Electromagnetic degrees of freedom of an optical system

Rafael Piestun and David A. B. Miller

E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305-4085

Received June 17, 1999; revised manuscript received December 14, 1999; accepted February 9, 2000

We present a rigorous electromagnetic formalism for defining, evaluating, and optimizing the degrees of freedom of an optical system. The analysis is valid for the delivery of information with electromagnetic waves under arbitrary boundary conditions communicating between domains in three-dimensional space. We show that, although in principle there is an infinity of degrees of freedom, the effective number is finite owing to the presence of noise. This is in agreement with the restricted classical theories that showed this property for specific optical systems and within the scalar and paraxial approximations. We further show that the best transmitting and receiving functions are the solutions of well-defined eigenvalue equations. The present approach is useful for understanding and designing modern optical systems for which the previous approaches are not applicable, as well as for application in inverse and synthesis problems. © 2000 Optical Society of America [S0740-3232(00)01005-X]

OCIS codes: 260.2110, 260.1960, 180.0180, 110.6880, 100.3190, 030.4070, 350.5730, 350.7420.

1. INTRODUCTION

The evaluation of the information that can be transmitted by an optical system is a problem of primary concern. The first step toward this goal is to determine the number of independent parameters needed to represent the output field or signal, which is known as the number of degrees of freedom (DOF). The significance of the DOF stems from its relation to performance measures such as resolution and capacity, in areas such as imaging or inverse problems from indirect measurements. In another approach, the DOF determine the number of useful independent communication channels available for the transmission of information.

In inverse problems, the DOF indicate the maximum amount of information that can be retrieved from the data in the presence of noise. In synthesis problems, the number of DOF permits the calculation of the number of different field distributions that it is possible to generate and, ultimately, the attainable resolution. Our approach will apply in all cases, regardless of the differences in terminology that may appear in different fields.

Classical theories were developed within the scalar approximations, considering transmission of information (generally images) between parallel planes of specific optical systems. Two main approaches were pursued, namely, that based on the sampling theory,[1,2] and that based on the theory of the prolate functions.[3] However, many modern photonic systems appear to operate well beyond the limits of validity of such theories. For example, systems utilizing nonparaxial waves or the near field or those involving nonplanar or three-dimensional (3D) domains cannot be analyzed in this framework.

Therefore in this paper we present a theory for the evaluation of the DOF or communication channels generated by arbitrary sources and transmitted by electromagnetic wave fields through systems with arbitrary boundary conditions. The difficulty here originates in dealing with vector fields in 3D space and in general with shift-variant systems. The conclusions are valid for one-dimensional (1D), two-dimensional (2D), or 3D sources and receivers, as well as for systems working from the near field to the far field, and involving more- or less-complicated architectures. The theory can be applied, for example, to near-field optical microscopy, dense optical interconnects, high-space–bandwidth diffractive optics, antennas, and in general various scattering and inverse problems.

The approach in this paper is similar to that recently discussed by one of us[4,5] for communication channels between volumes with scalar waves but considers the full vector electromagnetic case and generalizes to space-variant systems where there may be other fixed objects present. In addition, a rigorous mathematical approach in multidimensional Hilbert space is developed.

This paper is organized as follows: Section 2 presents a brief survey of previous research on the subject of DOF in optics and electromagnetism. Section 3 introduces the problem of EM DOF as a communication problem between domains in 3D space; we introduce the main definitions that are used throughout the paper. Section 4 presents a theorem stating the finiteness of the total interconnection strengths, which leads to the concept of a finite number of DOF in the presence of noise. Section 5 shows how to calculate the best source functions, leading to an eigenvalue problem. In Section 6 we show that the corresponding receiving functions are solutions of a dual eigenvalue problem. The main result of Section 4 is then specialized to the case of communication with these eigenfunctions, leading to a sum rule in terms of the eigenvalues. Section 7 shows an alternative and equivalent approach in terms of the singular system of the problem. In Section 8 we present two examples that illustrate the main features of the theory. Finally, the conclusions are presented in Section 9. For greater clarity in under-

0740-3232/2000/050892-11$15.00

standing the main ideas presented in this work, we have left most of the mathematical details and background information to (six) appendixes.

2. PREVIOUS RESEARCH

Toraldo di Francia[1] noted, using the sampling theorem, that an image formed by a finite pupil has a finite number of DOF. Gabor[2] extended a previous argument by Max von Laue in the optical domain and related it to his theory of information.[6] His basic interpretation involved the notion that each degree of freedom is associated with a Gaussian beam emerging from the object function and arriving at the image plane. The choice of a Gaussian beam is based on the following two facts: (a) The Gaussian function is the function with the most space-frequency localization,[6] and (b) Gaussian beams are invariant upon propagation in free space or what is known now as first-order optical systems.[7] He also extended his analysis to obtain a metric to evaluate the information content of an image based on the notion of entropy and considering the quantum character of light.

The analysis based on the sampling theorem presents some inconsistencies.[3,8,9] If the object were to have finite size, knowledge of its Fourier transform over a finite interval (pupil) would suffice to reconstruct the complete function by analytic continuation. Moreover, taking into account that no functions exist that are both space and band limited, sampling the Fourier transform of a finite object over a finite domain will not provide all the available information. Indeed, there are contributions of (not considered) sampling points outside the domain under consideration. Taking these contributions into account would lead to an infinite number of DOF. Another related aspect of the sampling theorem is that it is not applicable for small values of the space–bandwidth product,[9] a problem already raised by Gabor.[2]

The solution to these questions involves considering the inevitable existence of noise, and it was clarified by a later paper by Toraldo di Francia,[3] in which he applied the theory of the prolate spheroidal functions developed by Slepian and Pollack[10] and Landau and Pollack.[11] These functions are the eigenfunctions of the finite Fourier transform, i.e., the functions $\psi_n(x)$ that satisfy

$$\int_{-1}^{1} k(x - y)\psi_n(y)\mathrm{d}y = \lambda_n \psi_n(x), \tag{2.1}$$

where the convolution kernel is $k(x - y) = [\sin c(x - y)]/[\pi(x - y)]$. A notable property is that the set of functions $\psi_n(x)$ is complete and orthogonal in both $(-1, 1)$ and $(-\infty, \infty)$. The number $2c/\pi$ is associated with the so-called Shannon number obtained from the sampling theorem.[11,12]

The prolate functions can be directly applied to linear shift-invariant systems such as coherent imaging through an aberration-free slit aperture (Fig. 1). The object function can be expanded in the prolate functions as

$$o(x) = \sum_{n=0}^{\infty} \alpha_n \psi_n\left(\frac{x}{X}\right), \tag{2.2}$$

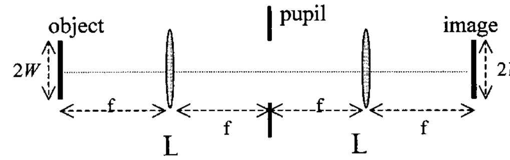

Fig. 1. Coherent imaging system used in the classical theory of DOF.

where $2X$ is the object width and α_n are expansion coefficients. The limiting pupil acts as an ideal low-pass filter, leading to an image function that can be also expanded in the $\psi_n(x)$:

$$i(y) = \sum_{n=0}^{\infty} \lambda_n \alpha_n \psi_n\left(\frac{y}{X}\right). \tag{2.3}$$

Each $\psi_n(x)$ can be associated with a degree of freedom, and knowledge of the eigenvalues suffices to reconstruct all the information related to an arbitrarily large number of DOF. However, it is known that the eigenvalues λ_n possess a steplike behavior, decreasing exponentially to zero. The conclusion is that in the presence of noise, only a finite number of coefficients $\lambda_n \alpha_n$ in Eq. (2.3) can be accurately determined.

Extension to 2D pupils other than square or circular[13] usually requires the use of numerical techniques, since there is no general analytical solution for the corresponding eigenvalue problem. The relation between the number of DOF and the Shannon number was further clarified in Refs. 14–16. In the extension to incoherent imaging[17] it was shown that the eigenvalues possess a triangle-function behavior; i.e., the eigenvalues decrease gradually to zero thus increasing the influence of noise. Reference 18 extended the previous concepts by introducing the singular-value formalism for imaging systems with asymmetric point-spread functions. The effects of noise and tolerances were discussed in Ref. 19 for finite convolution operators. Meanwhile, the singular-value decomposition was used in inverse problems as a means of determining the amount of meaningful information attainable from measured data.[20]

The DOF can also be defined for partially coherent sources[21,22] and polychromatic images.[23] The relation between DOF and capacity was discussed in Ref. 24. Recently, in Refs. 4 and 5 a different approach was presented: the problem was considered in terms of spatial channels for communication between volumes in free space by an analysis valid for scalar waves. This work showed that, for the scalar-wave case, considering the more general situation of volumes rather than only planar surfaces, there is a method for defining the orthogonal spatial channels. It also showed that there is an exact sum rule for the strengths of the connections between the two volumes, based only on a simple volume integral. This approach, when specialized to plane, parallel surfaces in the paraxial approximation, reproduces the prolate spheroidal function results quoted above.

