

Communicating with waves between volumes: evaluating orthogonal spatial channels and limits on coupling strengths

David A. B. Miller

A rigorous method for finding the best-connected orthogonal communication channels, modes, or degrees of freedom for scalar waves between two volumes of arbitrary shape and position is derived explicitly without assuming planar surfaces or paraxial approximations. The communication channels are the solutions of two eigenvalue problems and are identical to the cavity modes of a double phase-conjugate resonator. A sum rule for the connection strengths is also derived, the sum being a simple volume integral. These results are used to analyze rectangular prism volumes, small volumes, thin volumes in different relative orientations, and arbitrary near-field volumes: all situations in which previous planar approaches have failed for one or more reasons. Previous planar results are reproduced explicitly, extending them to finite depth. Depth is shown not to increase the number of communications modes unless the volumes are close when compared with their depths. How to estimate the connection strengths in some cases without a full solution of the eigenvalue problem is discussed so that estimates of the number of usable communications modes can be made from the sum rule. In general, the approach gives a rigorous basis for handling problems related to volume sources and receivers. It may be especially applicable in near-field problems and in situations in which volume is an intrinsic part of the problem. © 2000 Optical Society of America

OCIS codes: 260.1960, 110.2990.

1. Introduction

Much of the information that we communicate is sent in the form of waves. Often, as in vision or radio transmission, the waves propagate through free space. It is obviously important to know how many independent information channels we have available to us and how good a connection we can make on a given channel. These elements are among the factors that will limit our ability to communicate information, providing bounds, for example, on optical interconnection, optical memory access, and our ability to exploit techniques such as very-fine-line lithography or near-field microscopy. The details of such applications are beyond the scope of this paper, but here we deal with two particular underlying and related questions and some of their consequences: (i)

At a given frequency, what are the independent spatial channels available for sending information on waves between two volumes? (ii) How strong are these connections? These are simple questions, and the discussion is restricted to the simplest case of scalar waves.

Despite this simplicity, previous approaches, although useful, do not yield simple and universal answers; much such research addressed communication between regular plane surfaces only. Here we show that there is a rigorous and an exact approach to these problems for arbitrary volumes that

(i) Allows us to define uniquely the set of available spatial channels for communicating between arbitrary volumes¹ (the communications modes).

(ii) Gives us a very general sum rule for the connection strength and the number of such channels.¹

(iii) Enables us to deduce the approximate answers of previous models that were based on communicating paraxially between parallel planar surfaces.

(iv) Lets us draw clear conclusions about the effects of finite thickness on the number of communications channels.

The author (dabm@ee.stanford.edu) is with Ginzton Laboratory, 450 Via Palou, Stanford University, Stanford, California 94305-4085.

Received 9 June 1999; revised manuscript received 6 October 1999.

0003-6935/00/111681-19\$15.00/0

© 2000 Optical Society of America

(v) Enables us to estimate the number of usable communications modes in some cases.

(vi) Relates communications modes to resonator theory.¹

For many situations, especially those involving the near field, a scalar approach clearly is not sufficient for optical problems. In other research, my co-worker and I² analyzed the vector case and found analogous results. In this paper, I analyze the simpler scalar case in detail and build some concepts and intuition that underlie both the scalar and the vector cases.

Two important previous approaches to this problem (for plane surfaces) are summarized in Section 2. In Section 3 the model that is used in this paper is defined. The approach here is based on expanding the source function in the transmitting volume and the resultant wave in the receiving volume in complete sets. By use of linear algebra—here for the case of two spaces with different basis sets—the sum rule for the coupling strengths between the volumes is derived in Section 4. In Section 5 the concept of communications modes is introduced. The communications modes essentially are pairs of functions, with each pair consisting of one function for the transmitting volume and one for the receiving volume. The different communications modes are all orthogonal and define the two sets of functions with the best possible couplings between the volumes. These two sets of functions can be calculated as solutions to two eigenfunction problems, and they correspond to the best possible distinct communications channels between the volumes. The sum rule takes on a particularly simple form when the communications modes are used as the basis sets in the respective volumes. Explicit results for the communications modes between rectangular prism volumes and between small volumes are derived, and with an explicit calculation the communications modes between extreme volumes are illustrated. Also shown is the relation between communications modes and resonator modes, and the size of the connection strengths is rationalized on the basis of a heuristic approach. The main conclusions are summarized in Section 6.

2. Previous Approaches

The problem stated in Section 1 has been discussed mostly in the context of optics for which we might wish to know the number of separately resolvable spots when looking at an object with an imaging lens. We could also refer to this number of resolvable spots as the effective number of degrees of freedom in the image. A resolved spot corresponds effectively to a distinct channel from object to image, and each such spot could be used to send a separate channel of information, as in turning on and off a small light in the object and detecting the result with a small detector in the image space.

We can get an intuitive feel for the problem's solution by considering elementary diffraction theory.

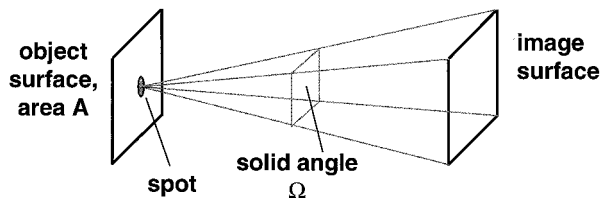


Fig. 1. Illustration of the diffraction approach to estimating the number of independent channels or resolvable spots for communicating between two surfaces. $N \sim \Omega A / \lambda^2$, where N is the number of channels and λ is the wavelength.

The smallest spot we can use will be the one from which the diffracted light approximately fills the aperture of the imaging lens or surface (see Fig. 1). The diffraction angle from a spot of size d in one direction is $\theta \sim \lambda/d$, so the diffraction solid angle from a spot of area a is $\Omega \sim \lambda^2/a$. If we try to use smaller spots, the diffraction angle will get larger, and we will start not to collect some significant fraction of the light with the lens or the image surface; hence we will lose strength in the interconnection. We will also have increasing difficulty in distinguishing the signal from two such adjacent spots if we attempt to image them by using a lens in the image surface. Hence we choose the spot area a such that its diffraction solid angle corresponds to the solid angle Ω , which is subtended by the image surface at the object surface. The number of distinct spots is $N \sim A/a$, i.e.,

$$N \sim \Omega A / \lambda^2. \quad (1)$$

Expression (1) therefore corresponds to the number of distinct channels for communication between these two surfaces. (It is easily shown that the same result is obtained for N if the roles of the two surfaces are interchanged.) This type of model was formalized to some extent by Gabor (see Ref. 3 and references therein), di Francia,^{4,5} and other authors (e.g., Walther⁶), and, for simplicity, we refer to it as the diffraction model. This kind of approach, which goes back at least to von Laue (see discussion in Ref. 3) still appears to be the main one for such problems. We also can view this kind of model equivalently in terms of the spatial frequencies in the object surface being Fourier transformed into the image surface by propagation, with the finite size of the image surface imposing a frequency cutoff. In such a picture, we can use the sampling theorem to deduce the effective number of independent channels for such a band-limited function.

The diffraction model is very useful but has several formal weaknesses. Gabor³ concluded that the best form of spot to consider was a Gaussian spot, which is well known to have a minimum uncertainty property—it has essentially the smallest diffraction angle for a given spot size—and he used this for his quantitative results. A Gaussian spot, however, extends sideways forever, so no finite object volume can actually generate a true Gaussian spot. Perhaps more importantly, no two Gaussian spots are truly independent in the mathematical sense: They al-

ways overlap to some degree and are not orthogonal. We cannot use such Gaussian spots as an orthogonal basis to describe the object. Of course, we can imagine intuitively that the object is made up of a series of spots, and using a set of approximately Gaussian spots is not an unreasonable way to approximate the object's emitted light.

The diffraction model's Gaussian spots therefore do not rigorously define independent communication channels between two volumes. It is reasonable to expect that there are rigorously independent communication channels (or communication modes) between two volumes (and how to evaluate these is described below). Essentially, Gabor took an intelligent guess at a set of functions that, although they are not the rigorous communications modes between two volumes, have many of the expected properties of such communications modes and can be used intuitively to generate a reasonable guess at some useful results.⁷ Gabor is aware of problems with the rigor of this approach; to quote Gabor, "It may be mentioned that the theorem has not yet been proved with a rigour which would satisfy mathematicians, but physicists have their own standards in these matters . . ." (Ref. 3, p. 120). In the sampling-theorem approach, to specify the function completely at all points within a finite domain (rather than at a discrete set of sampling points) actually also requires an infinite number of samples, so there may be channels beyond those counted in a simple sampling-theorem approach.

A second class of problem occurs if we try to extend the diffraction model (to regimes that violate its approximations and that it was never intended to model) to make predictions when the object is much smaller than a wavelength in size. If we were naively to follow Gabor's result there would then be no degrees of freedom or communications channels for such small volumes, which is clearly not the case. Many radio receivers and transmitters, for example, are much smaller than a wavelength in size, yet they communicate successfully, so there is at least one usable channel despite their small size. Another example for which scalar waves are an accurate approach is a loudspeaker and a microphone (or, for that matter, the human mouth and ear), both of which are substantially smaller than the wavelength of low-frequency sound.

The issue of the number of degrees of freedom was considered from another viewpoint, which, for simplicity here, we refer to as the eigenfunction approach, by di Francia.⁸ di Francia draws our attention to a mathematical fact that apparently was noted by several previous authors (see Ref. 8 for a discussion): The wave that emerges from a volume is presumably an analytic function for any real physical situation. If we know an analytic function to an arbitrarily high accuracy over any finite range, we can deduce its values everywhere else. This property is familiar, for example, in expanding a simple analytic function in a Taylor series about any given point; the value of the function and its derivatives at

that point are sufficient to establish the coefficients of a power series that defines the function everywhere, and the derivatives can certainly all be calculated if we know the function over some finite range about the point of interest. Equivalently, we could perform a multipole expansion of a given source or wave about the point of interest. Hence, if we know the wave over some finite aperture, we can calculate it everywhere else. We are therefore forced to conclude that the number of degrees of freedom in the wave that enters the aperture is not changed by the size of the aperture and must be the same as the number of degrees of freedom in the entire wave. This relation appears to contradict the diffraction result and our physical intuition; regardless of the size of the lens aperture used to observe the wave field, the number of degrees of freedom is apparently the same.

Fortunately, di Francia suggests a solution to this paradox: The entire wave may be expanded in some complete set of (orthonormal) basis functions. We can view the expansion coefficients as being the amplitudes of the degrees of freedom with one basis function corresponding to one degree of freedom. To describe the wave within only the aperture, we can use the same basis set, and we will have to use the same number of elements of the set as we did for the complete wave. Hence we still have the same number of degrees of freedom. But we will find that the couplings from many such basis functions on the object surface to basis functions on the image surfaces may be so small as to be negligible, and, if we count only those degrees of freedom for which the coupling coefficients are substantial, we essentially may retain the result that our intuition and the diffraction model's results suggest. This conclusion by di Francia is not proved in any rigorous sense for arbitrary apertures either, but it is strongly supported by specific calculations for rectangular and circular apertures.

In fact, considerable study in this area [e.g., Refs. 9–14 (including the research by di Francia⁸)] was stimulated by the mathematical results of Slepian and Pollak¹⁵ on prolate spheroidal functions. These functions are ideal for analyzing optical problems with rectangular or circular apertures and have several remarkable properties, one of which is that they are the eigenfunctions for imaging from one plane to another with such rectangular or circular apertures. (They also form the basis for the analysis of laser resonators with finite mirrors.) The eigenvalues are all approximately the same up to the number corresponding to the intuitive diffraction result [i.e., Gabor's result, expression (1)] after which they fall off rapidly. This prolate spheroidal eigenfunction work is reviewed extensively by Frieden.¹⁰ This work does therefore define an orthonormal basis set, at least for this class of problems, and through the introduction of eigenvalues has introduced the concept of what is essentially a coupling strength that is associated with a given basis function and a given aperture. (This body of research also goes on to discuss the effects of noise in cutting off the effective number

of degrees of freedom, although we do not discuss that topic here.) The eigenfunction approach can also be used to deduce useful results for the case of arbitrary apertures for which specific eigenfunction solutions are not known. Examples of such use include the study of atmospheric turbulence (see, e.g., Refs. 16–18).

