARE THEY RELATED?

Sergey Loyka

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

Abstract— *In this paper, we study the limitations imposed by the laws of electromagnetism on achievable MIMO channel capacity in its general form. Our approach is a three-fold one. First, we use the channel correlation argument to demonstrate that the minimum antenna spacing under any scattering conditions is at least half a wavelength. Secondly, using a plane-wave spectrum expansion of a generic electromagnetic wave combined with Nyquist sampling theorem in the spatial domain, we show that the laws of electromagnetism limit the minimum antenna spacing to half a wavelength,* $\lambda/2$*, (in the case of 1-D antenna apertures) only asymptotically, when the number of antennas* $n \to \infty$*. For a finite number of antennas, this limit is slightly less than* $\lambda/2$*. The number of antennas and, consequently, the MIMO capacity is limited for a given aperture size. This is a scenario-independent limit. Finally, we study the MIMO capacity of waveguide and cavity channels. The rationale for this is three-fold: (i) waveguide / cavity models can be used to model corridors, tunnels and other confined space channels, (ii) this is a canonical problem; its analysis allows to develop appropriate techniques, which can be further used for more complex problems, (iii) it allows to shed light on the relation between information theory and electromagnetism and, in particular, to establish the limits imposed by the laws of electromagnetism on achievable channel capacity.*

I. INTRODUCTION

It is well recognized that the wireless propagation channel has a profound impact on MIMO system performance. 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 [3]. Many practically-important scenarios have been studied and some design guidelines have been proposed as well.

In the present paper, we analyze the effect of propagation channel from a completely different perspective. 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 [5]. 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 this paper, we try to answer the second question. In other words, do the laws of electromagnetism impose any limitations on the achievable channel capacity? 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 two-fold one. 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 due to 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 (for linear antenna arrays) only asymptotically, when the number of antennas $n \to \infty$. 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, this limits the number of antennas and the MIMO capacity for a given aperture size. It should be emphasized that this is a scenario independent limit. It follows directly from Maxwell equations and is valid in any scenario.

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,

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

where N is the numbers of transmit/receive antennas, ρ is the average signal-to-noise ratio, **I** is n×n identity matrix, **G** is the normalized channel matrix (the entries are complex channel gains from each Tx to each Rx antenna), $\mathrm{tr}[\mathbf{GG}^{+}] = N$, which is considered to be frequency independent over the signal bandwidth, and "$^{+}$" denotes transpose conjugate. In an ideal case of uncorrelated full-rank channel (1) reduces to

$$C = N \log_2 (1 + \rho / N), \qquad (2)$$

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

III. THE LAWS OF ELECTROMAGNETISM

It follows from (1) that the MIMO channel capacity crucially depends the propagation channel **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 behaviour of **G** in different scenarios. Hence, we outline the laws of electromagnetism in a MIMO system perspective. In their most general form, they are expressed as Maxwell equations with charge and current densities as the field sources [5]. Appropriate boundary conditions must be applied in order to solve them. We are interested in application of Maxwell equations to find the channel matrix **G** in (1). Since the Rx antennas are located at some distance from Tx antennas (not at the same points in space), we are interested in source-free region of space, (i.e., electromagnetic waves). In this case, Maxwell equations simplifies to the system of two decoupled wave equations [5]:

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

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. Of course, only two of them survive in free space ("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,

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

(3) can be expressed as [5]

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

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) does not require for any significantly-restrictive assumptions. The source-free region assumption seems to be quite natural (i.e., Tx and Rx antennas are located in different points in space) and the narrowband assumption is simplifying but not restrictive since (5) can be solved for *any* frequency and, further, the capacity can be evaluated using well-known techniques.