Furthermore, the vector character of light fields has been traditionally considered in a duplication factor that takes into account the two polarization directions.[2] However, this is only an approximation, which is especially poor in the near field or in tightly focused beams. In electromagnetics there have been some partial efforts for determining the DOF of scattered fields. A sampling approach was presented in Ref. 25. In Ref. 26 a singular-value decomposition in free space for 1D planar domains was presented in the context of inverse scattering problems. The problem is thus simplified because it leads to a scalar, shift-invariant integral operator.

3. ELECTROMAGNETIC COMMUNICATION CHANNELS IN THREE-DIMENSIONAL DOMAINS

In this section we evaluate the DOF of an optical system with a rigorous EM approach. In this context it is useful to think of the optical system as a communication system. Mathematically, this involves a transmitting function or input signal, a receiving function or output signal, and an operational representation of the system or system response. Accordingly, determining the number of DOF is equivalent to determining the number of useful independent channels available for the transmission of information through that system. Thus our concern is not the DOF of the signal, image, or field but the DOF that can be identified after transmission through a well-defined system. Although this distinction is essential,[27] not much attention has been given to it in the past.

Let us consider a general system, as depicted in Fig. 2, composed of transmitting sources within a transmitting domain V_T and a separate (disjoint) receiving domain V_R. In addition, some material bodies may be present in the 3D space. Here we restrict the analysis to linear, homogeneous, and isotropic media and monochromatic radiation. In a separate communication we will deal with polychromatic radiation and time-varying signals. The fundamental concepts can also be extended to inhomogeneous and anisotropic media.

We consider electric (magnetic) current density sources represented by the complex vector field $\mathbf{J}(\mathbf{r}')[\mathbf{M}(\mathbf{r}')]$ within V_T. These sources can be real (impressed) sources, induced sources, or fictitious sources, i.e., sources that appear as a result of the application of equivalence theorems with the purpose of simplifying the problem under consideration.[28] The generated electric and magnetic complex vector fields within V_R are represented as $\mathbf{E}(\mathbf{r})$ and $\mathbf{H}(\mathbf{r})$. The full mathematical description of the problem is given by Maxwell equations, which in our case can be written as follows:

$$-\nabla \times \mathbf{E} = j\omega\mu\mathbf{H} + \mathbf{M},$$

$$\nabla \times \mathbf{H} = (\sigma + j\omega\epsilon)\mathbf{E} + \mathbf{J}, \qquad (3.1)$$

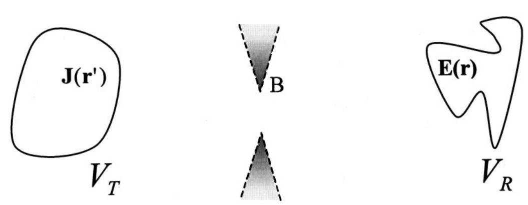

Fig. 2. Communication with EM waves between transmitting (V_T) and receiving (V_R) domains in the presence of material bodies (B).

where $\mu = \mu(\omega)$, $\epsilon = \epsilon(\omega)$, and $\sigma = \sigma(\omega)$ are the complex permeability, permittivity, and conductivity, respectively, of the medium and ω is the radiation frequency. A rigorous solution, under given boundary conditions imposed by material bodies, can be obtained by using the Green's function approach leading to integral equations[29] (see Appendix A). In our case this leads to the following expressions,

$$\mathbf{E}(\mathbf{r}) = \int_{V_T} \Gamma_{11}(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}' + \int_{V_T} \Gamma_{12}(\mathbf{r}, \mathbf{r}')\mathbf{M}(\mathbf{r}')\mathrm{d}\mathbf{r}',$$

$$\mathbf{H}(\mathbf{r}) = \int_{V_T} \Gamma_{21}(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}' + \int_{V_T} \Gamma_{22}(\mathbf{r}, \mathbf{r}')\mathbf{M}(\mathbf{r}')\mathrm{d}\mathbf{r}', \qquad (3.2)$$

where $\Gamma_{kl}(\mathbf{r}, \mathbf{r}')$ $(k, l = 1, 2)$ are the tensor Green's functions subject to the appropriate boundary conditions.[28–30] Note that these tensor (or dyadic) Green's functions are not independent, and only two suffice to determine the rest of them. A similar consideration applies to the relation between $\mathbf{E}$ and $\mathbf{H}$, i.e., either of them can be obtained from the other.

For brevity we consider only electric current sources [i.e., we presume for the moment that $\mathbf{M}(\mathbf{r}') = 0$], but the extension to magnetic sources is straightforward. Moreover, one can invoke an equivalence theorem to state an equivalent problem involving only one type of source (see also Section 9). Therefore we can write

$$\mathbf{E}(\mathbf{r}) = \int_{V_T} \Gamma(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}'. \qquad (3.3)$$

From a systems point of view, $\Gamma(\mathbf{r}, \mathbf{r}')$ is the impulse response of a linear, in general shift-variant, system.

For an adequate mathematical formalism, let us consider the Hilbert space $\mathcal{L}_2^3(V)$ composed of vectors of complex functions $\mathbf{f}(\mathbf{r}) \equiv [f_1(\mathbf{r})\, f_2(\mathbf{r})\, f_3(\mathbf{r})]^T$ defined on V, where $f_1(\mathbf{r}),\ f_2(\mathbf{r}),\ f_3(\mathbf{r}) \in \mathcal{L}_2(V)$ (the Lebesgue integrable functions defined on V), with the inner product defined as $(\mathbf{f}, \mathbf{g})_V = \int_V \mathbf{f}^{\mathbf{T}}(\mathbf{r})\mathbf{g}^*(\mathbf{r})\mathrm{d}\mathbf{r}$. Each element of the space $\mathcal{L}_2^3(V)$ represents a different vector field within V. In our case we consider field sources and electric fields belonging to Hilbert spaces $T(V_T)$ and $R(V_R)$, respectively.

Let us thus consider orthonormal bases in the domains V_T and V_R:

$$\mathbf{a}_{T1}(\mathbf{r}'), \mathbf{a}_{T2}(\mathbf{r}'), \ldots \mathbf{a}_{Ti}(\mathbf{r}'), \ldots,$$

$$\mathbf{a}_{R1}(\mathbf{r}), \mathbf{a}_{R2}(\mathbf{r}), \ldots \mathbf{a}_{Rj}(\mathbf{r}), \ldots, \qquad (3.4)$$

respectively. The $\mathbf{a}_{Ti}$ are source fields, and the $\mathbf{a}_{Rj}$ are electric fields. Accordingly, we can expand $\mathbf{J}(\mathbf{r}')$ and $\mathbf{E}(\mathbf{r})$ as follows,

$$\mathbf{J}(\mathbf{r}') = \sum_i b_i \mathbf{a}_{Ti}(\mathbf{r}'), \qquad \mathbf{E}(\mathbf{r}) = \sum_j d_j \mathbf{a}_{Rj}(\mathbf{r}), \qquad (3.5)$$

where $b_i = [\mathbf{J}(\mathbf{r}), \mathbf{a}_{Ti}(\mathbf{r})]_{V_T}$ and $d_j = [\mathbf{E}(\mathbf{r}), \mathbf{a}_{Rj}(\mathbf{r})]_{V_R}$. Substitution into Eq. (3.3) leads to

$$\sum_j d_j \mathbf{a}_{Rj}(\mathbf{r}) = \sum_i b_i \int_{V_T} \Gamma(\mathbf{r}, \mathbf{r}') \mathbf{a}_{Ti}(\mathbf{r}') \mathrm{d}\mathbf{r}' \quad (3.6)$$

where we have interchanged the order of integration and summation.[31] Performing the inner product with $\mathbf{a}_{Rj}(\mathbf{r})$, within V_R, on both sides of Eq. (3.6) gives the following relation between the coefficients of the transmitting and the receiving functions:

$$d_j = \sum_i g_{ji} b_i, \quad (3.7)$$

where

$$g_{ji} = \int_{V_R} \int_{V_T} \mathbf{a}_{Rj}^{*T}(\mathbf{r}) \Gamma(\mathbf{r}, \mathbf{r}') \mathbf{a}_{Ti}(\mathbf{r}') \mathrm{d}\mathbf{r}' \mathrm{d}\mathbf{r} = g[\mathbf{a}_{Ti}, \mathbf{a}_{Rj}] \quad (3.8)$$

g_{ji} are (complex) scalars with absolute values representing the strength of the coupling connection between sources at the transmitting domain and wave functions at the receiving domain. Thus the strength of the receiving function $\mathbf{a}_{Rj}(\mathbf{r})$ in V_R that results from a source $\mathbf{a}_{Ti}(\mathbf{r}')$ within V_T is $|g_{ji}|$. For a given system and once the bases are chosen, Eq. (3.7) gives the basic communications relation between receiving and transmitting waves. It is also possible to think of the $\{g_{ji}\}$ as an infinite square matrix of interconnection strengths. A heuristic interpretation of the coupling strengths in three dimensions for scalar waves can be found in Ref. 5.