Both the diffraction and the eigenfunction approaches discussed above deal with a situation in which we are communicating between an aperture on one surface and an aperture on another. They start with the commonly used optical approximation of the Huygens–Fresnel diffraction integral. Essentially, that approximation says that the wave field emitted from an aperture can be described if the surface within the aperture is considered to be filled with sources that are all of the same kind and of amplitudes proportional to the original wave amplitude within the aperture. Only the aperture is considered to have any sources. It is important to understand that, no matter how good a job we do of choosing the nature of these sources, this is still an approximation. The only truly exact answer that relates the wave outside a surface to the wave properties on a surface is the Kirchhoff integration of the wave equation, which requires integration over a complete closed surface (not merely an aperture) and also requires two kinds of sources on the surface whose amplitudes depend not only on the wave amplitude at the surface but also on its derivatives.

More recent results¹⁹ have shown that one kind of source can be used in the Kirchhoff integration, essentially in an extension of the Huygens principle, for the specific case of phase fronts on the surface, but this is still an approximation in the general case. The formal problems with the boundary conditions on the surface outside the aperture itself are also notorious. This is not to say that previous use of Huygens–Fresnel diffraction integrals is wrong in any sense in earlier research for the problems being addressed there. But if we wish to obtain absolutely rigorous results or if we wish to deal with, for example, volumes smaller than a wavelength in size or closer to one another than the lateral dimensions of the volumes, such approximations should be avoided.

Hence, although both the diffraction model and the previous eigenfunction approaches yield useful results and insights, they do not provide a complete answer to the questions of the communications modes and their strengths. The approach in this paper is to treat the communication as being between volumes and not between apertures in surfaces. This approach uses techniques of linear algebra similar to those used in the previous eigenfunction approaches but starts with formally exact integrals over volume sources. The method derived below can therefore be viewed as an extension to arbitrary volumes (as opposed to apertures in flat surfaces) of the eigenfunction approach. It is important to note that only by considering volumes (or complete surfaces that enclose volumes) can we achieve exact results. The method will reproduce the important results of



Fig. 2. Schematic diagram of the transmitting volume V_T , which contains sources $\Psi(\mathbf{r}_T)$ of waves, and the receiving volume V_R , which contains a wave field $\phi(\mathbf{r}_R)$ that arises from the sources in the transmitting volume.

both the previous eigenfunction and diffraction models. In addition, however, the present method also allows us to deduce results and insights that go beyond either of these previous approaches.

3. Formalism

A. Physical Problem and Wave Equation

First, we need to define the physical problem. We consider two volumes V_T and V_R , as sketched in Fig. 2. Note that we do not consider situations with an additional lens between the two volumes, i.e., a three-volume problem that could be considered as an extension to the present approach; it is not necessary to have any such lens to establish communications channels between two volumes, however. The discussion in this paper is restricted to the situation with only two volumes and no material bodies present. The approach can be extended to include material bodies (which makes the Green's functions different through the necessary inclusion of additional boundary conditions),² although only volumes in free space are considered here.

There are sources $\Psi(\mathbf{r}_T, t)$ in the transmitting volume V_T that generate waves $\Phi(\mathbf{r}_R, t)$ and obey the scalar wave equation

$$\nabla^2 \Phi(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial^2 \Phi(\mathbf{r}, t)}{\partial t^2} = -\Psi(\mathbf{r}, t), \quad (2)$$

where c is the wave velocity. We consider only monochromatic sources of the form

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) \exp(i\omega t) + \text{c.c.} \quad (3)$$

and scalar waves of the form

$$\Phi(\mathbf{r}, t) = \phi(\mathbf{r}) \exp(i\omega t) + \text{c.c.}, \quad (4)$$

where ω is the angular frequency. (This analysis could be extended to the general time-dependent case, although we do not do so in this paper.) We can usefully define the wave number as $k = c/\omega$. The scalar wave equation can now be written as the usual inhomogeneous Helmholtz equation:

$$\nabla^2 \phi(\mathbf{r}) + k^2 \phi(\mathbf{r}) = -\psi(\mathbf{r}). \quad (5)$$

Henceforth, we consider only the positive frequency component, i.e., $\exp(i\omega t)$, with the knowledge that we

can return to the real wave by adding the complex conjugate at the end of the calculation, as usual.

The Green's functions for Eq. (5), i.e., the possible waves resulting from a point source at position \mathbf{r}_1 , in otherwise empty space and with no other sources of waves are

$$G(\mathbf{r}, \mathbf{r}_1) = \frac{\exp(\pm ik|\mathbf{r} - \mathbf{r}_1|)}{4\pi|\mathbf{r} - \mathbf{r}_1|}. \quad (6)$$

The positive sign in Eq. (6) corresponds to inward-propagating waves (the advanced solution), and the negative sign corresponds to outgoing waves (the retarded solution). Because we are interested here in only the case of outgoing waves, we henceforth use only the retarded solution:

$$G(\mathbf{r}, \mathbf{r}_1) = \frac{\exp(-ik|\mathbf{r} - \mathbf{r}_1|)}{4\pi|\mathbf{r} - \mathbf{r}_1|}. \quad (7)$$

By using this Green's function [Eq. (7)], we can therefore formally add together all the waves from all the sources within volume V_T to obtain the resultant wave:

$$\phi(\mathbf{r}) = \int_{V_T} G(\mathbf{r}, \mathbf{r}_T)\psi(\mathbf{r}_T)d^3\mathbf{r}_T. \quad (8)$$

B. Mathematical Formalism

Now we can set up the mathematical formalism for the problem. To describe an arbitrary source function in the transmitting volume V_T , we choose a complete orthonormal basis set of functions that is defined within V_T , namely, $a_{T1}(\mathbf{r}_T)$, $a_{T2}(\mathbf{r}_T)$, $a_{T3}(\mathbf{r}_T)$, \dots . Similarly, to describe an arbitrary wave in the receiving volume V_R , we choose another complete orthonormal basis set of functions that is defined within V_R , namely, $a_{R1}(\mathbf{r}_R)$, $a_{R2}(\mathbf{r}_R)$, $a_{R3}(\mathbf{r}_R)$, \dots . At the moment it is not important to know what the sets of functions $a_{Ti}(\mathbf{r}_T)$ and $a_{Rj}(\mathbf{r}_R)$ are as long as they are complete orthonormal basis sets in their respective volumes. These basis sets satisfy the usual orthonormality relations:

$$\int_{V_T} a_{Tm}(\mathbf{r}_T)a_{Tn}^*(\mathbf{r}_T)d^3\mathbf{r}_T = \delta_{mn}, \quad (9)$$

$$\int_{V_R} a_{Rm}(\mathbf{r}_R)a_{Rn}^*(\mathbf{r}_R)d^3\mathbf{r}_R = \delta_{mn}, \quad (10)$$

where δ_{mn} is the Kronecker delta ($\delta_{mn} = 1$, $m = n$; $\delta_{mn} = 0$, $m \neq n$).

Using these basis sets [Eqs. (9) and (10)], we can expand $\psi(\mathbf{r}_T)$ and $\phi(\mathbf{r}_R)$ to obtain

$$\psi(\mathbf{r}_T) = \sum_i b_i a_{Ti}(\mathbf{r}_T), \quad (11)$$

$$\phi(\mathbf{r}_R) = \sum_j d_j a_{Rj}(\mathbf{r}_R). \quad (12)$$

Using these expansions [Eqs. (11) and (12)], we can formally rewrite Eq. (8) as

$$d_j = \sum_i g_{ji} b_i, \quad (13)$$

where

$$g_{ji} = \int_{V_R} \int_{V_T} a_{Rj}^*(\mathbf{r}_R)G(\mathbf{r}_R, \mathbf{r}_T)a_{Ti}(\mathbf{r}_T)d^3\mathbf{r}_T d^3\mathbf{r}_R. \quad (14)$$

The variables g_{ji} can now be thought of as coupling coefficients between the transmission from mode i in volume V_T and the reception by mode j in volume V_R . We can also think of g_{ji} as being a kind of transmittance that relates the wave amplitude in one volume to the source amplitude in the other.

Note that Eq. (13) [together with the expansions in Eqs. (11) and (12)] is exactly equivalent to Eq. (8). Instead of expressing the value of the wave at a given point in V_R directly, we have given the coefficients d_j for the basis functions $a_{Rj}(\mathbf{r}_R)$. It can be conceptually useful to think of Eq. (13) in matrix terms, in which case we can rewrite it as

$$\begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} & g_{13} & \cdots \\ g_{21} & g_{22} & g_{23} & \cdots \\ g_{31} & g_{32} & g_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \end{bmatrix}, \quad (15)$$

where it should be noted that this matrix will be infinite, or we can rewrite with ϕ and ψ as the vectors of elements d_1, d_2, d_3, \dots and b_1, b_2, b_3, \dots , respectively, and Γ_{RT} as the matrix with elements g_{ji} :

$$\phi = \Gamma_{RT}\psi. \quad (16)$$

(In practice, the matrix form is the one that we will likely use in actual calculations.) We can now view Γ_{RT} in general as the communications operator between the transmitting and the receiving volumes.

C. Interpretation and Measurement of the Coupling Coefficients g_{ji}

Before proceeding to the main results presented in this paper, we should formally clarify the physical meaning of the coupling coefficients g_{ji} . They have specific meanings in terms of the power-transfer coefficients between the two volumes.

One way of completing the transfer of energy into the receiving volume from the transmitting volume is to set up a receiving source $\Psi_R(\mathbf{r}_R, t)$ in the receiving volume V_R . Our goal is to have the wave $\Phi(\mathbf{r}_R, t)$ transfer energy to this source. To keep matters simple, we formally require that the source $\Psi_R(\mathbf{r}_R, t)$ be weak so that it does not itself generate any wave that is significant compared with the wave $\Phi(\mathbf{r}_R, t)$ that is arriving from sources in the transmitting volume. Hence the total wave in the receiving volume is, in the limit of weak $\Psi_R(\mathbf{r}_R, t)$, the original wave $\Phi(\mathbf{r}_R, t)$.

In the wave equation [Eq. (2)] Ψ can be viewed as corresponding to force (per unit volume) and Φ as corresponding to displacement in the wave. As a

result, the work done per unit time per unit volume by the force $\Psi_R(\mathbf{r}_R, t)$ on the wave is the usual product of force and velocity, i.e.,

$$W = \Psi_R(\mathbf{r}_R, t) \frac{\partial \Phi(\mathbf{r}_R, t)}{\partial t}. \quad (17)$$

Averaging over complete half-cycles of the wave (or over times that are long compared with a cycle) for our monochromatic wave yields

$$W = 2\omega \text{Im}[\psi_R(\mathbf{r}_R)\phi^*(\mathbf{r}_R)]. \quad (18)$$

The total average power transferred from the *wave* to the *source* in V_R is therefore

$$P = - \int_{V_R} 2\omega \text{Im}[\psi_R(\mathbf{r}_R)\phi^*(\mathbf{r}_R)] d^3\mathbf{r}_R. \quad (19)$$

If we consider the specific case of a normalized source in V_T that corresponds to one of our basis functions, $\psi(\mathbf{r}_T) = a_{T_i}(\mathbf{r}_T)$, in the transmitting volume and choose a receiving-source function that is proportional to the complex conjugate of one of the basis functions, $a_{R_j}(\mathbf{r}_R)$, i.e., $\psi_R(\mathbf{r}_R) = Ca_{R_j}^*(\mathbf{r}_R)$, we obtain

$$\begin{aligned} P &= -2\omega C \int_{V_R} \text{Im}[a_{R_j}^*(\mathbf{r}_R)\phi(\mathbf{r}_R)] d^3\mathbf{r}_R \\ &= -2\omega C \text{Im}(g_{ji}). \end{aligned} \quad (20)$$

If we choose $\psi_R(\mathbf{r}_R) = iCa_{R_j}^*(\mathbf{r}_R)$, we have

$$\begin{aligned} P &= -2\omega C \int_{V_R} \text{Im}[ia_{R_j}^*(\mathbf{r}_R)\phi(\mathbf{r}_R)] d^3\mathbf{r}_R \\ &= -2\omega C \text{Re}(g_{ji}). \end{aligned} \quad (21)$$

Hence the real Re and the imaginary Im parts of g_{ji} are essentially the power-coupling coefficients between the source function $a_{T_i}(\mathbf{r}_T)$ in V_T and a receiving-source function $a_{R_j}(\mathbf{r}_R)$ in V_R for the two possible orthogonal phases.