Unfortunately, the link between (5) and the channel matrix $\mathbf{G}$ is implicit. A convenient way to study this link is to use the space domain Fourier transform, i.e. the plane-wave spectrum expansion,

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

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

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

Hence, $|\mathbf{k}| = \omega/c$ and the electromagnetic filed 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 [6],

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

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, $P_x = \left\langle \mathbf{x}\mathbf{x}^+ \right\rangle \le P_t$. Under some conditions, this results in (1). In order to study the impact of the electromagnetism laws on the channel capacity, we definite the *spatial capacity* S as the maximum mutual information between the Tx vector on one side and the pair of the Rx vector and the channel (assuming perfect CSI at the Rx) on the other, the maximum being taken over both the Tx vector and EM field distributions,

$$S = \max_{p(\mathbf{x}),\,\mathbf{E}}\left\{I\left(\mathbf{x},\{\mathbf{y},\mathbf{G}(\mathbf{E})\}\right)\right\},\ \text{const.: } \left\langle \mathbf{x}^+\mathbf{x}\right\rangle \le P_T,$$
$$\nabla^2\mathbf{E} - \frac{1}{c^2}\frac{\partial^2\mathbf{E}}{\partial t^2} = 0,\ \mathbf{E}\text{=}\mathbf{E}_0 \forall\{\mathbf{r},t\}\in B \tag{9}$$

where, to be specific, we assumed that the electric field $\mathbf{E}$ is used to transmit data ($\mathbf{H}$ field can be used in the same way), B is the boundary condition (due to the scattering environment), and the last constraint is due to the boundary condition. 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. This maximum is difficult to find in general since one of the constraint is a partial differential equation with an arbitrary boundary condition.

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, the capacity is maximized by changing $\mathbf{G}$ (within some limits), for example, by appropriate positioning of antennas,

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

where the constraint $\mathcal{S}(\text{Maxwell})$ is due to the Maxwell (wave) equations. Unfortunately, the explicit form of this constraint is not known. Additional constraints may be included (due to a limited aperture, for example). Note that this definition will give a capacity, which is, in general, less than that in the first definition.

Using the analogy with the channel capacity definition, one can call this maximum (if it exists) "capacity of a given space" or "spatial capacity" (since we have to vary 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 which have an impact on it? Using the ray tracing (geometrical optics) arguments and the recent result on the MIMO capacity, it can be further demonstrated that there exists an optimal distribution of scatterers and Tx/Rx antennas that provides the maximum possible capacity in a given region of space. Hence, the MIMO capacity per unit space volume can be defined in a fashion similar to the traditional definition of the channel capacity per unit bandwidth.

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), 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 **G** is very implicit since a lot depends on Tx and Rx antenna designs and many other details.

We further 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 ("ideal" scattering) in the entire space (including the space region considered) and which do not absorb EM field. 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 (this 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 [4]. This minimum distance is about half a wavelength. It should be noted that the model in [4] is a two-dimensional (2D) one. However, it can be applied to both orthogonal planes and, due to the symmetry of the problem (no preferred direction), the same result should hold in 3D as well. We note that, under the assumptions above, the angle-of-arrival (AOA) of multipath components is uniformly distributed over $[0, 2\pi]$ in both planes. Thus, the model above can be applied and the minimum distance is about half a wavelength. Due to the assumption of uniform scattering media, all the antennas experience the same multipath environment.

When we increase the number of antennas the capacity at first increases. But at some point, due to aperture limitation, we have to decrease the distance between adjacent antennas to accommodate new antennas within the given aperture. When the adjacent antenna spacing decreases, the capacity increase slows down and finally, when the antenna spacing is less than the minimum distance, the capacity begins to decrease. Hence, there is an optimal number of antennas, for which the capacity is maximum. An argument similar to the present one has already been presented earlier [8]. However, the optimal number of antennas has not been evaluated. Using the model in [4], which results in the minimum distance be equal to approximately half a wavelength, 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):

$$N_{opt} \approx 2L/\lambda + 1 \tag{11}$$

where λ is the wavelength. Similar expressions can be obtained for 2-D and 3-D apertures as well. This is consistent with the diversity combining analysis, where the minimum distance is about half a wavelength as well [10], and with an earlier speculation in [1].