In Sections 4 and 5 we discuss the choice of the optimal bases, i.e., those that lead to a minimum number of coupling coefficients. In the matrix representation, the optimal coefficients are associated with the bases that diagonalize the infinite matrix.

4. EVALUATION OF THE TOTAL CONNECTION STRENGTHS

In this section we derive a sum rule for evaluating the connections' strengths. As will be shown, this sum rule implies the finiteness and invariance (for a given system) of the total interconnection strengths. From this result we are led to the notion of a finite number of DOF in the presence of noise.

From the properties of the Green's functions we know that $\Gamma(\mathbf{r}, \mathbf{r}')$ is continuous except at $\mathbf{r} = \mathbf{r}'$. Since we are considering disjoint domains V_R and V_T, $\Gamma(\mathbf{r}, \mathbf{r}')$ can be expanded bilinearly as

$$\Gamma(\mathbf{r}, \mathbf{r}') = \sum_{ij} g_{ji} \mathbf{a}_{Rj}(\mathbf{r}) \otimes \mathbf{a}_{Ti}^{*T}(\mathbf{r}'), \quad (4.1)$$

where $\otimes$ represents the tensor (or outer) product defined as follows: If $\mathbf{u} = (u_1, u_2, u_3)^T$ and $\mathbf{v} = (v_1, v_2, v_3)^T$, then $\{\mathbf{u} \otimes \mathbf{v}^T\}_{kl} = u_k v_l$, with the notation $\{\mathbf{M}\}_{kl}$ as the elements of a tensor $\mathbf{M}$. Note that if the $\mathbf{a}_{Ti}(\mathbf{r})$ form an orthonormal basis, the $\mathbf{a}_{Ti}^*(\mathbf{r})$ do so also. We also define the norm: $\|\mathbf{M}\|^2 = \Sigma_{kl=1}^3 |\{\mathbf{M}\}_{kl}|^2$. Applying Parseval's theorem, we prove in Appendix B that the sum of the squared strengths S satisfies the following relation:

$$S = \sum_{ij} |g_{ji}|^2 = \int_{V_T} \int_{V_R} \|\Gamma(\mathbf{r}, \mathbf{r}')\|^2 \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}'. \quad (4.2)$$

The integral on the right-hand side of Eq. (4.2) is finite, stating that the total strength of the interconnections is bounded; i.e., the strength of the interconnections is negligible for all but a finite number of DOF. Therefore, although in principle the number of DOF is infinite, the number of practically useful channels is finite owing to the presence of noise. In effect, weakly connected communication channels will lead to weak receiving waves that become useless when the inevitable presence of noise is considered.

Observe that S depends only on the evaluation of the Green's function $\Gamma(\mathbf{r}, \mathbf{r}')$ over the domains V_T and V_R. The boundary conditions are implicit in $\Gamma(\mathbf{r}, \mathbf{r}')$. Thus in free space we obtain that S depends only on the geometry of the domains.

Note also that the result of Eq. (4.2) is valid regardless of the specific selected communication functions, implying that even the best strategy for the selection of the transmitting and receiving functions cannot lead to more than a finite number of effective DOF. However, proper system design can improve the number of channels with connection strengths above the thresholds imposed by noise levels. Note also that a high value of S does not necessarily indicate that the number of channels with $|g_{ji}|^2$ above a certain threshold is high, since such S may be due to a large number of weakly connected channels.

It should be noted that S has dimensions, but we can consider the nondimensional magnitude $\bar{S} = S/(\omega\epsilon)^2$ or the normalized number $\tilde{S} = S/|g_{ji}|^2_{\mathrm{MAX}}$. In the classical case of a (scalar) imaging system with a steplike behavior of the eigenvalues, the latter can be associated with a generalized Shannon number. This is true because most of the eigenvalues above a given threshold are close to unity, and then $\tilde{S}$ approximates the number of effective DOF. However, in general, this is not the case because the eigenvalues are not unity and they decrease gradually to zero.

In this section we have shown that the total connection strength S is finite and invariant upon changes in the transmitting and receiving functions. In addition, (once normalized) it can be understood as a generalized Shannon number.

5. SELECTION OF THE SOURCE FUNCTIONS

A natural criterion for selecting source and receiving functions is to choose the most strongly connected functions. We shall see below that the functions selected in this way generate two orthogonal sets of vector functions. Therefore our strategy leads to one-to-one communication channels; i.e., activation of such a source function will lead to the excitation of the corresponding receiving function and to a null excitation of the rest of the receiving functions.

To find these best communication functions, we maximize the connection strengths between pairs of receiving and transmitting functions. Formally, the problem is to maximize $|g_{ji}|^2$, as defined in Eq. (3.8), for normalized source functions. Let us consider the normalized source function $\mathbf{J}_N(\mathbf{r})$. It will generate an unnormalized receiving field $\mathbf{E}(\mathbf{r})$ as given by Eq. (3.3). In Appendix C we prove that $\mathbf{E}_N(\mathbf{r}) = \mathbf{E}/\sqrt{(\mathbf{E},\mathbf{E})_{V_R}}$ is, among the normalized receiving functions $\mathbf{a}_R(\mathbf{r})$, the one that maximizes the corresponding coupling coefficient $|g|^2 = |g[\mathbf{J}_N, \mathbf{a}_R]|^2$ for this given $\mathbf{J}_N(\mathbf{r})$; i.e.,

$$|g[\mathbf{J}_N, \mathbf{a}_R]|^2 \leqslant |g[\mathbf{J}_N, \mathbf{E}_N]|^2,$$
$$\forall \mathbf{a}_R : \mathbf{a}_R \in R(V_R), \qquad \|\mathbf{a}_R\| = 1. \tag{5.1}$$

Therefore we have reduced the problem to find the *normalized* source $\mathbf{J}_N(\mathbf{r})$ that maximizes

$$\begin{aligned} |g|^2 &= |g[\mathbf{J}_N, \mathbf{E}_N]|^2 \\ &= (\mathbf{E}, \mathbf{E})_{V_R} \\ &= \int_{V_R} \mathbf{E}^T(\mathbf{r})\mathbf{E}^*(\mathbf{r})\mathrm{d}\mathbf{r} \\ &= \int_{V_T}\int_{V_T} {\mathbf{J}_N}^T(\mathbf{r}')\mathbf{K}(\mathbf{r}',\mathbf{r}''){\mathbf{J}_N}^*(\mathbf{r}'')\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r}'', \end{aligned} \tag{5.2}$$

where

$$\mathbf{K}(\mathbf{r}', \mathbf{r}'') = \int_{V_R} \Gamma^T(\mathbf{r}, \mathbf{r}')\Gamma^*(\mathbf{r}, \mathbf{r}'')\mathrm{d}\mathbf{r}. \tag{5.3}$$

In other words, the best communication sources are those that maximize $|g|^2$, which turn out to be those that maximize the intensity of the electric field within the receiving volume V_R. The functional to be maximized can be expressed explicitly in terms of $\mathbf{J}_N(\mathbf{r})$ as in Eq. (5.2).

Let us define the operator $\mathcal{K}$ as the following tensor integral operator:

$$\mathcal{K}\mathbf{J} = \int_{V_T} \mathbf{K}^*(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}'. \tag{5.4}$$

In Appendix D we show that this is a compact, nonnegative, and self-adjoint operator. These properties enable us to derive important conclusions regarding the properties of the solutions of the problem stated in Eq. (5.2). First, note that we can rewrite Eq. (5.2) as

$$|g|^2 = (\mathbf{J}, \mathcal{K}\mathbf{J}) = (\mathcal{K}\mathbf{J}, \mathbf{J}), \tag{5.5}$$

the last equality being satisfied because of the self-adjointness of $\mathcal{K}$. According to known results from functional analysis,[32] $\mathcal{K}$ possesses nonnegative eigenvalues: $|g_0|^2 \geqslant |g_1|^2 \geqslant \cdots \geqslant 0$, and its eigenfunctions ς_n form an orthogonal set.