An alternative way, which is mathematically very similar to the preceding power-transfer method, to detect the signals in the receiving volume is to measure the wave at every point in the volume, weight these measurements by a weighting function $w(\mathbf{r}_R)$, and add (i.e., integrate) to obtain a result M (which is, in general, complex). If we choose $w(\mathbf{r}_R) = a_j^*(\mathbf{r}_R)$, then, for a normalized source $\psi(\mathbf{r}_T) = a_{T_i}(\mathbf{r}_T)$, the result is

$$\begin{aligned} M &= \int_{V_R} a_j^*(\mathbf{r}_R)\phi(\mathbf{r}_R) d^3\mathbf{r}_R \\ &= g_{ji}. \end{aligned} \quad (22)$$

4. Sum Rule

To this point all we have done mathematically is to recast the problem in terms of basis sets, coupling coefficients, and an abstract operator and matrix notation. Now we can proceed to use this formalism to

derive a key result, which is a sum rule on the strengths of the coupling coefficients. Specifically, we can state the following theorem:

$$\gamma_{RT} \equiv \sum_{i,j} |g_{ji}|^2 = \frac{1}{(4\pi)^2} \int_{V_R} \int_{V_T} \frac{1}{|\mathbf{r}_R - \mathbf{r}_T|^2} d^3\mathbf{r}_T d^3\mathbf{r}_R. \quad (23)$$

(Here we have, for convenience, defined the quantity γ_{RT} , which we can refer to as the magnitude or the norm of the communications operator.) Theorem (23) can be proved by expansion of $G(\mathbf{r}_R, \mathbf{r}_T)$ in the basis sets. First, we expand $G(\mathbf{r}_R, \mathbf{r}_T)$ in the complete set of functions $a_{T_i}^*(\mathbf{r}_T)$. (Note that, given that the set $a_{T_i}(\mathbf{r}_T)$ is complete, it follows that the set $a_{T_i}^*(\mathbf{r}_T)$ is also complete.) Hence we obtain for points \mathbf{r}_R in V_R and \mathbf{r}_T in V_T

$$G(\mathbf{r}_R, \mathbf{r}_T) = \sum_i v_i(\mathbf{r}_R) a_{T_i}^*(\mathbf{r}_T), \quad (24)$$

where

$$v_i(\mathbf{r}_R) = \int_{V_T} G(\mathbf{r}_R, \mathbf{r}_T) a_{T_i}(\mathbf{r}_T) d^3\mathbf{r}_T. \quad (25)$$

Further expanding $v_i(\mathbf{r}_R)$ by use of the complete set $a_{R_j}(\mathbf{r}_R)$ yields

$$v_i(\mathbf{r}_R) = \sum_j g_{ji} a_{R_j}(\mathbf{r}_R), \quad (26)$$

where we made use of the definition of g_{ji} from Eq. (14). Consequently, by using Eqs. (24) and (26), we have

$$G(\mathbf{r}_R, \mathbf{r}_T) = \sum_{i,j} g_{ji} a_{R_j}(\mathbf{r}_R) a_{T_i}^*(\mathbf{r}_T). \quad (27)$$

[The steps from Eq. (24) to Eq. (27) technically correspond to a similarity transformation from a continuous to a discrete basis, which constitutes a standard approach that is given here completely for clarity.] Now we have

$$\begin{aligned} |G(\mathbf{r}_R, \mathbf{r}_T)|^2 &= \left[\sum_{i,j} g_{ji} a_{R_j}(\mathbf{r}_R) a_{T_i}^*(\mathbf{r}_T) \right] \\ &\quad \times \left[\sum_{p,q} g_{qp}^* a_{R_q}^*(\mathbf{r}_R) a_{T_p}(\mathbf{r}_T) \right]. \end{aligned} \quad (28)$$

Integrating the right-hand side of Eq. (28) over both volumes V_T and V_R and using the orthonormality relations of Eqs. (9) and (10) of the basis sets, we obtain

$$\int_{V_R} \int_{V_T} |G(\mathbf{r}_R, \mathbf{r}_T)|^2 d^3\mathbf{r}_T d^3\mathbf{r}_R = \sum_{i,j} |g_{ji}|^2. \quad (29)$$

From Eq. (7), we can deduce that

$$|G(\mathbf{r}_R, \mathbf{r}_T)|^2 = \frac{1}{(4\pi)^2 |\mathbf{r}_R - \mathbf{r}_T|^2}, \quad (30)$$

and, finally, from Eqs. (29) and (30), we deduce the sum rule [Eq. (23)].

Equation (23) has the remarkable property of the sum of the modulus squared of the coupling coefficients depending on *only* the shapes and the relative positions of the two volumes. This result is also true independently of the basis sets we choose in each volume (as long as they are complete and orthonormal) because we placed no restrictions on the choice of basis sets.

5. Communications Modes

So far, we have used only arbitrary basis sets for the functions in the two volumes V_T and V_R . We can reasonably ask whether there is a natural pair of basis sets for this problem in which the mathematical results become particularly simple and the physical interpretation becomes clearer. In fact, for any pair of volumes there are such natural sets, and we can view these as defining communications modes. We now proceed to derive these sets. We first take a heuristic approach that shows that the communications modes are the pairs of functions that diagonalize the communications operator. Second, we start afresh mathematically to derive the eigenvalue problems whose eigenfunctions will yield the communications modes. The mathematical approach is essentially similar to that of the singular-value decomposition of matrices (although we have to derive the results in integral-equation form for formal reasons that are related to the infinite basis sets).

A. Heuristic Approach to Communications Modes

It is immediately clear from Eq. (23) that there is some maximum value of $|g_{ji}|^2$; in the worst case $|g_{ji}|^2$ cannot possibly exceed γ_{RT} . Hence there must be some pair $[\psi_1(\mathbf{r}_T), \phi_1(\mathbf{r}_R)]$ of (normalized) transmitting and receiving functions that are most strongly coupled, i.e., a pair for which the coupling coefficient g has the largest squared modulus $|g|^2$, and we can label this coupling coefficient g_1 .

There are many ways with which we could attempt to evaluate such a pair of functions, including, for example, some variational method to maximize $|g|^2$. (It is also conceivable, especially in situations with high symmetry, that there is more than one such pair with the same, i.e., the largest value of $|g|^2$. Such degeneracy causes no problems here, and it does not matter which of these pairs of degenerate functions we choose first. We know, anyway, that the number of such degenerate-function pairs must be finite because of the sum rule.) We now choose $\psi_1(\mathbf{r}_T)$ as the first member of a new basis set for the transmitting volume V_T and similarly choose $\phi_1(\mathbf{r}_R)$ as the first member of a new basis set for the receiving volume V_R .

We can now find the second members $[\psi_2(\mathbf{r}_T), \phi_2(\mathbf{r}_R)]$ of the basis sets by a similar procedure, except at this point, we also require that these (normalized) functions be orthogonal to the corresponding first members. The corresponding coupling coefficient g_2 will have the next-largest value of $|g|^2$. (In the case of degenerate functions, $|g_1|^2 = |g_2|^2$.) We can continue to proceed in this manner to find all the other

members of the basis set, requiring that each successive member of each set be orthogonal to all the previous members of that set.

An important question is whether the off-diagonal coupling coefficients (i.e., $g_{ji}, j \neq i$) are zero when we use this basis. In fact, they are all zero, as is easily proved. Essentially, if the off-diagonal elements were not zero, we would be able to find functions with larger coupling coefficients than those of these basis sets, which is impossible by definition. A second important question is whether this set of communications modes is unique (except for the usual arbitrariness associated with linear combinations of degenerate functions). In fact, they are (within a phase factor); this point will become clear from the formal mathematical solution described in Subsection 5.B. In this new representation, therefore, the matrix Γ_{RT} is diagonal with its diagonal elements being the coupling coefficients g_1, g_2, g_3, \dots , arranged in descending order of their squared magnitudes.

The communications modes, then, are those pairs of functions $[\psi_j(\mathbf{r}_T), \phi_j(\mathbf{r}_R)]$ that are unique (within a phase factor) except for the arbitrary linear combinations of possible degenerate solutions that, when used as the complete orthogonal basis functions for their respective volumes, diagonalize the communications operator Γ_{RT} . Such functions also yield the largest possible coupling coefficients.

B. Formal Eigenfunction Solution for Communications Modes

The maximization procedure described in Subsection 5.A is not necessarily a convenient one in practice for calculating the functions. Also, such an approach does not prove the completeness and the uniqueness of the solutions. Hence it is useful to recast the solution of this problem in terms of eigenfunction problems; then we can use standard linear-algebra techniques to arrive at the solutions. In fact, this kind of problem is well known mathematically, and some aspects of it have already been applied (in the two-dimensional case and on the basis of diffraction-theory approximations) to the theory of optical propagation in turbulent atmospheres.¹⁶⁻¹⁸ In matrix form the solutions we seek would reduce to the results of a singular-value decomposition of the matrix Γ_{RT} of the coefficients g_{ji} , although, to justify the results here, we need some of the results of functional analysis to be applied to integral equations. Thus we briefly derive the solution in the integral-equation form.

Formally, we wish to deduce which (normalized) source function $\psi(\mathbf{r}_T)$ yields the (unnormalized) wave $\phi(\mathbf{r}_R)$ with the largest magnitude in volume V_R , where, by magnitude, we mean the quantity

$$|g|^2 = \int_{V_R} \phi^*(\mathbf{r}_R)\phi(\mathbf{r}_R)d^3\mathbf{r}. \quad (31)$$

[Note that, in Eq. (31), g must, by definition from Eq. (14), then be the coupling coefficient between the nor-

malized source function $\psi(\mathbf{r}_T)$ and the corresponding normalized wave function $\phi_N(\mathbf{r}_R)$ of the same form as $\phi(\mathbf{r}_R)$, i.e., $\phi(\mathbf{r}_R) = g\phi_N(\mathbf{r}_R)$.]