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 refereeing 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 receive antenna properties, one may further compute the signal at the antenna output and, hence, the channel matrix **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 earlier, we assume that the receive antennas are ideal field sensors (with no size, no coupling between antennas 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 filed itself. In general, different Tx antennas will produce different plane-wave spectra around the Rx antennas and, hence, the wave equation is:

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

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: $\omega=const$, 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:

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

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 **r**. The integration in (13) is performed on a hypersurface $|\mathbf{k}|=\omega/c$. As we show below, it 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

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

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

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

where $k_{\max}=\omega/c$ due to (14). We ignored the evanescent waves with $|k|>k_{\max}$ because they decay exponentially and can be ignored at distances more than few λ from the source [5]. 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) [6]. In our case, the Nyquist rate is $2k_{\max}$ and the sampling interval is

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

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 $x=0,\pm\Delta x_{\min},\pm 2\Delta x_{\min},...$. This means that by locating the field sensors at sampling points, which are separated by $\Delta x_{\min}$, we are able to recover all the information transmitted by electromagnetic waves to the receiver. Hence, channel capacity is not altered. This means, in turn, that the minimum spacing between antennas is half a wavelength:

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

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 [7] (when evaluating correlation, we have to use it to make 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 $\Delta x_{\min}$) 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 [11]. If one uses a 2-D antenna aperture (i.e. 2-D sampling), the sampling interval is

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

and in the case of 3-D aperture,

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

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 theorem 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 note is in order on using the sampling theorem to find the minimum antenna spacing. The sampling theorem guarantees that the original band-limited function can be recovered from its samples provided that the infinite number of samples is used (band-limited function cannot be time limited!). Hence, the half wavelength limit, as derived using the sampling theorem, holds true only asymptotically, when $n\rightarrow\infty$. When n is finite, the optimal number of antennas may be larger than that given by (11), i.e. the minimum spacing may be less than half a wavelength 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 [9]. In this case, one is able to recover almost all the information conveyed by the EM field to the antenna aperture (but not outside of the aperture). Hence, one may expect that the actual minimum antenna spacing is quite close to half a wavelength for a 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. Fig. 1 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 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.

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.

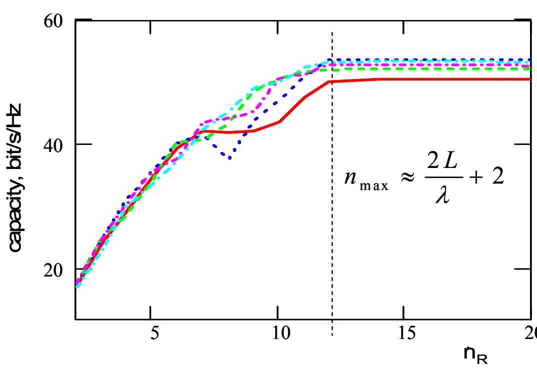

Fig. 1 MIMO channel capacity versus the number of antennas for $L=5\lambda$.

It should be noted that, in some cases, increasing n over N_{opt} in (11) may result in SNR increase due to antenna gain increase and, consequently, in logarithmic increase in capacity. However, this increase 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. Besides, 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). This limit is approximately 30% larger than the array gain at $d=\lambda/2$. Keeping in mind that the capacity depends logarithmically on SNR and, consequently, the antenna gain, we see that this increase in capacity is very small.

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].

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 $\Delta \ll 1$, $d_{\min} \approx \lambda/(2\Delta) \gg \lambda/2$ [4]. Hence, less antennas can be accommodated within given aperture and, consequently, the MIMO capacity is smaller for a given aperture size.

VI. MIMO Capacity of Waveguide Channels

The case of an ideal waveguide MIMO channel is especially interesting because the relationship between information theory and electromagnetics manifests itself in the most clear form.