The source function that maximizes $|g|^2$ is the eigenfunction of $\mathcal{K}$ with the largest eigenvalue. Moreover, the set of orthogonal functions with the successive values of $|g|^2$ is the set of eigenfunctions arranged in descending order of their eigenvalues. We can write the corresponding eigenvalue equation as

$$\mathcal{K}\varsigma_n = |g_n|^2\varsigma_n. \tag{5.6}$$

Finally, note that we can expand the kernel $\mathbf{K}(\mathbf{r}, \mathbf{r}')$ as

$$\mathbf{K}(\mathbf{r}, \mathbf{r}') = \sum_n^{\infty} |g_n|^2\varsigma_n(\mathbf{r}) \otimes \varsigma_n^{*T}(\mathbf{r}'). \tag{5.7}$$

6. SELECTION OF THE RECEIVING FUNCTIONS

In Section 5 we showed how to find the best source functions based on the solution of an eigenvalue equation. Obviously, the corresponding receiving functions can be obtained from Eq. (3.3). However, the best receiving functions also satisfy a dual eigenvalue equation to Eq. (5.6) and thus are the solutions of a maximization problem dual to Eq. (5.2). Therefore the receiving functions possess properties similar to those of the source functions. Moreover, the dual problem presents an alternative way of calculating both sets of functions.

Let us consider the set of eigenfunctions of $\mathcal{K}$: ς_n ($n = 0, 1, 2 \ldots$). By definition, the corresponding normalized field functions at the receiving domain $[\epsilon_n(\mathbf{r})]$ satisfy

$$|g_n|\epsilon_n(\mathbf{r}) = \mathbf{E}_n(\mathbf{r}) = \int_{V_T} \Gamma(\mathbf{r}, \mathbf{r}')\varsigma_n(\mathbf{r}')\mathrm{d}\mathbf{r}' \tag{6.1}$$

where we used the fact that

$$\epsilon_n = \frac{\mathbf{E}_n}{(\mathbf{E}_n, \mathbf{E}_n)_{V_R}^{1/2}} = \frac{\mathbf{E}_n}{|g_n|}. \tag{6.2}$$

Multiplying from the left both sides of Eq. (6.1) by $\Gamma^{*T}(\mathbf{r}, \mathbf{r}')$, integrating over V_R, and changing the order of integration, we get

$$\int_{V_R} \Gamma^{*T}(\mathbf{r}, \mathbf{r}')|g_n|\epsilon_n(\mathbf{r})\mathrm{d}\mathbf{r} = \int_{V_T} K^*(\mathbf{r}', \mathbf{r}'')\varsigma_n(\mathbf{r}'')\mathrm{d}\mathbf{r}'', \tag{6.3}$$

and according to Eq. (5.6),

$$\int_{V_R} \Gamma^{*T}(\mathbf{r}, \mathbf{r}')\epsilon_n(\mathbf{r})\mathrm{d}\mathbf{r} = |g_n|\varsigma_n(\mathbf{r}'), \tag{6.4}$$

that is, the dual of Eq. (3.3) but only for the case of the eigenfunctions ς_n. This equation shows how to calculate $\varsigma_n(\mathbf{r}')$ once $\epsilon_n(\mathbf{r})$ and $|g_n|$ are known.

Note that although there is freedom to determine the square root of $|g_n|^2$ up to an arbitrary phase, we considered the positive root in Eqs. (6.1)–(6.4), without loss of generality. In effect, note also that $\epsilon_n(\mathbf{r})$ and $\varsigma_n(\mathbf{r}')$ are determined up to a constant phase.

Substitution into Eq. (6.1) and changing the order of integration leads to

$$|g_n|^2\epsilon_n(\mathbf{r}) = \int_{V_R} \mathbf{L}^*(\mathbf{r}_1, \mathbf{r})\epsilon_n(\mathbf{r}_1)\mathrm{d}\mathbf{r}_1 = \mathcal{L}\epsilon_n(\mathbf{r}), \tag{6.5}$$

where

$$\mathbf{L}(\mathbf{r}_1, \mathbf{r}) = \int_{V_T} \Gamma^*(\mathbf{r}, \mathbf{r}')\Gamma^T(\mathbf{r}_1, \mathbf{r}')\mathrm{d}\mathbf{r}', \tag{6.6}$$

and $\mathcal{L}$ is the corresponding tensor integral operator. Equation (6.5) is an eigenvalue equation, analogous to Eq. (5.6), for the case of the best receiving functions.

Finally, the sum rule [Eq. (4.2)] can be expressed in terms of the eigenvalues, which are the coupling strengths when the source and receiving functions are selected as the eigenfunctions of Eqs. (5.6) and (6.5):

$$\sum_n |g_n|^2 = \int_{V_T}\int_{V_R} \|\Gamma(\mathbf{r}, \mathbf{r}')\|^2 \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'. \tag{6.7}$$

7. ALTERNATIVE APPROACH: SINGULAR-VALUE DECOMPOSITION

We can state an alternative and equivalent approach to the determination of the DOF based on the singular-value decomposition of the source field (see Appendix E). Although mathematically more elegant, this approach is somewhat less intuitive for justifying the definition of the DOF, and thus we introduce it only after our previous discussion.

Accordingly, we consider the linear tensor operator $\mathcal{G}$ defined by the tensor Green's function Γ

$$\mathcal{G}\mathbf{J} = \int_{V_T} \Gamma(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{7.1}$$

and its adjoint [it is easy to verify that $(\mathcal{G}\mathbf{J}, \mathbf{E})_{V_T} = \mathbf{J}, \mathcal{G}^{+}\mathbf{E})_{V_R}$]:

$$\mathcal{G}^{+}\mathbf{E} = \int_{V_R} \Gamma^{*T}(\mathbf{r}, \mathbf{r}')\mathbf{E}(\mathbf{r}')\mathrm{d}\mathbf{r}. \tag{7.2}$$

By definition, the singular values of $\mathcal{G}$ are the eigenvalues of $\mathcal{G}^{+}\mathcal{G}$. However, it is easy to prove that $\mathcal{K} = \mathcal{G}^{+}\mathcal{G}$ and $\mathcal{L} = \mathcal{G}\mathcal{G}^{+}$. Therefore the previously defined $|g_n|$, square roots of the eigenvalues of the operators $\mathcal{K}$ and $\mathcal{L}$, are the singular values of both $\mathcal{G}$ and $\mathcal{G}^{+}$. Then we can state directly the existence of eigenfunctions $\varsigma_n(\mathbf{r}') \in T(V_T)$ and $\epsilon_n(\mathbf{r}) \in R(V_R)$, such that (see Appendix F)

$$\mathcal{G}\varsigma_n(\mathbf{r}') = |g_n|\epsilon_n(\mathbf{r}), \qquad \mathcal{G}^{+}\epsilon_n(\mathbf{r}) = |g_n|\varsigma_n \quad \forall n, \tag{7.3}$$

which correspond to Eqs. (6.4) and (3.3), respectively. The singular system is then defined as $(|g_n|, \varsigma_n, \epsilon_n)$.

In addition, we can expand source and receiving functions as follows,

$$\mathbf{J} = \sum_{n=1}^{\infty} (\mathbf{J}, \varsigma_n)\varsigma_n + \boldsymbol{P}\mathbf{J}, \qquad \mathcal{G}\mathbf{J} = \sum_{n=1}^{\infty} |g_n|(\mathbf{J}, \varsigma_n)\epsilon_n, \tag{7.4}$$

where $\boldsymbol{P}: V_T \rightarrow \boldsymbol{N}(\mathcal{G})$ is the orthogonal projection operator of V_T onto the nullspace $\boldsymbol{N}(\mathcal{G}) \equiv \{\mathbf{J} \in V_R : \mathcal{G}\mathbf{J} = 0\}$; i.e., $\boldsymbol{P}\mathbf{J}$ represents those source functions leading to a null response in the system defined by the operator $\mathcal{G}$.