Substituting from Eq. (8) into Eq. (31) yields

$$|g|^2 = \int_{V_R} \int_{V_T} \int_{V_T} G^*(\mathbf{r}_R, \mathbf{r}'_T) \psi^*(\mathbf{r}'_T) G(\mathbf{r}_R, \mathbf{r}_T) \times \psi(\mathbf{r}_T) d^3\mathbf{r}'_T d^3\mathbf{r}_T d^3\mathbf{r}_R, \quad (32)$$

that is,

$$|g|^2 = \int_{V_T} \psi^*(\mathbf{r}'_T) \int_{V_T} K(\mathbf{r}'_T, \mathbf{r}_T) \psi(\mathbf{r}_T) d^3\mathbf{r}'_T d^3\mathbf{r}_T, \quad (33)$$

where

$$K(\mathbf{r}'_T, \mathbf{r}_T) = \int_{V_R} G^*(\mathbf{r}_R, \mathbf{r}'_T) G(\mathbf{r}_R, \mathbf{r}_T) d^3\mathbf{r}_R. \quad (34)$$

Note that $K(\mathbf{r}'_T, \mathbf{r}_T)$ is Hermitian, i.e.,

$$K(\mathbf{r}'_T, \mathbf{r}_T) = K^*(\mathbf{r}_T, \mathbf{r}'_T). \quad (35)$$

Also, because V_R and V_T are completely separate and finite volumes, the terms $|\mathbf{r}_R - \mathbf{r}_T|$ and $|\mathbf{r}_R - \mathbf{r}'_T|$ are never zero, and $K(\mathbf{r}'_T, \mathbf{r}_T)$ can have no singularities and is a continuous, bounded function. Additionally, with V_T finite, we have

$$\int_{V_T} \int_{V_T} |K(\mathbf{r}'_T, \mathbf{r}_T)|^2 d^3\mathbf{r}'_T d^3\mathbf{r}_T < \infty.$$

This boundedness of the integral and the integrand comprises sufficient conditions to make the linear map that corresponds to this kernel $K(\mathbf{r}'_T, \mathbf{r}_T)$ technically compact,²⁰ which means that, in practice, $K(\mathbf{r}'_T, \mathbf{r}_T)$ can be expanded bilinearly in a complete set of functions $a_{Ti}(\mathbf{r}_T)$:

$$K(\mathbf{r}'_T, \mathbf{r}_T) = \sum_{j,i} k_{ji} a_{Tj}(\mathbf{r}'_T) a_{Ti}^*(\mathbf{r}_T). \quad (36)$$

The eigenfunctions of such a compact Hermitian kernel themselves form a complete set for functions in V_T and have real eigenvalues,²⁰ and it is also a standard result that the Hermitian form [Eq. (33)] is maximized when $\psi(\mathbf{r}_T)$ is chosen as the eigenfunction of $K(\mathbf{r}'_T, \mathbf{r}_T)$ with the largest eigenvalue,^{20,21} i.e., the function $\psi(\mathbf{r}_T)$ that solves the integral equation

$$|g|^2 \psi(\mathbf{r}'_T) = \int_{V_T} K(\mathbf{r}'_T, \mathbf{r}_T) \psi(\mathbf{r}_T) d^3\mathbf{r}_T \quad (37)$$

with the largest value of $|g|^2$. Indeed, the set of orthogonal functions with the largest successive values of $|g|^2$ is the set of eigenfunctions $\psi_1(\mathbf{r}_T)$, $\psi_2(\mathbf{r}_T)$, $\psi_3(\mathbf{r}_T)$, . . . , arranged in descending order of their eigenvalues $|g_1|^2$, $|g_2|^2$, $|g_3|^2$, [This result becomes relatively obvious if we choose the eigenfunctions for the expansion in Eq. (36), which yields a diagonal expansion (i.e., $k_{j \neq i} = 0$) with the eigenvalues as the diagonal expansion coefficients (i.e., $k_{j=i} = |g_i|^2$).]

For any given normalized source function $\psi_n(\mathbf{r}_T)$ there is a corresponding (unnormalized) wave $\phi_{nU}(\mathbf{r}_R)$ generated in the receiving volume, which allows us to define normalized waves as $\phi_n(\mathbf{r}_R) = \phi_{nU}(\mathbf{r}_R)/g_n$. We know that the proportionality constant here is g_n because $|g_n|^2$ is the value that we deduced for $|g|^2$ for the case of source function $\psi_n(\mathbf{r}_T)$ and $|g|^2$ is, by definition, the magnitude (squared) of the generated wave from Eq. (33). Note, incidentally, that the phase of g_n is arbitrary because wave $\phi_n(\mathbf{r}_R)$ is itself arbitrary within a phase factor. It is not immediately obvious, but these waves $\phi_n(\mathbf{r}_R)$ also form an orthonormal complete set in the receiving volume V_R because, as we now show, they are also the eigenfunctions of a compact Hermitian kernel. We have by definition

$$g_n \phi_n(\mathbf{r}_R) = \phi_{nU}(\mathbf{r}_R) = \int_{V_T} G(\mathbf{r}_R, \mathbf{r}_T) \psi_n(\mathbf{r}_T) d^3\mathbf{r}_T. \quad (38)$$

Now, by using Eqs. (34) and (37), changing the order of the integrations, and interchanging the use of the variable names \mathbf{r}_T and \mathbf{r}'_T , we have

$$\int_{V_R} G^*(\mathbf{r}_R, \mathbf{r}_T) \left[\int_{V_T} G(\mathbf{r}_R, \mathbf{r}'_T) \psi_n(\mathbf{r}'_T) d^3\mathbf{r}'_T \right] d^3\mathbf{r}_R = |g_n|^2 \psi_n(\mathbf{r}_T) = \int_{V_R} G^*(\mathbf{r}_R, \mathbf{r}_T) g_n \phi_n(\mathbf{r}_R) d^3\mathbf{r}_R, \quad (39)$$

where we have made use of the definition of $\phi_n(\mathbf{r}_R)$ from Eq. (38). Hence we have

$$g_n^* \psi_n(\mathbf{r}_T) = \int_{V_R} G^*(\mathbf{r}_R, \mathbf{r}_T) \phi_n(\mathbf{r}_R) d^3\mathbf{r}_R. \quad (40)$$

Substituting Eq. (40) into Eq. (38) yields

$$|g_n|^2 \phi_n(\mathbf{r}_R) = \int_{V_R} J(\mathbf{r}_R, \mathbf{r}'_R) \phi_n(\mathbf{r}'_R) d^3\mathbf{r}'_R, \quad (41)$$

where

$$J(\mathbf{r}_R, \mathbf{r}'_R) = \int_{V_T} G(\mathbf{r}_R, \mathbf{r}_T) G^*(\mathbf{r}'_R, \mathbf{r}_T) d^3\mathbf{r}_T. \quad (42)$$

Equation (42) is a compact Hermitian kernel by the same arguments as given above and therefore has a complete set of eigenfunctions with real eigenvalues.

Note that this problem of communicating from one volume to another has turned out to be a remarkably symmetrical and reciprocal one. There are two sets of eigenfunctions, one set for each volume, with a one-to-one correspondence between the members of the sets in the two volumes and the same eigenvalues for the eigenfunctions in the different volumes. It is also true, incidentally, that, by taking the complex conjugate of Eq. (40), we have

$$g_n \psi_n^*(\mathbf{r}_T) = \int_{V_R} G(\mathbf{r}_R, \mathbf{r}_T) \phi_n^*(\mathbf{r}_R) d^3\mathbf{r}_R, \quad (43)$$

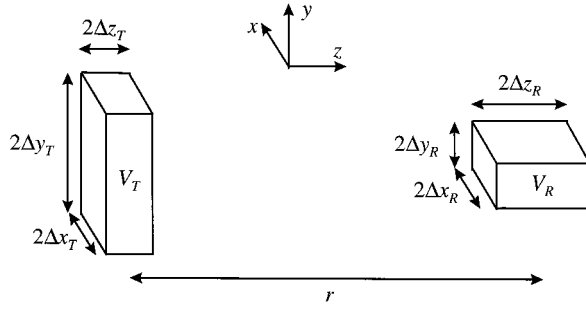


Fig. 3. Transmitting volume V_T , receiving volume V_R , rectangular prism volumes, and their coordinates and dimensions.

which means that, if we had a source of the form $\phi_n^*(\mathbf{r}_R)$ in volume V_R , it would generate a wave of the form $\psi_n^*(\mathbf{r}_R)$ in volume V_T . Such a source could be generated by the well-known process of phase conjugation, whereby, usually through some nonlinear optical process, we generate a source that is proportional to the (spatial) complex conjugate of the wave in a volume. We return to this point below in Subsection 5.F.

C. Communications Modes for Two Rectangular Prism Volumes

Now let us consider the specific case of the communications modes between two rectangular prism volumes (volumes in which each face is rectangular). This case gives us an explicit connection to the eigenfunction approach presented in Subsection 5.B and yields analytic results for cases of practical interest.

Consider two rectangular prism volumes V_T and V_R that are oriented along the same axis and are a distance r apart (center to center), as shown in Fig. 3. The volumes are of size $2\Delta x_T$, $2\Delta y_T$, and $2\Delta z_T$, in the x , the y , and the z directions, respectively, for V_T and are similar for V_R . We presume that the volumes are far apart compared with their sizes:

$$r \gg 2\Delta x_T, 2\Delta y_T, 2\Delta z_T, 2\Delta x_R, 2\Delta y_R, 2\Delta z_R. \quad (44)$$

We can formally find the communications modes by solving Eqs. (37) and (41). Explicitly, from Eq. (37), we have

$$|g|^2 \psi(\mathbf{r}_T) = \int_{V_T} \left[\int_{V_R} \frac{\exp(ik|\mathbf{r}_R - \mathbf{r}_T|) \exp(-ik|\mathbf{r}_R - \mathbf{r}'_T|)}{(4\pi)^2 |\mathbf{r}_R - \mathbf{r}_T| |\mathbf{r}_R - \mathbf{r}'_T|} d^3 \mathbf{r}_R \right] \psi(\mathbf{r}'_T) d^3 \mathbf{r}'_T. \quad (45)$$

Because of condition (44), we can approximately replace $|\mathbf{r}_R - \mathbf{r}_T|$ and $|\mathbf{r}_R - \mathbf{r}'_T|$ with r in the denominator in Eq. (45). In the exponent, for example, we have

$$\begin{aligned} |\mathbf{r}_R - \mathbf{r}_T| &= [(r + z_R - z_T)^2 + (x_R - x_T)^2 + (y_R - y_T)^2]^{1/2} \\ &\cong r + (z_R - z_T) + \frac{(x_R - x_T)^2}{2r} + \frac{(y_R - y_T)^2}{2r}, \end{aligned} \quad (46)$$

where x_R , y_R , and z_R are the coordinates relative to the center of V_R and similarly x_T , y_T , and z_T are the coordinates relative to the center of V_T , which can be viewed as a paraxial approximation. Hence

$$\begin{aligned} &\exp(ik|\mathbf{r}_R - \mathbf{r}_T|) \exp(-ik|\mathbf{r}_R - \mathbf{r}'_T|) \\ &\cong \exp \left(-ik \left[z_T - \frac{1}{2r} (x_T^2 + y_T^2) - z'_T + \frac{1}{2r} (x_T'^2 + y_T'^2) \right. \right. \\ &\quad \left. \left. + \frac{1}{r} [x_R(x_T - x'_T) + y_R(y_T - y'_T)] \right] \right). \end{aligned} \quad (47)$$

Now we can choose to write

$$\psi(\mathbf{r}_T) = F_T(\mathbf{r}_T) \beta_T(\mathbf{r}_T), \quad (48)$$

where

$$F_T(\mathbf{r}_T) = \exp \left\{ -ik \left[z_T - \frac{1}{2r} (x_T^2 + y_T^2) \right] \right\}. \quad (49)$$

The term $F_T(\mathbf{r}_T)$ can be described as a focusing function; it corresponds to a spherical wave that is centered on the receiving volume (see Fig. 4). The purpose of this separation [Eq. (48)] is to take out the underlying spherical focusing of the source from the mathematical problem that we now solve for the function $\beta_T(\mathbf{r}_T)$.