The main idea for a waveguide channel is to use the eigenmodes (or simply modes) as independent sub-channels since they are orthogonal (if the waveguide is lossless and uniform) and it is well-known that the MIMO capacity is maximum for independent sub-channels. Since any field inside of the waveguide can be presented as a linear combination of the modes [5], 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 [5] and, hence, the capacity is smaller (due to the power loss as well as to the 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 sub-channel correlation less than approximately 0.5, the capacity decrease is not significant [13]. We further assume that the waveguide is lossless and is matched at both ends. In this case, the transverse electric fields for two different E modes, or two different H modes, or one E and one H mode are mutually orthogonal [5]

$$\iint_S \mathbf{E}_\mu \mathbf{E}_\nu dS = c\delta_{\mu\nu}, \tag{20}$$

where the integral is over the waveguide cross-sectional area S, μ and ν are composite mode indices, $\delta_{\mu\nu}$ is Kronecker delta, and c is a constant (which depends on the power transmitted in each mode). (20) immediately suggests the system architecture to achieve the maximum MIMO capacity using the modes: at the Tx end, all the possible modes are excited using any of the well-known techniques 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, see Fig. 2. 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 equivalent channel matrix (i.e., Tx end-Rx end-correlator outputs) is $\mathbf{H}=\mathbf{I}_N$ (recall that the waveguide is assumed to be matched and lossless), where $\mathbf{I}_N$ is NxN identity matrix, and the capacity achieves its maximum (2). 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

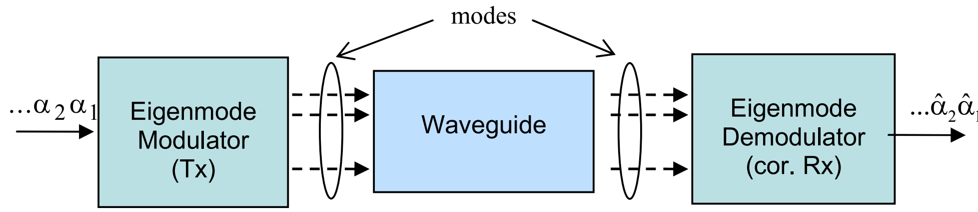

Fig. 2 MIMO system architecture for a waveguide channel.

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.

A. Rectangular Waveguide Capacity

Let us consider first a rectangular waveguide located along OZ axis (see Fig. 3). The field distribution at XY plane

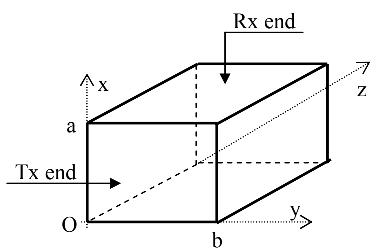

Fig. 3 Rectangular waveguide geometry.

(cross-section) for E and H modes is given by well-known expressions [5] 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:

$$k_z = \sqrt{\left(\frac{\omega}{c_0}\right)^2 - \gamma_{mn}^2}, \quad \gamma_{mn}^2 = \left(\frac{\pi m}{a}\right)^2 + \left(\frac{\pi n}{b}\right)^2, \tag{21}$$

where ω is the 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 filed 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 [5]. 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 (21):

$$\left(\frac{m}{a'}\right)^2 + \left(\frac{n}{b'}\right)^2 \le 4, \tag{22}$$

where $a' = a/\lambda$, $b' = b/\lambda$ and λ is the free-space wavelength; and $m,n = 1,2,...$ for E mode and $m,n = 0,1,...$, $m+n \neq 0$ for H mode. Using a numerical procedure and (22), the number of modes N can be easily evaluated. A closed-from approximate expression can be obtained for large a' and b' by observing that (22) 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:

$$N \approx 2\frac{S_e/4}{S_0} = \frac{2\pi ab}{\lambda^2} = \frac{2\pi S_w}{\lambda^2}, \tag{23}$$

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 ¼ 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 (23) 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 further, 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, (23) gives the number of degrees of freedom the rectangular waveguide is able to support and which can be further used for MIMO communication. Fig. 4 compares the exact number of modes computed numerically using (22) and the approximate number (23). As one may see, (23) is quite accurate when a and b are greater then approximately a wavelength. Note that the number of modes has a step-like behavior with a/λ, which is consistent with (22). Using (2) and (23), the maximum capacity of the rectangular waveguide channel can be easily evaluated.