In this context the role of the invariants is clear in the definition of the DOF. This fact determines the connection between this work and classical theories. Heuristically, information can be regarded as an entity that, in the absence of noise, is propagated from the input to the output of the system without destruction. For example, in the case of the prolate DOF, this invariance is apparent in the property that, within a finite domain, the eigenfunctions of Eq. (2.1) keep their shape after passing through the ideal optical system of Fig. 1. In our more general case, the invariance is manifested in the property of Eqs. (7.3): The DOF are defined by those source functions ς_n that, after diffraction in space, generate field functions ϵ_n that contain the same information, i.e., that upon an adjoint transformation $\mathcal{G}^{+}$ lead to the original source function. In this sense the operator $\mathcal{G}^{+}$ can be regarded as a generalized EM phase-conjugate operator.

8. EXAMPLES

As pointed out above, classical theories cannot be applied in cases where the scalar approximation is not valid. Therefore we will present two examples showing how to apply the proposed approach to the problem of evaluating the DOF. Previous approaches fail in these cases because we deal with dimensions of the order of the wavelength, with 3D domains, and the near field.

To apply our previous results to a real problem we must first define the domain of the sources and the domain of the receiving functions. The second step is to calculate the Green's tensor, which in some cases such as in free space, possesses an analytical expression (see Appendix A and Ref. 29). Otherwise, a numerical technique must be used to calculate it. The third step is to solve the eigenvalue equation [Eq. (5.6) or alternatively Eq. (6.5)] that leads to the singular system of the problem. Since these equations are defined in an infinite-dimensional space, we must transform them into a finite-dimensional space such that the solutions agree in the limit of an infinitely increasing dimension. A well-known numerical technique to solve this kind of eigenvalue equation is the Galerkin method,[32] which is explained in Appendix F.

Accordingly, we have to choose a finite set of normalized source functions $\mathbf{a}_{T1}(\mathbf{r}'), \mathbf{a}_{T2}(\mathbf{r}'), \ldots \mathbf{a}_{TN}(\mathbf{r}')$ that define a subspace in V_T. In the examples that follow we used pulse functions. In the new finite-dimensional space the problem is reduced to finding the eigenvalues and eigenvectors of a matrix $\mathbf{M}_{nm} = (\mathcal{K}\mathbf{a}_{Tn}, \mathbf{a}_{Tm})_{V_T}$ (see Appendix F). The calculation of the matrix is simplified if we consider that $\mathbf{M}_{nm} = (\mathbf{E}_n, \mathbf{E}_m)_{V_R}$, where $\mathbf{E}_n$ is the field generated by $\mathbf{a}_{Tn}$ within V_R. The eigenvalues of $\mathbf{M}$ are approximations to the eigenvalues of $\mathcal{K}$, and the eigenvectors of $\mathbf{M}$ are the expansion coefficients (in the $\mathbf{a}_{Ti}$ functions) of an approximation to the eigenfunctions of $\mathcal{K}$. Once the source functions are obtained, the corresponding receiving functions are calculated with Eq. (3.3).

In the first example we consider 1D transmitting and receiving domains in 3D free space, i.e., infinitely thin cylinders, as shown in Fig. 3. This implies that the source currents can flow in only one direction, x, leading to a reduction in the DOF. The electric field within the receiving domain can assume any value or direction.

The calculated eigenvalues are represented in Fig. 4. As expected, owing to the small dimensions of both do-

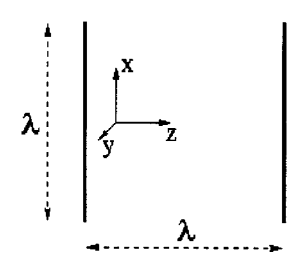

Fig. 3. 1D transmitting and receiving domains used in the first example of EM DOF calculation in 3D space ($\lambda = 500\,\mathrm{nm}$).

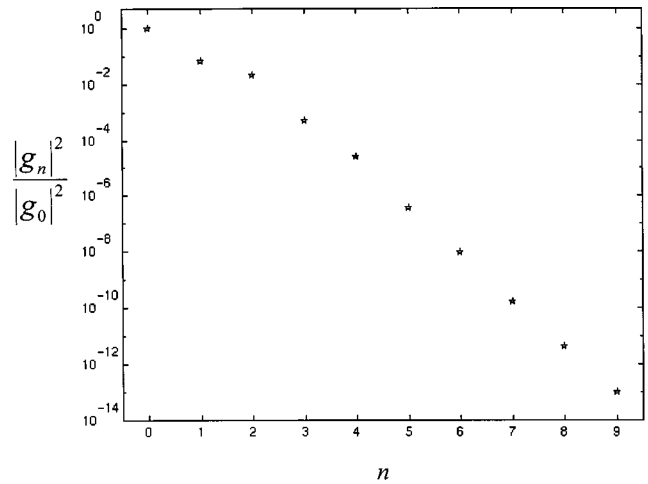

Fig. 4. Calculated eigenvalues for the system of Fig. 3. Note the rapid reduction of the coupling strengths.

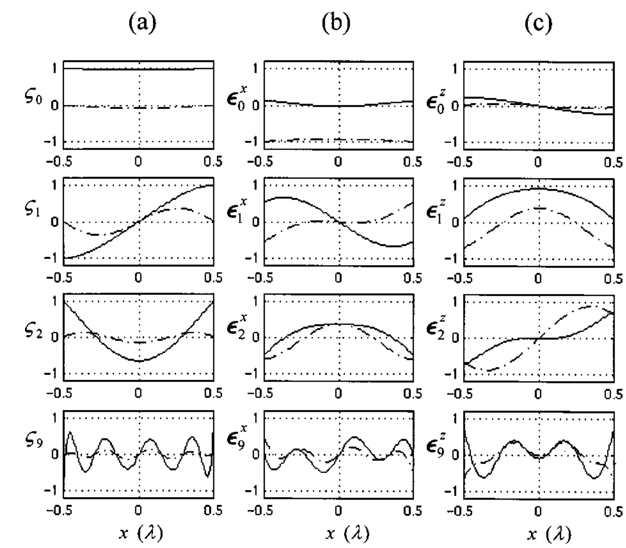

Fig. 5. Singular functions for the system of Fig. 3. (a) Source functions ς_n , (b) x component of the receiving functions $\epsilon_n{}^x$, (c) z component of the receiving functions $\epsilon_n{}^z$. The solid curves represent the real part of the functions; the dashed curves represent the imaginary part.

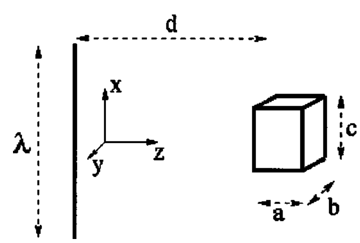

Fig. 6. 1D transmitting and 3D receiving domains used in the second example of EM DOF $d = 1.04\lambda$, $a = 0.04\lambda$, $b = 0.07\lambda$, $c = 0.11\lambda$ ($\lambda = 500\,\text{nm}$).

mains, the eigenvalues decrease rapidly to zero (although they never reach it). The first three eigenvalues account for approximately 93% of the total sum of Eq. (6.7). The coupling strengths will be practical only for a finite number of DOF. The information carried by the receiving singular functions corresponding to low eigenvalues will be lost in the presence of noise. Different criteria can be considered to determine the threshold values, which depend on the noise level and statistics. However, it is apparent in our case that for $n > 3$ the coupling strengths are too low to be of any practical value, leading to no more than three DOF for any reasonable criterion.

As stated, Galerkin's method can also be used to calculate the singular functions of the system. In Fig. 5(a) we show some of the source singular functions. In this case, we see that higher-order eigenfunctions lead to a higher degree of detail or higher harmonics. In effect, the number of zero crossings is equal to n.

In addition, the functions show a high degree of symmetry: For n even the functions are even, while for n odd the functions are odd. Figures 5(b) and 5(c) show the x and z components of the electric field along the receiving domain ($\epsilon_n{}^x$ and $\epsilon_n{}^z$ respectively). Note that the y component $\epsilon_n{}^y \equiv 0$ ($n = 1, 2, \ldots$).

In a second example we considered the same transmitting domain but a 3D receiving domain of subwavelength dimensions, as shown in Fig. 6. The eigenvalues are

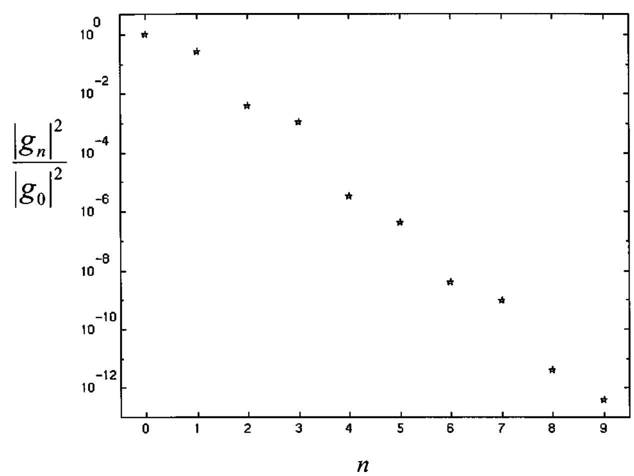

Fig. 7. Calculated eigenvalues for the system of Fig. 6.