With the definitions of Eqs. (48) and (49), we can now rewrite Eq. (45) with our approximation (47) as

$$\begin{aligned} |g|^2 \beta_T(\mathbf{r}_T) &= \frac{1}{(4\pi r)^2} \int_{V_T} \int_{V_R} \exp \left\{ -\frac{ik}{r} [x_R(x_T - x'_T) \right. \\ &\quad \left. + y_R(y_T - y'_T)] \right\} \beta(\mathbf{r}'_T) d^3 \mathbf{r}_R d^3 \mathbf{r}'_T. \end{aligned} \quad (50)$$

Now we postulate that we can write

$$\beta_T(\mathbf{r}_T) = \alpha_{T_x}(x_T) \alpha_{T_y}(y_T) \alpha_{T_z}(z_T). \quad (51)$$

This separation [Eq. (51)] is justified *a posteriori* because we show that functions of this form are eigenfunctions of Eq. (50). With the form of Eq. (51), the problem can be rewritten as

$$\begin{aligned} |g|^2 \alpha_{T_x}(x_T) \alpha_{T_y}(y_T) \alpha_{T_z}(z_T) &= \frac{1}{(4\pi r)^2} \int_{V_T} \int_{V_R} \\ &\times \left(\exp \left\{ -\frac{ik}{r} [x_R(x_T - x'_T) + y_R(y_T - y'_T)] \right\} \right. \\ &\quad \left. \times \alpha_{T_x}(x'_T) \alpha_{T_y}(y'_T) \alpha_{T_z}(z'_T) \right) d^3 \mathbf{r}_R d^3 \mathbf{r}'_T, \end{aligned} \quad (52)$$

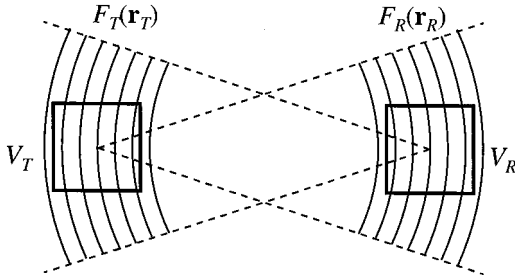


Fig. 4. Illustration of the focusing functions $F_T(\mathbf{r}_T)$ and $F_R(\mathbf{r}_R)$ that arise in the analysis of communications between rectangular prism volumes.

which is now separable into three eigenfunction equations in the three directions:

$$\eta_{Tx}\alpha_{Tx}(x_T) = \int_{-\Delta x_T}^{\Delta x_T} \left\{ \int_{-\Delta x_R}^{\Delta x_R} \exp\left[-\frac{ik}{r}x_R(x_T - x'_T)\right] dx_R \right\} \times \alpha_{Tx}(x'_T) dx'_T. \quad (53)$$

The equation for the y direction is similar, and for the z direction it is

$$\eta_{Tz}\alpha_{Tz}(z_T) = \int_{-\Delta z_T}^{\Delta z_T} \int_{-\Delta z_R}^{\Delta z_R} \alpha_{Tz}(z'_T) dz_R dz'_T, \quad (54)$$

where

$$\eta_{Tx}\eta_{Ty}\eta_{Tz} = (4\pi r)^2 |g|^2. \quad (55)$$

The eigenequation [Eq. (54)] for the z direction is trivially solved. The left-hand side of the equation depends on z_T , whereas the right-hand side does not. The only solution to Eq. (54) is for $\alpha_{Tz}(z_T)$ to be a constant, in which case

$$\eta_{Tz} = (2\Delta z_R)(2\Delta z_T). \quad (56)$$

[Note that we now do not have a complete set of eigenfunctions for the z direction, which is a consequence of the approximations that we have made here. The resultant kernel in Eq. (54) is itself a constant, which therefore maps all functions onto constants and thus is a many-to-one mapping that cannot be inverted. A full solution of this problem without approximations would presumably result in a set of functions in the z direction.]

The situation for the x and the y directions is more interesting. The integration over x_R can be performed in Eq. (53), yielding

$$\begin{aligned} & \int_{-\Delta x_R}^{\Delta x_R} \exp\left[-\frac{ik}{r}x_R(x_T - x'_T)\right] dx_R \\ &= \frac{r}{ik(x_T - x'_T)} \left\{ \exp\left[\frac{ik}{r}\Delta x_R(x_T - x'_T)\right] \right. \\ & \quad \left. - \exp\left[-\frac{ik}{r}\Delta x_R(x_T - x'_T)\right] \right\} \\ &= \frac{2\pi\Delta x_R}{\Omega_{Tx}} \frac{\sin[\Omega_{Tx}(x_T - x'_T)]}{\pi(x_T - x'_T)}, \quad (57) \end{aligned}$$

where

$$\Omega_{Tx} = \frac{k\Delta x_R}{r}. \quad (58)$$

Note, incidentally, that

$$\frac{2\pi\Delta x_R}{\Omega_{Tx}} = \frac{2\pi r}{k} = \lambda r, \quad (59)$$

where λ is the (free-space) wavelength. Equation (57) leads to a rewriting of Eq. (53) to yield

$$\nu_{Tx}\alpha_{Tx}(x_T) = \int_{-\Delta x_T}^{\Delta x_T} \frac{\sin[\Omega_{Tx}(x_T - x'_T)]}{\pi(x_T - x'_T)} \alpha_{Tx}(x'_T) dx'_T, \quad (60)$$

where

$$\nu_{Tx} = \frac{\eta_{Tx}}{\lambda r}. \quad (61)$$

Now Eq. (60) is the defining eigenequation for the prolate spheroidal wave functions.¹⁰ Hence we have found the desired solutions to Eq. (45) (within the approximations made). Explicitly, we have

$$\psi_{n_x n_y}(\mathbf{r}_T) = F_T(\mathbf{r}_T)\alpha_{n_x}(x_T)\alpha_{n_y}(y_T), \quad (62)$$

where $\alpha_{n_x}(x_T)$ is the n_x th (linear) prolate spheroidal function with a bandwidth Ω_{Tx} , a scale Δx_T , and an eigenvalue ν_{n_x} ; the situation is similar for the y direction. These sets of functions are known to be complete over the associated intervals $\pm\Delta x$ and $\pm\Delta y$, respectively.

To relate to the standard notation of Flammer²² for prolate spheroidal wave functions, we scale the lengths in units of Δx_T (and similarly for the y and the z directions) to obtain instead of (but equivalent to) Eq. (60)

$$\nu_{Tn_x} S_{0n_x}(c_x, \xi_T) = \int_{-1}^1 \frac{\sin[c_x(\xi_T - \xi'_T)]}{\pi(\xi_T - \xi'_T)} S_{0n_x}(c_x, \xi'_T) d\xi'_T, \quad (63)$$

where

$$\xi_T = \frac{x_T}{\Delta x_T}, \quad (64)$$

the dimensionless parameter c_x is given by

$$c_x = \Delta x_T \Omega_{Tx}, \quad (65)$$

and $S_{0m}(c, \xi)$ is the $(0, m)$ th angular prolate spheroidal function.

The eigenvalues ν_{Tn_x} are known to have the properties¹⁰

$$1 \geq \nu_0 > \nu_1 > \nu_2 > \dots > 0, \quad (66)$$

and for small nx the ν_{Tnx} fall off slowly until they reach a critical value

$$nx_{\text{crit}} = \frac{2\Delta x_T \Omega_x}{\pi} = \frac{2}{\pi} c_x = \frac{(2\Delta x_T)(2\Delta x_R)}{\lambda r}, \quad (67)$$

after which they fall off rapidly. We refer to this critical value as the effective number of degrees of freedom (in the x direction) because, as we see below, this is the number of degrees of freedom that we would expect from the usual diffraction models of communications between surfaces.

Note, however, that nx_{crit} can be less than 1 here if sizes Δx_T and Δx_R are too small, the meaning of which is that the communications modes are not as well connected as they might otherwise be for volumes that are spaced such a distance apart. Again, this situation agrees with our intuitive notions of diffraction: If the volumes are so small that, according to a simple diffraction model, the wave from the source volume cannot all be focused efficiently onto the receiving volume, then the communications mode is not well connected.

Incidentally, Eq. (67) can be viewed as having a simple physical meaning. The ratio $2\Delta x_R/\lambda$ is the number of wavelengths that correspond to the size of the receiving volume (in this x direction), and the ratio $2\Delta x_T/r$ is the angle subtended by the transmitting volume (in this x direction). Expression (67) therefore corresponds to the notion that the number of resolvable spots is the area expressed in units of wavelength and weighted by the subtended angle, as we would expect from the results of diffraction theory. Equivalently, $\lambda/2\Delta x_T$ is the diffraction angle of the radiation that emanates from the aperture of the transmitting volume (in the x direction), $2\Delta x_R/r$ is the angle subtended by the receiving volume at the transmitting volume (in the x direction), and nx_{crit} is the number of times this diffraction angle can be fitted into this subtended angle. It is also clear from expression (67) that we can view the size, the diffraction angle, and the subtended angle from the perspective of the other (y and z) volumes and obtain the same answer.

Thus we will find, equivalently to expression (67),

$$\eta_{Tx} \leq \lambda r, \quad (68)$$

with η_{Tx} staying close to λr up to the critical number nx_{crit} . If we take expression (68) for η_{Ty} , the equivalent expression for η_{Tz} , and Eq. (56) for η_{Tz} and substitute these into Eq. (55), we have

$$|g|^2 \leq \frac{\lambda^2(2\Delta z_T)(2\Delta z_R)}{(4\pi r)^2}. \quad (69)$$

Now, for the case in which the volumes are far apart compared with their dimensions the sum rule [Eq. (23)] reduces to

$$\sum |g|^2 = \frac{V_T V_R}{(4\pi r)^2}. \quad (70)$$

Hence the maximum number of communications modes that we could have that are as well connected as is the best mode is

$$N_{\text{max}} = \frac{V_R V_T}{(4\pi r)^2} \frac{(4\pi)^2}{\lambda^2(2\Delta z_T)(2\Delta z_R)} = n_{x\text{crit}} n_{y\text{crit}}, \quad (71)$$

which is exactly the number we would deduce directly from the properties of the eigenvalues of the prolate spheroidal functions. Hence our result is consistent with the sum rule.

To complete the problem, we need also to find the associated wave functions in the receiving volume V_R . The solution proceeds with a function that is exactly analogous to the source functions derived above, except that we start with Eq. (41). By analogy with Eq. (48), we can write

$$\phi(\mathbf{r}_R) = F_R(\mathbf{r}_R)\beta_R(\mathbf{r}_R), \quad (72)$$

where

$$F_R(\mathbf{r}_R) = \exp\left\{ik\left[z_R - \frac{1}{2r}(x_R^2 + y_R^2)\right]\right\} \quad (73)$$

is the receiving focusing function. By choosing the separation

$$\beta_R(\mathbf{r}_R) = \alpha_{Rx}(x_R)\alpha_{Ry}(y_R)\alpha_{Rz}(z_R) \quad (74)$$

as in Eq. (51), we similarly deduce that $\alpha_{Rz}(z_T)$ is a constant with the associated eigenvalue

$$\eta_{Rz} = (2\Delta z_R)(2\Delta z_T). \quad (75)$$

Exactly following steps analogous to those above, we have

$$\phi_{n_x n_y}(\mathbf{r}_R) = F_R(\mathbf{r}_R)\alpha_{n_x}(x_R)\alpha_{n_y}(y_R), \quad (76)$$

where $\alpha_{n_x}(x_R)$ is the n_x th (linear) prolate spheroidal function with a bandwidth of $\Omega_{Rx} = k\Delta x_T/r$, a scale Δx_R , and an eigenvalue ν_{n_x} ; the situation is similar for the y direction. All the results that we deduced above about the eigenvalues and the numbers of strongly coupled modes can be deduced identically if we consider this eigen problem for the received waves. Note that the x functions in the transmitting and the receiving volumes are identical in form to one another; the c_x parameter is the same number for both sets of prolate spheroidal functions. The scale factors of the functions will, however, in general be different if the sizes of the volumes are different in the x direction. Similar conclusions apply to the functions in the y direction. It does not matter which volume we consider first; having solved the problem for one volume, we already know the solutions for the other volume.

Hence we see that, by considering rectangular prism volumes that are far apart compared with their dimensions, we (i) reproduce the results of the previous analysis⁸ for surfaces and obtain prolate spheroidal wave functions as the forms of the source and the

wave functions in the two volumes, (ii) obtain results from the direct solution of the eigen problems for the number of strongly connected modes that agree with the sum rule [Eq. (23)], and (iii) find that the inclusion of the finite thickness of the two volumes has no effect on the number of degrees of freedom (or on the number of strongly connected modes) for communicating between the two volumes (under the assumption that the separation of the volumes is large compared with their dimensions), although the communications modes in thicker volumes are more strongly connected [see, e.g., expression (69)] in proportion to the thicknesses of both volumes.

As an illustrative numerical example, let us consider two volumes that are separated by 81λ , a transmitting volume V_T of dimensions $9\lambda \times 9\lambda \times 27\lambda$, and a receiving volume of dimensions $9\lambda \times 9\lambda \times 18\lambda$. We find that we will have to deal with numbers of degrees of freedom of 1, 2, 3, and 6 as we analyze this problem. The prolate spheroidal functions $S_{0n}(c, \xi)$ for 1, 2, and 3 degrees of freedom ($c = \pi/2, \pi, 3\pi/2$, respectively) are shown in Fig. 5. These functions were calculated by use of the routines described in Refs. 23 and 24. Note that, as might be expected, the number of zeros in the functions over the range -1 to 1 is equal to the index n . Note also that the set of functions (with $n = 0, 1, 2, \dots$) for a given value of c is complete and orthogonal over the interval -1 to 1 .