The analysis above assumes that the E-field (including both E_x and E_y components) is measured on the entire

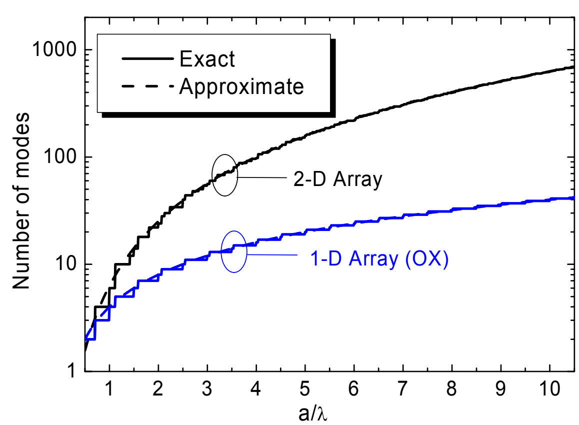

Fig. 4. Number of modes in a rectangular waveguide for *a=b*.

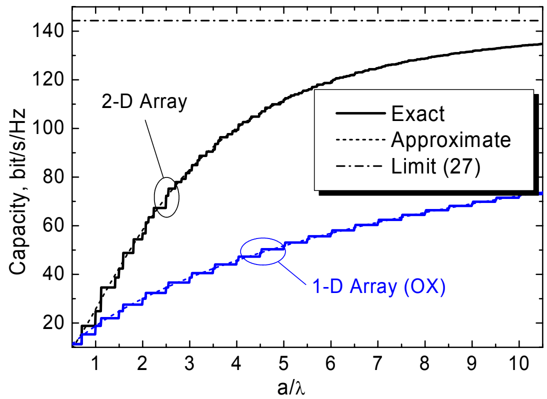

Fig. 5. MIMO capacity in a rectangular waveguide for *a=b* and SNR=20 dB.

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. This is analysed below in details.

Let us assume that the E-field (both components) is measured along the OX axis only (this corresponds to 1-D antenna array located along OX). Due to this limitation, one can compute the correlations at the Rx using the integration over 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:

$$I = \int_0^a \mathbf{E}_\mu \mathbf{E}_\nu dx = c\delta_{\mu\nu}, \quad (24)$$

In this case, one finds that two different E-modes $E_{m_1 n_1}$ and $E_{m_2 n_2}$ 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 $m_1 \neq m_2$ <u>*or*</u> $n_1 \neq n_2$. 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 measured 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 (22):

$$N_x \approx 4a/\lambda, \quad (25)$$

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 $\lambda/2$.

The 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

$$N_y \approx 4b/\lambda, \quad (26)$$

Fig. 5 shows the MIMO capacity of a rectangular waveguide (the same as in Fig. 3) for SNR $\rho = 20$ dB. Note that the capacity saturates as a/λ increases. This is because (2) saturates as well as N increases:

$$\lim_{N\to\infty} C = \rho/\ln 2 \quad (27)$$

C in (2) can be expanded as

$$C = \frac{\rho}{\ln 2}\sum_{i=0}^{\infty}\frac{(-1)^i}{i+1}\left(\frac{\rho}{N}\right)^i \quad (28)$$

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

$$C \approx \frac{\rho}{\ln 2}\left(1 - \frac{\rho}{2N}\right) \quad (29)$$

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

$$\frac{\rho}{2N} << 1 \Rightarrow N > N_{\max} \approx \rho \quad (30)$$

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 (23) and (25), one finds the maximum "reasonable" size of the waveguide for the case of 2-D and 1-D arrays correspondingly:

$$\frac{a_{\max}}{\lambda} \approx \sqrt{\frac{\rho}{2\pi}} \text{ (2-D array)}, \quad \frac{a_{\max}}{\lambda} \approx \frac{\rho}{4} \text{ (1-D OX array)}, \quad (31)$$

Note that Fig. 5 shows, in fact, the fundamental limit of the waveguide capacity, which is imposed jointly by the laws of information theory and electromagnetism.