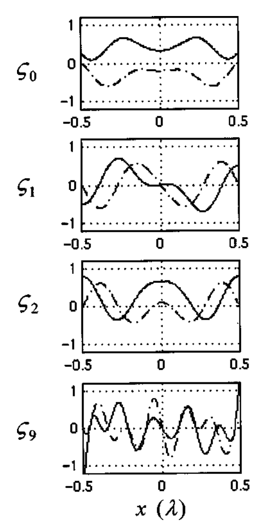

Fig. 8. Source singular functions ς_n for the system of Fig. 6. Solid curve, real part of the function; dashed curve, imaginary part.

shown in Fig. 7. Although in this case there are two relevant DOF, the absolute coupling strengths (not represented in the plot) are higher than in the previous case. This is quantified by the sum rule result of Eq. (6.7), which is approximately five times larger than in the previous example. The corresponding source singular functions are represented in Fig. 8.

9. DISCUSSION AND CONCLUSIONS

We justified our definition of DOF starting from a general approach of interconnection among modes within input and output domains. In this context we defined coupling coefficients and showed the relation between input and output in terms of these coefficients that form an infinite square coupling matrix.

A sum rule, which can be regarded as a generalization of the Shannon number, implies that independent of the basis functions used to calculate the coupling matrix, the total strength of the interconnections is invariant. This means that using functions different from the eigenfunctions leads to the use of more parameters with lower strengths.

Imposing a criterion for the best communication modes led us to an eigenvalue equation that defines the best communication functions (the eigenfunctions) and also the interconnection strengths (the eigenvalues). It also leads to a minimum number of coupling coefficients different from zero. Thus the best communication strategy is attained when there is a one-to-one correspondence between transmitting and receiving functions. The best receiving functions are obtained by solving a dual eigenvalue equation.

We showed that this approach is equivalent to a singular-value decomposition with the corresponding tensor operator. We then proposed an interpretation of the DOF in terms of invariants in the diffraction process.

The relevant singular values according to the specific signal-to-noise level can be associated with an effective number of DOF. Hence the (complex) coefficients associated with the corresponding source functions [in the expansion of Eq. (7.4)] suffice to represent the effective information transmitted by the system.

It should be noted that from Eq. (3.3) we have assumed that we are considering only the electric field $\mathbf{E}(\mathbf{r})$. The assumption was that once $\mathbf{E}(\mathbf{r})$ is known, the magnetic field $\mathbf{H}(\mathbf{r})$ can be completely determined, and thus specification of the magnetic field does not introduce new information. However, this is true only in the absence of noise, as we explain in what follows. In effect, the presence of noise led us to the conclusion that only a finite number of DOF are relevant; i.e., not all the information concerning $\mathbf{E}(\mathbf{r})$ can be retrieved. Therefore neither can all the information concerning $\mathbf{H}(\mathbf{r})$ be retrieved by measuring $\mathbf{E}(\mathbf{r})$. Would it be possible to retrieve additional information by measuring $\mathbf{H}(\mathbf{r})$? Let us assume that we consider the magnetic field as given in the second part of Eq. (3.2), with $\mathbf{M} = \mathbf{0}$. A completely analogous development of the theory for the DOF generated by the determination of $\mathbf{H}(\mathbf{r})$ would lead in general to a different set of singular source functions $\tilde{\varsigma}_n(\mathbf{r}')$. This is true because we are dealing with a different operator $\tilde{\mathcal{G}}$ such that

$$\tilde{\mathcal{G}} = \int_{V_T} \Gamma_{21}(\mathbf{r}, \mathbf{r}')\mathbf{J}(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{9.1}$$

where, in general, $\Gamma_{11} \neq \Gamma_{21}$ (Ref. 29). Thus, in the presence of noise, the relevant singular functions after determination of $\mathbf{H}(\mathbf{r})$ may deliver information different from that delivered by the relevant singular functions after determination of $\mathbf{E}(\mathbf{r})$. Nevertheless, in general, this additional information will not be completely independent from that obtained from $\mathbf{E}(\mathbf{r})$; i.e., there might be some redundancy. In conclusion, in the presence of noise, measuring both the electric and the magnetic fields may lead to additional useful information in comparison with determining only one of them. In a synthesis problem, this reasoning implies that after a specifying $\mathbf{E}(\mathbf{r})$ with use of the available DOF, one still has some freedom to specify $\mathbf{H}(\mathbf{r})$.

In Section 3 we assumed null magnetic sources $[\mathbf{M}(\mathbf{r}') = 0]$. This is by no means a limitation, since in an analogous way we could have developed the theory exclusively for magnetic sources. Moreover, if both types of source are present, we can always transform the equations mathematically to a single type of source.[28] Therefore, owing to this equivalence, the consideration of a second type of source does not add new DOF. Indeed, if we consider the receiving functions that are due to a given type of source, the additional receiving functions that are due to the second type of source will not be orthogonal to the functions already considered.

Another important aspect is that when the theory is developed for the scalar case and taken to the limit of thin planar surfaces and paraxial approximation, it coincides with the theory presented by Toraldo di Francia, leading to the prolate spheroidal functions and the Shannon number.[5]

The examples demonstrate the calculation of DOF of EM fields in 3D space. They also show the predicted behavior of the singular values and eigenfunctions. The numerical approach is feasible with current computing capabilities for reasonably large problems. The achievement of lower-complexity numerical procedures that would allow the solution of more complex structures and systems deserves further investigation. Note also that the complexity of the problem increases with complicated boundary conditions or nonuniform media.

For future work, starting from this point, it is interesting to take into consideration the signal-to-noise ratios and their statistics in order to extend the theory to establish a general theory of information in optics. This could be achieved for example by (a) establishing criteria for determining the effective DOF, i.e., defining the optimal number of DOF for a given signal-to-noise ratio; (b) calculating the capacity of given systems; and (c) defining strategies for information encoding, e.g., by assigning higher probability to events associated with the strongly connected eigenfunctions.

Another issue, which is beyond the scope of this paper, is that of determining the way in which the source functions can be generated and the receiving functions detected. Of course, this task is related to the design of specific systems for particular applications. However, the theory we have presented here establishes an appropriate framework to deal with this kind of problem. This approach is also useful for the solution of fundamental and practical problems arising in synthesis problems of 3D wave fields.[33–35]

In conclusion, we have presented a framework for the study of the DOF of optical systems, based on a rigorous EM formalism. This theory may prove useful for understanding and designing systems for which previous approaches are not applicable.

APPENDIX A: TENSOR GREEN'S FUNCTIONS

The linearity of the EM field equations [Eqs. (3.1)] implies that the field can be represented in terms of the source currents as in Eqs. (3.2) (Ref. 30). The Green's tensor functions can be interpreted as the impulse response of this linear system. Therefore $\Gamma_{11}(\mathbf{r}, \mathbf{r}')\hat{\mathbf{u}}$, with $\hat{\mathbf{u}}$ a unit vector, corresponds to the electric field produced at $\mathbf{r}$ by a unit (harmonic) electric current density at $\mathbf{r}'$, in the direction $\hat{\mathbf{u}}$.

The equations defining the tensor Green's functions are obtained on substitution of Eqs. (3.2) into field equations (3.1) and the specific boundary conditions.[29] In free space, for example, we get

$$\begin{aligned} \nabla \times \Gamma_{21} - j\omega\epsilon_0\Gamma_{11} &= \delta(\mathbf{r} - \mathbf{r}')\mathbf{I}, \\ \nabla \times \Gamma_{11} + j\omega\mu\Gamma_{21} &= 0, \\ \nabla \times \Gamma_{22} - j\omega\epsilon_0\Gamma_{12} &= 0, \\ \nabla \times \Gamma_{12} + j\omega\mu_0\Gamma_{22} &= -\delta(\mathbf{r} - \mathbf{r}')\mathbf{I}, \end{aligned} \tag{A1}$$

where $\mathbf{I}$ is the unit dyadic.