Table 1 lists the calculated eigenvalues ν_n that correspond to the various functions shown in Fig. 3. (Note that these are not the eigenvalues quoted by Flammer²² for prolate spheroidal wave functions; as discussed by Frieden,¹⁰ the relevant eigenvalues are different here and are actually derived from the radial prolate wave functions.) Note that the functions with large eigenvalues have small amplitudes at $+1$ and -1 and also have gradients that appear to become asymptotic parallel to the axis. This asymptotic behavior is consistent with the function's being relatively well confined within this range. Functions with small eigenvalues correspond to functions that are not well focused into the volume, and they will have substantial amplitudes outside the volume.

There are many different possible relative configurations for the example volumes. For example, if we line up the long directions of the two volumes in the y direction (i.e., $2\Delta x_T = 9\lambda, 2\Delta y_T = 27\lambda, 2\Delta z_T = 9\lambda, 2\Delta x_R = 9\lambda, 2\Delta y_R = 18\lambda$, and $2\Delta z_R = 9\lambda$), which we might consider to be the most conventional optical configuration, we find that $n_{x_{\text{crit}}} = 1$ and $n_{y_{\text{crit}}} = 6$. We would find approximately six strongly connected communications modes, each corresponding to different patterns in the y direction in each volume.

Another relatively conventional optical configuration is the crossed transverse configuration with the long (27λ) axis in the y direction in the transmitting volume and the receiving volume's long (18λ) axis in the x direction (i.e., $2\Delta x_T = 9\lambda, 2\Delta y_T = 27\lambda, 2\Delta z_T = 9\lambda, 2\Delta x_R = 18\lambda, 2\Delta y_R = 9\lambda$, and $2\Delta z_R = 9\lambda$). This configuration has values of $n_{x_{\text{crit}}} = 2$ and $n_{y_{\text{crit}}} = 3$. For each of three patterns in the y direction there are

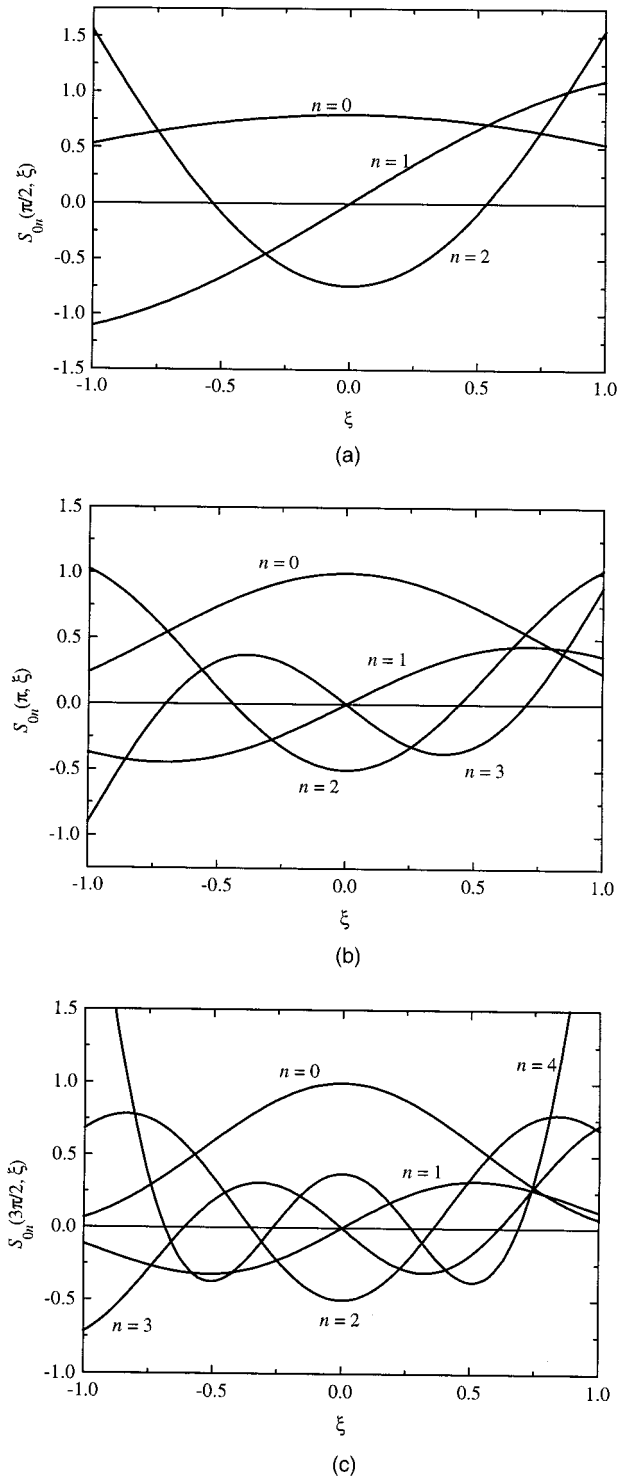


Fig. 5. Normalized prolate spheroidal functions for various parameter values: (a) 1 degree of freedom ($c = \pi/2$), (b) 2 degrees of freedom ($c = \pi$), (c) 3 degrees of freedom ($c = 3\pi/2$).

two distinct patterns in the x direction, again giving six strongly connected communications modes.

There are several other distinct configurations. Figure 6 illustrates a configuration in which the long (18λ) axis is in the z direction in the receiving volume (i.e., $2\Delta x_T = 9\lambda, 2\Delta y_T = 27\lambda, 2\Delta z_T = 9\lambda, 2\Delta x_R = 9\lambda,$

Table 1. Eigenvalues ν_n That Correspond to the Various Functions Shown in Fig. 3^a

Eigenvalue ν_n	Degrees of Freedom		
	1	2	3
ν_0	0.783	0.981	0.999
ν_1	0.205	0.750	0.969
ν_2	0.011	0.244	0.733
ν_3		0.025	0.263
ν_4			0.0005
Total	1.00	2.00	2.96

^aBlank entries indicate that the values are negligibly small.

$2\Delta y_R = 9\lambda$, and $2\Delta z_R = 18\lambda$) with values of $nx_{\text{crit}} = 1$ and $ny_{\text{crit}} = 3$. The most extreme, well-connected communications mode, $nx = 0$ and $ny = 2$, is illustrated in Fig. 6. Note in this case that there are only three well-connected communications modes because of the greater depth in the z direction in the receiving volume, although each of these three modes is more strongly connected than in the previously discussed cases that each had six strongly connected modes.

Note that, in the example above (Fig. 6), we chose the dimensions such that nx_{crit} and ny_{crit} were always greater than or equal to 1. This choice ensures that we have strongly connected modes for all the relative orientations discussed above. The situation in which one or more of the dimensions is too thin to satisfy this criterion yields somewhat different results, which we discuss below.

D. Communications Modes for Very Small Volumes

Consider now the case of volumes of arbitrary shape that are very small. Formally, for the volume to be very small, we require that the diffraction angle from one volume be much larger than the angle subtended by the other volume:

$$\frac{\Delta x_{TM} \Delta x_{RM}}{\lambda_0 r} \ll 1, \quad (77)$$

where Δx_{TM} (Δx_{RM}) is now the maximum size of the transmitting (receiving) volume in the x direction. There is an equivalent relation for the y direction.

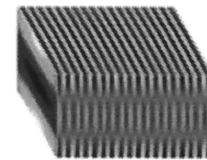
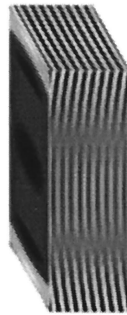


Fig. 6. Most extreme well-connected communications mode with $nx = 0$ and $ny = 2$ between volumes with dimensions of $9\lambda \times 27\lambda \times 9\lambda$ (left-hand image) and $9\lambda \times 9\lambda \times 18\lambda$ (right-hand image). The two volumes are separated by 81λ (center to center). The volumes are shown to scale in an isometric projection.

The diffraction angle should also be such that the maximum size of the volume in the z direction (the direction of the axis between the volumes) is still much less than the separation r between the volumes. Then we have

$$\exp(ik|\mathbf{r}_R - \mathbf{r}_T|)\exp(-ik|\mathbf{r}_R - \mathbf{r}'_T|) \cong \exp(-ik\{z_T - z'_T\}), \quad (78)$$

and the eigenequations of Eqs. (37) and (41) become very simple. Specifically, we have

$$|g|^2 \psi(\mathbf{r}'_T) = \frac{1}{(4\pi r)^2} \int_{V_T} \int_{V_R} \exp(ik[z_T - z'_T]) d^3 \mathbf{r}_R \psi(\mathbf{r}_T) d^3 \mathbf{r}_T, \quad (79)$$

$$|g|^2 \phi(\mathbf{r}'_R) = \frac{1}{(4\pi r)^2} \int_{V_R} \int_{V_T} \exp(ik[z_R - z'_R]) d^3 \mathbf{r}_T \phi(\mathbf{r}_R) d^3 \mathbf{r}_R, \quad (80)$$

respectively. By writing

$$\psi(\mathbf{r}_T) = \beta_T(\mathbf{r}_T) \exp(-ikz_T), \quad (81)$$

we obtain the trivial integral equation

$$|g|^2 \beta_T(\mathbf{r}_T) = \frac{1}{(4\pi r)^2} \int_{V_T} \int_{V_R} \beta_T(\mathbf{r}'_T) d^3 \mathbf{r}_R d^3 \mathbf{r}'_T. \quad (82)$$

Nothing in the integrand of Eq. (82) depends on \mathbf{r}_R , so the integral over \mathbf{r}_R simply becomes V_R . Similarly, nothing on the right-hand side of Eq. (82) depends on \mathbf{r}_T , so the only solution is that $\beta_T(\mathbf{r}_T)$ is a constant. Hence the only solution is

$$\psi(\mathbf{r}_T) = \frac{1}{\sqrt{V_T}} \exp(-ikz_T), \quad (83)$$

where we also normalized this source function with

$$|g|^2 = \frac{V_R V_T}{(4\pi r)^2}. \quad (84)$$

Note that the value of $|g|^2$ from Eq. (84) is sufficient to satisfy the sum rule completely, so there are no

other communications modes with finite connection strength in this limit.

Similarly, we deduce that

$$\phi(\mathbf{r}_R) = \frac{1}{\sqrt{V_R}} \exp(-ikz_R) \quad (85)$$

for the normalized wave function $\phi(\mathbf{r}_R)$. Hence we conclude that, for such small volumes, there is essentially only one communications mode with finite connection strength and that the optimum choice of source is one with the same form as a plane wave that propagates along the direction between the transmitting and the receiving volumes. This propagation direction gives rise to a wave that is essentially a plane wave within the receiving volume (although it cannot, in general, be a plane wave with such a finite source, as would be quite apparent if we were to look at the wave substantially outside the receiving volume).

E. Communications Modes for Arbitrary Volumes

We can use the general method presented in Subsection 5.B to deal numerically with extreme situations or cases that simply cannot be approached by the methods of previous authors. An example is shown in Fig. 7. We have two very thin ($1/10$ wavelength) volumes at right angles to each other and only one wavelength apart. A conventional picture based on plane-parallel surfaces can tell us nothing about the communications modes in this situation. Note that the separation between these volumes is less than the thickness (horizontal length) of the transmitting volume and that the receiving volume effectively has a large numerical aperture in the vertical direction, especially as seen from the nearer end of the transmitting volume.

In general, we can find an approximate numerical solution by taking a finite number of basis functions in each volume; in the limit of large numbers of basis functions this approach can be proved to converge toward the actual solution and is known to yield reliably an underestimate of the coupling strengths.²⁰ Solving numerically for the communications modes by use of Eqs. (37) and (41) yields the functions illustrated in Fig. 7 for the two strongest modes. For this integration, we formally used a finite Fourier basis in each volume initially, constructed the matrix of the coefficients g_{ji} , and then solved the resulting matrix eigenequations for the eigenfunctions and the eigenvalues. In each case, the functions do not vary significantly along the thin directions, so the functions in these directions were taken to be constant.