B. Rectangular Cavity Capacity

The analysis of MIMO capacity in cavities is very 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 [5]:

$$k^2 = \left(\frac{\pi m}{a}\right)^2 + \left(\frac{\pi n}{b}\right)^2 + \left(\frac{\pi p}{c}\right)^2 = \left(\frac{\omega}{c_0}\right)^2, \quad (32)$$

where c is the waveguide length (along OZ axis in Fig. 1), and p is a non-negative integer; $m, n = 1, 2, 3, \ldots,\ p = 0, 1, 2, \ldots$ for E-modes, and $m, n = 0, 2, 3, \ldots,\ p = 1, 2, \ldots$ for H-modes ($m = n = 0$ is not allowed). Noting that (32) is a equation of a sphere in terms of (m,n,p), 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:

$$N_c \approx 2\frac{V_e/8}{V_0} = \frac{8\pi V_c}{\lambda^3}\frac{\Delta f}{f_0}, \quad (33)$$

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 (33) 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 (33) 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),

$$\iiint_{V_c} \mathbf{E}_\mu \mathbf{E}_\nu dV = c\delta_{\mu\nu}\,, \tag{34}$$

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. (20), 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. 6 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 (32)). 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 equation:

$$N_c \approx N_w \Rightarrow \frac{c_t}{\lambda} = \frac{f_0}{4\Delta f}\,, \tag{35}$$

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. 6 shows the 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 (27). It should be noted that (27) is the capacity limit due to the information theory laws, and (23), (25), (26), and (33) 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).

VII. 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. 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.

[4] 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.

[5] E.D. Rothwell, M.J. Cloud, Electromagnetics, CRC Press, Boca Raton, 2001.

[6] J.D. Gibson (Ed.), The Communications Handbook, CRC Press, Boca Raton, 2002.

[7] S.R. Saunders, Antennas and Propagation for Wireless Communication Systems, Wiley, Chichester, 1999.

[8] 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.

[9] 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.

[10] Jakes, W.C. Jr.: ‘Microwave Mobile Communications’, John Wiley and Sons, New York, 1974.

[11] 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

[12] S.L. Loyka, Multi-Antenna Capacities of Waveguide and Cavity Channels, IEEE CCECE’03, Montreal, May 2003

[13] 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.

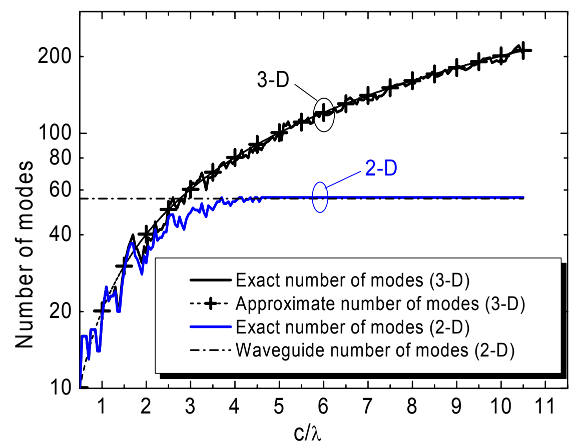

Fig. 6. Number of orthogonal modes in a rectangular cavity for $a = 4\lambda,\ b = 2\lambda$ and $\Delta f / f_0 = 0.01$.

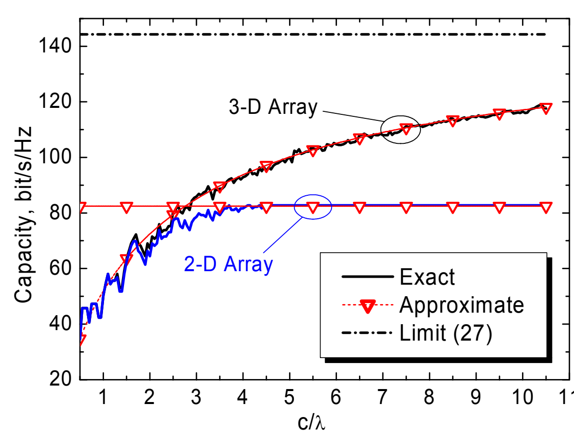

Fig. 7. Capacity in a rectangular cavity for $a = 4\lambda,\ b = 2\lambda$ and $\Delta f / f_0 = 0.01$.