In free space the tensor Green's functions can be expressed in terms of a single scalar Green's function. For illustration, we consider the case of electric current sources. The components of the tensor $\gamma_{kl} = \{\Gamma_{11}\}_{kl}$ can be calculated as [28]

$$\begin{aligned} \gamma_{kk} &= \left(-j\omega\mu + \frac{1}{j\omega\epsilon}\frac{\partial^2}{\partial k^2}\right)G, \\ \gamma_{kl} &= \frac{1}{j\omega\epsilon}\frac{\partial^2 G}{\partial k\,\partial l}, \qquad k \neq l, \end{aligned} \tag{A2}$$

where

$$G = \frac{\exp(-jk|\mathbf{r} - \mathbf{r}'|)}{4\pi|\mathbf{r} - \mathbf{r}'|}. \tag{A3}$$

These expressions were used in the calculations of Section 8.

APPENDIX B: SUM RULE THEOREM

We prove here the result of Eq. (4.2).

Since $\Gamma : V_R \times V_T \rightarrow C$ is continuous, we can expand it as follows:

$$\Gamma(\mathbf{r}, \mathbf{r}') = \sum_i \mathbf{c}_i(\mathbf{r}) \otimes \mathbf{a}_{Ti}^{*T}(\mathbf{r}'). \tag{B1}$$

By Parseval's theorem we have that

$$\int_{V_T} \|\Gamma(\mathbf{r}, \mathbf{r}')\|^2 \mathrm{d}\mathbf{r}' = \sum_i |\mathbf{c}_i(\mathbf{r})|^2. \tag{B2}$$

We can also expand each $\mathbf{c}_i(\mathbf{r})$ as

$$\mathbf{c}_i(\mathbf{r}) = \sum_j g_{ji}\mathbf{a}_{Rj}(\mathbf{r}) \tag{B3}$$

and again apply Parseval's theorem:

$$\int_{V_R} |\mathbf{c}_i(\mathbf{r})|^2 \mathrm{d}\mathbf{r} = \sum_j |g_{ji}|^2. \tag{B4}$$

Integrating both sides of Eq. (B2) in V_R, we obtain

$$\int_{V_R}\int_{V_T} \|\Gamma(\mathbf{r}, \mathbf{r}')\|^2 \mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r} = \int_{V_R}\sum_i |\mathbf{c}_i(\mathbf{r})|^2 \mathrm{d}\mathbf{r}. \tag{B5}$$

By applying the monotone convergence theorem[32] we can interchange the order of integration and summation, which leads to

$$\begin{aligned} \int_{V_R}\int_{V_T} \|\Gamma(\mathbf{r}, \mathbf{r}')\|^2 \mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{r} &= \sum_i \int_{V_R} |\mathbf{c}_i(\mathbf{r})|^2 \mathrm{d}\mathbf{r} \\ &= \sum_i \sum_j |g_{ji}|^2, \end{aligned} \tag{B6}$$

where we have used Eq. (B4) in the last equality. This completes our proof.

APPENDIX C: MAXIMIZATION OF COUPLING COEFFICIENTS

We want to prove here the statement of Eq. (5.1) that shows that the normalized receiving function that maximizes the coupling coefficient with a given source function is the (normalized) function generated by the same source. First note that $|g|^2 = |(\mathbf{E}, \mathbf{a}_R)_{V_R}|^2$. Let us then consider any normalized receiving function $\mathbf{a}_R(\mathbf{r})$ other than $\mathbf{E}_N(\mathbf{r})$. According to the Schwarz inequality we can write

$$|(\mathbf{E}_N, \mathbf{a}_R)_{V_R}|^2 \leqslant (\mathbf{E}_N, \mathbf{E}_N)_{V_R}(\mathbf{a}_R, \mathbf{a}_R)_{V_R} = 1, \tag{C1}$$

and multiplying by $(\mathbf{E}, \mathbf{E})_{V_R} = |(\mathbf{E}, \mathbf{E}_N)_{V_R}|^2$, we get

$$|(\mathbf{E}, \mathbf{a}_R)_{V_R}|^2 \leqslant |(\mathbf{E}, \mathbf{E}_N)_{V_R}|^2, \tag{C2}$$

i.e., what we wanted to prove:

$$|g[\mathbf{J}_N, \mathbf{a}_R]|^2 \leqslant |g[\mathbf{J}_N, \mathbf{E}_N]|^2. \tag{C3}$$

APPENDIX D: MATHEMATICAL PROPERTIES OF $\mathcal{K}$

1. Compact

There is more than one way to prove that $\mathcal{K}$ is a compact operator. Here we base the proof on the fact that each component $\mathscr{k}_{pq}$ of the tensor operator is a scalar compact integral operator. $\Gamma(\mathbf{r}, \mathbf{r}')$ is composed of continuous functions on $\mathbf{r} \in V_R$; therefore $\mathbf{K}(\mathbf{r}', \mathbf{r}'')$ is also composed of continuous functions, and thus $\mathcal{K}$ is a bounded operator. To prove compactness we need to prove the existence of a sequence of bounded finite-rank operators $[\mathcal{K}^{(n)}]$ such that $\|\mathcal{K}^{(n)} - \mathcal{K}\| \xrightarrow{n\rightarrow\infty} 0$, where the norm of an operator is defined as $\|\mathcal{K}\| = \sup\{\|\mathcal{K}\mathbf{J}\| : \|\mathbf{J}\| \leq 1\}$. For this purpose we use a known theorem that states the compactness of bounded integral operators generated by $\mathcal{L}_2$ kernels.[32]

Let us consider the components of $\mathcal{K} : \{\mathcal{K}\}_{pq} = \mathscr{k}_{pq}$. The operators $\mathscr{k}_{pq}$ operate from $\mathcal{L}_2$ to itself; they are generated by kernels on $V_T \times V_T$, and they are bounded. Thus each one of them is a compact operator.[32] Therefore by definition there are sequences of bounded finite-rank operators $[\mathscr{k}_{kl}^{(n)}]$ such that $\|\mathscr{k}_{kl}^{(n)} - \mathscr{k}_{kl}\| \xrightarrow{n\to\infty} 0$. Let us consider the sequence of operators $\mathcal{K}^{(n)}$ in $\mathcal{L}_2{}^3$, defined as $\{\mathcal{K}^{(n)}\}_{kl} = \mathscr{k}_{kl}^{(n)}$. We now show that $\|\mathcal{K}^{(n)} - \mathcal{K}\| \xrightarrow{n\to\infty} 0$, leading to the conclusion that $\mathcal{K}$ is compact:

$$\begin{aligned}\|\mathcal{K}^{(n)} - \mathcal{K}\| &= \sup_{\|\mathbf{J}\| \leq 1} \|(\mathcal{K}^{(n)} - \mathcal{K})\mathbf{J}\| \\ &\leq \sum_{p,q=1}^{3} \sup_{\|\mathbf{J}\| \leq 1} \sqrt{\int_{V_T} |[k_{pq}^{(n)} - k_{pq}] j_q|^2 \mathrm{d}\mathbf{r}'}, \\ &\leq \sum_{p,q=1}^{3} \sup_{\|j_q\| \leq 1} \|[k_{pq}^{(n)} - k_{pq}] j_q\| \\ &= \sum_{p,q=1}^{3} \|[k_{pq}^{(n)} - k_{pq}]\| \xrightarrow{n\to\infty} 0, \end{aligned} \tag{D1}$$

where we denoted $\{\mathbf{J}\}_q = j_q$.

2. Nonnegative

We have to prove that $(\mathcal{K}\mathbf{J}, \mathbf{J}) \geq 0$, and $(\mathcal{K}\mathbf{J}, \mathbf{J}) = 0$ if and only if $\mathbf{J} = 0$. This is obvious from Eqs. (5.2) and (5.5).

3. Self-Adjoint

Given an operator $\mathbf{A} : X \to Y$, the adjoint operator $\mathbf{A}^+ : Y \to X$ is defined as the operator satisfying $(\mathbf{A}x, y)_Y = (x, \mathbf{A}^+ y)_X$ for every $x \in X$, $y \in Y$.

In our case, since $\mathcal{L}_2^3$ is a complex Hilbert space and $(\mathcal{K}\mathbf{J}, \mathbf{J})$ is always real, then $\mathcal{K}$ is self-adjoint; i.e., $(\mathcal{K}\mathbf{J}_1, \mathbf{J}_2) = (\mathbf{J}_1, \mathcal{K}\mathbf{J}_2)$ for all $\mathbf{J}_1, \mathbf{J}_2 \in \mathcal{L}_2^3$.[29]

APPENDIX E: SINGULAR-VALUE DECOMPOSITION

We recall the definition of singular values[36]: Given two Hilbert spaces X and Y, and $\mathcal{A} : X \to Y$ a compact linear operator, the square roots of the eigenvalues of the (self-adjoint compact) operator $\mathcal{A}^+\mathcal{A} : X \to X$ are called the singular values of $\mathcal{A}$.