The first mode [Fig. 7(b)] takes at least approximately 86% of the available communications strength (i.e., $|g_1|^2 \cong 0.86\gamma_{RT}$), and the second mode [Fig. 7(c)] takes at least approximately 11% (i.e., $|g_2|^2 \cong 0.11\gamma_{RT}$). There are apparently no other modes of significant strength (the 3% of strength that is unaccounted for may be from limitations of the numerical technique). Incidentally, this problem is symmetric, yielding essentially the same solutions if the roles of

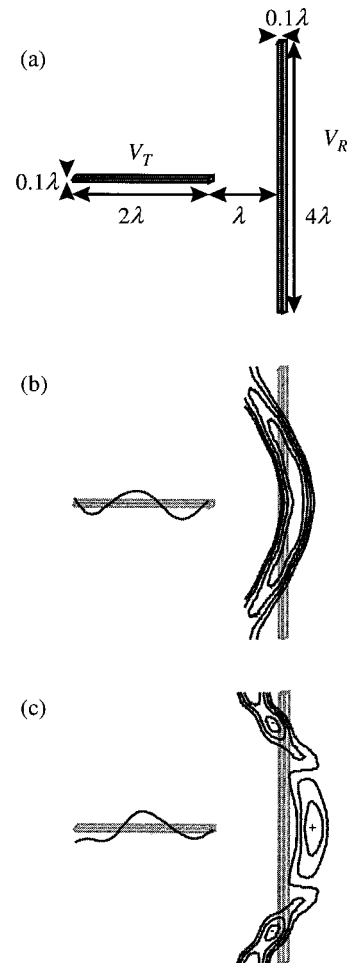


Fig. 7. Illustration of (a) two thin volumes, (b) the strongest communications mode, (c) the second-strongest communications mode. For the transmitting volume V_T [the thin horizontal volume shown in (a)] the real part of the wave amplitude along the length of the volume is shown for a particular arbitrary phase. For the receiving volume V_R [the thin vertical volume shown in (a)] the real part of the wave is shown in a contour plot that illustrates approximately half of a period of the wave with a horizontal scale such that 2π of the phase is the same size as one wavelength on the diagram. With this choice of scale the curvatures of the phase fronts correspond approximately to the actual curvature of the propagating waves. Dimensions are in wavelengths λ . Note that the second communications mode changes sign between the peak in the center and those in the upper and the lower lobes. Note also that these upper and lower lobes are more intense than the center peak. At least 86% of the available communications strength is in the first mode, and at least 11% of the strength is in the second mode.

the transmitting and the receiving volumes and of the source and the wave functions are interchanged. The orthogonality of the functions is relatively obvious for the waves (the second mode has strong side lobes of opposite sign to the main peak); the source functions are orthogonal also, although it is necessary to look at the entire complex function to see this clearly. We note that the second mode has greater intensity in the wave at the edges of the receiving volume and a somewhat stronger contribution from

the far end of the source, as we might expect intuitively.

F. Relation to Resonator Theory and Phase Conjugation

Suppose that we start out with the source $\psi_n(\mathbf{r}_T)$ in the transmitting volume V_T , which generates the wave $\phi(\mathbf{r}_R) = g_n \phi_n(\mathbf{r}_R)$ in the receiving volume V_R . Now suppose that we have some physical process in V_R that generates a new source of amplitude, $A\phi^*(\mathbf{r}_R) = Ag_n^* \phi_n^*(\mathbf{r}_R)$. This source is, by definition, the phase conjugate of the wave in V_R . Then we know from Eq. (43) that the wave generated in the volume V_T is $\phi'(\mathbf{r}_T) = A|g_n|^2 \psi_n^*(\mathbf{r}_T)$. If we then similarly have a mechanism in V_T that generates a new source with an amplitude $B\phi'(\mathbf{r}_T)$, the resulting source is $BA^*|g_n|^2 \psi_n(\mathbf{r}_T)$, which is simply a constant times the original source. Hence this system constitutes a resonator, and the communications modes are essentially the modes of this resonator that reproduce themselves on multiple reflections inside the resonator—in other words, we have proved that the communications modes can be generated physically if we set up a phase-conjugate resonator system in which each mirror is a phase-conjugate reflector.²⁵ The communications modes are, quite generally, the modes of the phase-conjugate resonator that is formed from the two volumes; this condition applies without approximations (the coefficients A and B must, of course, be constants), giving a physical interpretation of communications modes.

One very simple case of a resonator is the confocal resonator formed from two mirrors whose radius of curvature is equal to the separation between the mirror centers. If the phase front arriving at the mirror has a curvature that matches that of the mirror the resulting reflected phase front is actually the phase conjugate of the incident phase front and has a phase front that is the exact reversed version of the incident wave. Such a wave is focused initially on the center of the other mirror (although it subsequently diffracts). The reader can see connection to the situation depicted in Fig. 4. In fact, it is now obvious from the above discussion that, in the paraxial case [i.e., with conditions similar to those of expression (44)], just as the communications modes between two parallel, thin, rectangular volumes that are spaced far apart are the prolate spheroidal functions, so also are the modes of the confocal resonator the prolate spheroidal functions, a fact that is already well known from previous work.¹⁰ Hence this present discussion of the modes of phase-conjugate resonators and communications modes is consistent with this previous work.

G. Rationalization of the Size of the Connection Strengths

Above we showed formally how to solve for the communications modes and their (squared) connection strengths $|g|^2$. We also showed explicit results for the case of rectangular prism volumes that are far apart. In addition we saw in this rectangular prism case that mathematically the connection

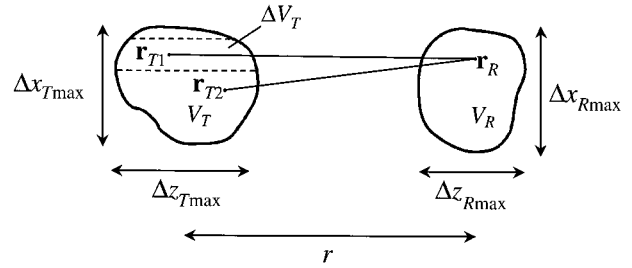


Fig. 8. Illustration in cross section of the source points \mathbf{r}_{T1} and \mathbf{r}_{T2} in volume V_T and the volume V_R , which contains the resultant waves of interest. Also shown is a volume ΔV_T near \mathbf{r}_{T1} . Other sources in ΔV_T , with their phases appropriately chosen, are expected to produce waves in V_R that are substantially similar (not orthogonal) to the wave from the source at point \mathbf{r}_{T1} , whereas sources outside ΔV_T are expected to produce waves in V_R that are substantially orthogonal to those from the source at point \mathbf{r}_{T1} .

strengths are all similar in size up to a critical number, above which they fall off dramatically. We showed too that the product of this size and the critical number would satisfy the sum rule [Eq. (23)]. This behavior of the connection strengths raises the following questions: (i) Is there a physical explanation of why the connection strengths exhibit this behavior, an explanation that might give us some more general physical insight? (ii) Is there a more general way to make intelligent guesses for what will constitute reasonable communications modes and to estimate their connection strengths without the need to calculate an entire exact solution? Here we attempt to answer these questions. The arguments are necessarily approximate and not completely rigorous; exactness is not the point, however, and we already have exact methods for any particular case anyway.

A key concept to understanding these questions is to understand the extent to which the waves in the receiving volume that are generated from different points in the transmitted volume are orthogonal to one another. We can imagine two sources at points \mathbf{r}_{T1} and \mathbf{r}_{T2} in the transmitting volume, as shown in Fig. 8. Obviously, if these points are close the waves that they generate in V_R will be almost identical; hence they will not be orthogonal. As we move the points further apart for the particular monochromatic case we consider here, there is, of course, a rapidly varying phase that would cause the waves to go through alternating constructive and destructive interference. Let us presume, however, that we always make the best possible choice of relative phase of the two sources to eliminate this particular interference effect. Even with this optimal phase choice, as we move points \mathbf{r}_{T1} and \mathbf{r}_{T2} further apart, we can expect that the waves will become progressively different from one another.

Formally, to assess the orthogonality of the waves, we should evaluate their overlap integral in V_R . Because the wave at point \mathbf{r}_R in V_R from a point source at \mathbf{r}_T is simply $G(\mathbf{r}_R, \mathbf{r}_T)$, the overlap integral of the

two waves (within a phase factor for the relative phase of the two sources) is

$$K(\mathbf{r}_{T2}, \mathbf{r}_{T1}) = \int_{V_R} G^*(\mathbf{r}_R, \mathbf{r}_{T2})G(\mathbf{r}_R, \mathbf{r}_{T1})d^3\mathbf{r}_R. \quad (86)$$

Note that Eq. (86) represents the kernel [Eq. (34)] of the integral eigenequation that gives the communications modes and is also involved in calculating the connection strength $|g|^2$ in Eq. (33).

We now use a simplistic model to obtain some approximate results. As we said above, we expect the overlap to be large and then to fall off as we separate the two source points, so we now presume that, for any point \mathbf{r}_{T1} , there is a finite volume ΔV_T for which $K(\mathbf{r}_{T2}, \mathbf{r}_{T1})$ is finite (neglecting the overall rapid destructive and constructive interference) and, for simplicity, is approximately constant (in magnitude), and again for simplicity, for \mathbf{r}_{T2} outside this volume $K(\mathbf{r}_{T2}, \mathbf{r}_{T1})$ is approximately zero. This approach will give us an estimate of $|g|^2$. To proceed, we need to estimate the dimensions of this volume ΔV_T .

We consider only volumes that are far apart compared with their linear dimensions, so we can take paraxial approximations for simplicity. Then we have

$$\begin{aligned} G^*(\mathbf{r}_R, \mathbf{r}_{T2})G(\mathbf{r}_R, \mathbf{r}_{T1}) &= \frac{\exp[-ik(|\mathbf{r}_R - \mathbf{r}_{T1}| - |\mathbf{r}_R - \mathbf{r}_{T2}|)]}{(4\pi r_0)^2} \\ &\cong \frac{1}{(4\pi r_0)^2} F_T(\mathbf{r}_{T2})F_T^*(\mathbf{r}_{T1})\exp\left[-\frac{ikx_R(x_{T2} - x_{T1})}{r_0}\right]\exp\left[-\frac{iky_R(y_{T2} - y_{T1})}{r_0}\right], \end{aligned} \quad (87)$$

where $F_T(\mathbf{r}_T)$ is the focusing phase-factor function [defined above by (Eq. (49)], which takes care of much of the rapidly varying phase. The remaining exponentials in expression (87) are unity for small arguments; however, after $(x_{T2} - x_{T1})$ or $(y_{T2} - y_{T1})$ becomes sufficiently large, the function overall becomes oscillatory as we move through V_R , and the integral [expression (86)] will tend to average to zero. Essentially, we then cannot avoid the effects of interference as we integrate over V_R no matter how well we choose the relative phases of the sources. A characteristic size of $(x_{T2} - x_{T1})$ for which this interference becomes strong occurs when

$$\left|\frac{k\Delta x_{R\max}(x_{T2} - x_{T1})}{r}\right| \approx \frac{\pi}{2}, \quad (88)$$

that is, when

$$|x_{T2} - x_{T1}| \approx \frac{\pi r}{2k\Delta x_{R\max}} = \frac{1}{2} \frac{\lambda r}{\Delta x_{R\max}}. \quad (89)$$

In this case, as we integrate from the middle of V_R to one extreme in the x direction, the phase of the integrand will change by $\pi/2$. For the purposes of our argument, we can use a simplistic approximation and replace these remaining exponentials by functions

that are unity up to arguments of $\pm\pi/2$ [i.e., no (destructive) interference] and zero otherwise (i.e., total destructive interference on the average). Note, incidentally, that, given that we make the best choice of phase for our sources, the sources may have any position in the z direction. Hence we conclude that our volume ΔV_T has dimensions of approximately $\lambda r/2\Delta x_{R\max}$ and $\lambda r/2\Delta y_{R\max}$ in the x and the y directions, respectively, and of $2\Delta z_{T\max}$ in the z direction for a total volume of

$$V_T \approx \frac{\lambda r}{2\Delta x_{R\max}} \frac{\lambda r}{2\Delta y_{R\max}} 2\Delta z_{T\max}, \quad (90)$$

and by use of Eq. (86) and expression (87), we find

$$K(\mathbf{r}_{T2}, \mathbf{r}_{T1}) \approx F_T^*(\mathbf{r}_{T2})F_T(\mathbf{r}_{T1}) \frac{V_R}{(4\pi r)^2} \quad (91)$$

for \mathbf{r}_{T2} within the volume ΔV_T near \mathbf{r}_{T1} and zero otherwise.