Singular-value decomposition theorem: If $\{\mu_n\}$ is the sequence of the nonzero singular values of $\mathcal{A}$ repeated according to their multiplicity; then there exist orthonomal sequences $\{x_n\} \in X$ and $\{y_n\} \in Y$, such that

$$\mathcal{A}x_n = \mu_n y_n, \qquad \mathcal{A}^+ y_n = \mu_n x \qquad (n = 1, 2, \ldots). \tag{E1}$$

For every $x \in X$, it holds that the singular-value decomposition is

$$x = \sum_{n=1}^{\infty} (x, x_n) x_n + Px, \tag{E2}$$

with the orthogonal projection $P : X \to N(\mathcal{A})$. $N(\mathcal{A})$ is the nullspace of $\mathcal{A}$.

In addition,

$$\mathcal{A}x = \sum_{n=1}^{\infty} \mu_n (x, x_n) y_n. \tag{E3}$$

The system (μ_n, x_n, y_n) is called a singular system of $\mathcal{A}$.

APPENDIX F: GALERKIN'S METHOD

Galerkin's method is a method to solve operator equations in infinite-dimensional Hilbert spaces by orthogonal projection into finite-dimensional subspaces.[32,36] The original equation in a Hilbert space H is replaced by another equation defined in a finite-dimensional subspace E_N. E_N is thus spanned by a finite set of functions χ_n, called the test functions.

We denote P_N the orthogonal projection onto E_N. Upon projection, the eigenvalue equation

$$\mathcal{A}x = \mu x \qquad (x \in H) \tag{F1}$$

leads to

$$P_N \mathcal{A} y = v y, \tag{F2}$$

where $y = P_N x$.
The function y can be expanded in the χ_n as $y = \sum_{n=1}^{N} \alpha_n \chi_n$.

A solution to Eq. (F2) is obtained by solving[32]

$$\sum_{n=1}^{N} (\mathbf{M}_{nm} - v\mathbf{I}_{nm}) \alpha_n = 0, \qquad m = 1 \ldots N, \tag{F3}$$

where $\mathbf{M}_{nm} = (\mathcal{A}\chi_n, \chi_m)$ and $\mathbf{I}_{nm} = (\chi_n, \chi_m)$.

The solutions for v are second-order approximations to the eigenvalues and are given by

$$\det(\mathbf{M}_{nm} - v\mathbf{I}_{nm}) = 0. \tag{F4}$$

The associated approximations to the eigenfunctions are calculated from Eq. (F3).

ACKNOWLEDGMENTS

R. Piestun thankfully acknowledges support from a Fullbright fellowship. This work was supported by the U.S. Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant F49620-97-1-0517 (DARPA Ultra-Photonics).

REFERENCES AND NOTES

1. G. Toraldo di Francia, "Resolving power and information," J. Opt. Soc. Am. **45**, 497–501 (1955).
2. D. Gabor, "Light and information," in *Progress in Optics*, E. Wolf, ed. (North-Holland, Amsterdam, 1961), Vol. 1, pp. 109–153.
3. G. Toraldo di Francia, "Degrees of freedom of an image," J. Opt. Soc. Am. **59**, 799–804 (1969).
4. D. A. B. Miller, "Spatial channels for communicating with waves between volumes," Opt. Lett. **23**, 1645–1647 (1998).
5. D. A. B. Miller, "Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths," Appl. Opt. **39**, 1681–1699 (2000).
6. D. Gabor, "Theory of information," J. IEE **93**, 429–153 (1946).

7. A. Siegman, *Lasers* (University Science Books, Mill Valley, Calif., 1986).
8. H. Wolter, "On basic analogies and principal differences between optical and electronic information," in *Progress in Optics*, E. Wolf, ed. (North-Holland, Amsterdam, 1961), Vol. 1, pp. 157–210.
9. D. Slepian, "On bandwidth," Proc. IEEE **64**, 292–300 (1976).
10. D. Slepian and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty. I," Bell Syst. Tech. J. **40**, 43–64 (1961).
11. H. J. Landau and H. O. Pollack, "Prolate spheroidal wave functions, Fourier analysis and uncertainty. II," Bell Syst. Tech. J. **40**, 65–84 (1961).
12. B. R. Frieden, "Evaluation, design, and extrapolation methods for optical signals, based on use of the prolate functions," in *Progress in Optics*, E. Wolf, ed. (North-Holland, Amsterdam, 1971), Vol. 9, pp. 313–407.
13. F. Gori and L. Ronchi, "Degrees of freedom for scatterers with circular cross section," J. Opt. Soc. Am. **71**, 250–258 (1981).
14. F. Gori and G. Guattari, "Shannon number and degrees of freedom of an image," Opt. Commun. **7**, 163–165 (1973).
15. W. Lukosz, "Optical systems with resolving powers exceeding the classical limit," J. Opt. Soc. Am. **56**, 1463–1472 (1966).
16. A. W. Lohmann, "The space–bandwidth product applied to spatial filtering and holography," Research Paper RJ-438 (IBM San Jose Research Laboratory, San Jose, Calif., 1967), pp. 1–23.
17. M. Bendinelli, A. Consortini, L. Ronchi, and B. Roy Frieden, "Degrees of freedom, and eigenfunctions, for the noisy image," J. Opt. Soc. Am. **64**, 1498–1502 (1974).
18. R. Barakat, "Shannon numbers of diffraction images," Opt. Commun. **6**, 391–394 (1982).
19. G. Newsam and R. Barakat, "Essential dimension as a well-defined number of degrees of freedom of finite-convolution operators appearing in optics," J. Opt. Soc. Am. A **2**, 2040–2045 (1985).
20. E. R. Pike, J. G. McWhirter, M. Bertero, and C. de Mol, "Generalised information theory for inverse problems in signal processing," IEE Proc. F, **131**, 660–670 (1984).
21. E. Wolf, "New theory of partial coherence in the space–frequency domain. Part I: spectra and cross-spectra of steady-state sources," J. Opt. Soc. Am. **72**, 343–351 (1982).
22. A. Starikov, "Effective number of degrees of freedom of partially coherent sources," J. Opt. Soc. Am. **72**, 1538–1544 (1982).
23. S. A. Basinger, E. Michielssen, and D. J. Brady, "Degrees of freedom of polychromatic images," J. Opt. Soc. Am. A **12**, 704–714 (1995).
24. I. J. Cox and C. J. R. Sheppard, "Information capacity and resolution in an optical system," J. Opt. Soc. Am. A **3**, 1152–1158 (1986).
25. O. M. Bucci and G. Franceschetti, "On the degrees of freedom of scattered fields," IEEE Trans. Antennas Propag. **37**, 918–926 (1989).
26. R. Pierri and F. Soldovieri, "On the information content of the radiated fields in the near zone over bounded domains," Inverse Probl. **14**, 321–337 (1998).
27. A. W. Lohmann, R. G. Dorsch, D. Mendlovic, Z. Zalevsky, and C. Ferreira, "Space–bandwidth product of optical signals and systems," J. Opt. Soc. Am. A **13**, 470–473 (1996).
28. R. Harrington, *Time Harmonic Electromagnetic Fields* (McGraw-Hill, New York, 1961).
29. L. Felsen and N. Markuvitz, *Radiation and Scattering of Waves* (IEEE Press, Piscataway, N.J., 1994).
30. R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Interscience, New York, 1961).
31. Note that by Lebesgue's (dominated convergence) theorem[32] this is possible if there exists an integrable majorant function $m(\mathbf{r})$ such that $\|\Sigma_i b_i \mathbf{a}_{Ti}(\mathbf{r}')\| \leqslant m(\mathbf{r})$ for all i. This is satisfied, for example, if the basis functions $\mathbf{a}_{Ti}$ are bounded almost everywhere.
32. See, for example, D. Porter and D. S. G. Stirling, *Integral Equations* (Cambridge U. Press, Cambridge, UK, 1990).
33. R. Piestun and J. Shamir, "Control of wave-front propagation with diffractive elements," Opt. Lett. **19**, 771–773 (1994).
34. R. Piestun, B. Spektor, and J. Shamir, "Wave fields in three dimensions: analysis and synthesis," J. Opt. Soc. Am. A **13**, 1837–1848 (1996).
35. R. Piestun and D. A. B. Miller, "Ill-posedness of the three-dimensional wave-field synthesis problem," presented at the Annual Meeting of the Optical Society of America, Santa Clara, California, September 26–October 1, 1999, paper WLL27.
36. See, for example, R. Kress, *Linear Integral Equations* (Springer-Verlag, Berlin, 1989).