Now we make an intelligent guess at the eigenfunctions for this problem. We choose a function that is essentially uniform within a volume of size ΔV_T near some point \mathbf{r}_{T0} [except for having a phase factor $F_T(\mathbf{r}_T)$ to ensure that we have the correct choice of

phase for the source] and is zero elsewhere, i.e., within ΔV_T , we have

$$\psi_1(\mathbf{r}_T) = \frac{1}{(\Delta V_T)^{1/2}} F_T(\mathbf{r}_T), \quad (92)$$

where we have also normalized the function $\psi_1(\mathbf{r}_T)$. Now we can calculate $|g|^2$ from Eq. (33):

$$\begin{aligned} |g|^2 &= \int_{V_T} \psi_1^*(\mathbf{r}_{T2}) \int_{V_T} K(\mathbf{r}_{T2}, \mathbf{r}_{T1})\psi_1(\mathbf{r}_{T1})d^3\mathbf{r}_{T1}d^3\mathbf{r}_{T2} \\ &= \int_{\Delta V_T} \frac{1}{(\Delta V_T)^{1/2}} \frac{V_R\Delta V_T}{(4\pi r)^2} \frac{1}{(\Delta V_T)^{1/2}} d^3\mathbf{r}_{T2} \\ &= \frac{V_R\Delta V_T}{(4\pi r)^2}. \end{aligned} \quad (93)$$

To form another eigenfunction orthogonal to that of Eq. (93), we can simply move sideways to an adjacent volume of essentially the same shape and construct the second eigenfunction similarly in that volume. The second eigenfunction is orthogonal to the first because there is exactly no overlap between such functions. It will yield exactly the same result as

above for $|g|^2$. We can continue this process until we have used up all the volume, which will give us a number N of orthogonal functions

$$N \approx \frac{V_T}{\Delta V_T}, \quad (94)$$

all with the same value of $|g|^2$. Note that

$$N|g|^2 = \frac{V_R V_T}{(4\pi r)^2}, \quad (95)$$

which satisfies the sum rule exactly, suggesting that these are all the modes with strong connections. Actually, to construct a function orthogonal to all these functions would require that the function change sign in the middle of one of the volumes ΔV_T , which in turn would mean that there would be no net wave in V_R from this source function, hence no connection strength $|g|^2$, consistent with our arguments.

The above argument works particularly well if the volume is of uniform thickness in the z direction because then the various ΔV_T volumes will all be the same length. Otherwise, they will be of different lengths with proportionately different coupling strengths. We have, however, now successfully rationalized why it is that rectangular prism volumes have a set of N communications modes all of approximately the same strength, and we can conjecture that other volumes of uniform thickness will also have a set of N communications modes of substantially equal coupling strength. We can conversely conjecture that, for volumes that are not uniform in thickness, the coupling strengths of the various communications modes will not be substantially equal. In other words, we can hypothesize that the uniformity of the thickness of the transmitting and the receiving volumes leads to substantially equal connection strengths for the strongly coupled communications modes.

One might object that we chose simplistic approximate eigenfunctions that are not in reality very similar to the exact solutions of the rectangular prism volume problem; these exact solutions are, after all, prolate spheroidal functions that each fill the entire volume rather than the solutions given here that are localized each within a subvolume ΔV_T . Note, however, that all these simplistic approximate eigenfunction solutions are degenerate, corresponding to the same eigenvalue; hence we can take orthogonal linear combinations of them instead, and it is straightforward to construct sets that each fill the volume. The number of orthogonal functions remains unchanged by such linear combinations. Of course, the functions should not, in fact, have quite the abrupt cutoff at the edges that our simple model has, but smoothing these out somewhat does not change the basic argument. Our simplistic approximate eigenfunctions, which are constant (except for a phase factor) within a given ΔV_T and zero outside it, can be viewed as a generalization for our volume case of the spots used by Gabor for the planar case. In fact, if

we perform our approximate argument here for the case of thin rectangular prism volumes, we essentially recover Gabor's diffraction-spot picture. Our simplistic eigenfunctions are no more the true exact eigenfunctions than are Gabor's spots, but they do correctly count the available degrees of freedom and yield a substantially correct intuitive picture.

Note that our argument here does not apply to the case in which the volumes are close compared with their thickness. In that case other modes may appear that utilize the depth of the volume, as was found above in the numerical example in Subsection 5.E.

6. Conclusions

We have shown that there is a rigorous and complete method for finding the orthogonal communications channels or modes for scalar waves between two arbitrary volumes in free space. In contrast to previous approaches that were restricted to planar surfaces that are far apart, this approach is exact and is valid for volumes of arbitrary shape. The resulting communications modes are found as the results of two well-defined eigenvalue problems. (We have, incidentally, shown that these communications modes are the same as the cavity modes of the resonator that results if the two volumes are filled with a perfect phase-conjugating material.) The present approach also provides the strengths of the connections formed by the channels, and we have derived a sum rule for those strengths, with the sum depending on only a simple integral over the two volumes. The connection strengths can be interpreted in terms of power-transfer coefficients. The sum rule shows that, although there may be an infinite number of communications modes, only a finite number of these modes can be strongly connected.

We have demonstrated explicitly that the approach presented here will reproduce previous research results for rectangular surfaces that are far apart, showing both that the resulting eigenfunctions (prolate spheroidal functions) correspond to those deduced previously and that the sum rule does indeed predict the number of well-connected modes in the system. We have been able to extend the results to show the effect of the depth of the volume for both rectangular prism volumes and, by illustration, more general volumes. Briefly, when the volumes are far apart compared with their depth, the depth does not generate any increase in the number of communications modes, although it can increase the connection strengths. When volumes are close together compared with their depths, the numerical aperture or the solid angle subtended by one volume at the other is large, or both, new modes are possible.

Incidentally, it is important to make the distinction here between communications channels and stored information. This question arises, for example, in considering the information capacity of a volume hologram. All we have considered here is the number of channels for communicating information, not the amount of information that can be communicated.

In a simple system in which, for example, we make only binary decisions about whether there is a signal in a channel, our number of strongly connected communications modes essentially tells us how many bits of information can be read out in the receiving volume from any given source in the transmitting volume—in optical storage terms it tells us the maximum number of bits per page (for such binary decisions). We can, of course, choose some other source function in the transmitting volume, in which case a different page of information will be transmitted. (In reading out a volume hologram, we usually generate the different source function by exciting the medium with a different readout beam.) Deciding the maximum number of bits that can be stored in the hologram is therefore quite a different problem. One should not therefore be surprised at our conclusion that the depth of the volume under usual optical conditions makes no difference to the number of communications channels. It may well be that increasing the depth of the volume does increase the number of different source functions that can be excited in the volume by a readout beam (i.e., the number of pages), but that is quite a different calculation.

We have also shown how to calculate communications modes in arbitrary or extreme situations. For example, we have calculated communications modes for volumes that (i) are smaller than a wavelength in at least one lateral dimension, (ii) are closer together than the depth of one volume, (iii) have one volume presented edge-on to the other, (iv) have one volume that subtends a large solid angle to the other, and (v) have volumes that are only a wavelength apart. Any one of these five conditions would have been enough to invalidate the results of previous methods.

Although any such calculation of modes for arbitrary volumes will be approximate because only a finite basis set can generally be used for numerical calculations, the sum rule gives us a measure of how good any such finite basis approximation is and how close we are to finding all the communications modes of substantial interconnection strength. It can be proved that any such finite basis will underestimate the connection strengths; so, if we are close to satisfying the sum rule with our finite basis, we can be sure that we have identified all the significant communications modes and have included enough functions in our basis set.

We have also analyzed the case of communications modes between small volumes that are far apart. We found one mode (essentially the one that has a plane-wave source and generates an approximately plane wave in the receiving volume); we proved by use of the sum rule that this is the only mode with significant connection strength.

In general, the approach presented here establishes a rigorous basis for handling problems that are related to volume sources and receivers of waves. We have not attempted to address any specific practical problem, but this approach could be useful in understanding limits and opportunities in areas in which (i) conventional diffraction approaches break

down, such as in near-field microscopy or very-fine-line lithography, or (ii) volume is an intrinsic part of the problem, such as in volume holography or certain types of optical interconnects or information storage and possibly modes in photonic bandgap crystals or other volume scatterers. It should be emphasized, however, that the present study discusses only the modes or the channels for communicating information into and out of volumes and not the information that is stored inside the volume or that can be read out of the volume. Other extensions of this work could include the addition of the time dimension to the problem (which can be done relatively straightforwardly from a mathematical point of view), and solution of the vector (rather than only the scalar) wave problem is already being considered to allow extensions of the present approach to be applied rigorously to electromagnetic waves.²

I am pleased to acknowledge Haldun Ozaktas and Rafael Piestun for critical readings of versions of this manuscript.

References and Notes

1. I previously have briefly presented the derivation of the communications modes and the sum rule in D. A. B. Miller, "Spatial channels for communicating with waves between volumes," *Opt. Lett.* **23**, 1645–1647 (1998), and in summary form in D. A. B. Miller, "Communicating with waves between volumes—how many different spatial channels are there?" in *Optics in Computing '98*, P. Chavel, D. A. B. Miller, and H. Thienpont, eds., *Proc. SPIE* **3490**, 111–114 (1998).
2. R. Piestun and D. A. B. Miller, "Degrees of freedom of an electromagnetic wave," in *Eighteenth Congress of the International Commission for Optics*, A. J. Glass, J. W. Goodman, M. Chang, A. H. Guenther, and T. Asakura, eds., *Proc. SPIE* **3749**, 110–111 (1999); R. Piestun and D. A. B. Miller, "Electromagnetic degrees of freedom of an optical system," *J. Opt. Soc. Am. A* (to be published).
3. D. Gabor, "Light and information," in *Progress in Optics*, Vol. 1, E. Wolf, ed. (North-Holland, Amsterdam, The Netherlands, 1961), pp. 109–153.
4. G. Toraldo di Francia, "Resolving power and information," *J. Opt. Soc. Am.* **45**, 497–501 (1955).
5. G. Toraldo di Francia, "Directivity, supergain, and information," *Trans. IRE* **4**, 473–478 (1956).
6. A. Walther, "Gabor's theorem and energy transfer through lenses," *J. Opt. Soc. Am.* **57**, 639–644 (1967).
7. I omit here for simplicity the two distinct phases and, in the case of light, the two distinct polarizations, the effects of each of which can be viewed as doubling the number of degrees of freedom.
8. G. Toraldo di Francia, "Degrees of freedom of an image," *J. Opt. Soc. Am.* **59**, 799–804 (1969).
9. C. W. Barnes, "Object restoration in a diffraction-limited imaging system," *J. Opt. Soc. Am.* **56**, 575–578 (1966).
10. B. R. Frieden, "Evaluation, design, and extrapolation methods for optical signals, based on the use of the prolate functions," in *Progress in Optics IX*, E. Wolf, ed. (North-Holland, Amsterdam, The Netherlands, 1971), pp. 311–407.
11. F. Gori and G. Guattari, "Shannon number and degrees of freedom of an image," *Opt. Commun.* **7**, 163–165 (1973).
12. F. Gori and G. Guattari, "Degrees of freedom of images from point-like-element pupils," *J. Opt. Soc. Am.* **64**, 453–458 (1974).
13. L. Ronchi and A. Consortini, "Degrees of freedom of images in

- coherent and incoherent illumination," *Alta Frequenza* **43**, 1034–1036 (1974).
14. M. Bendinelli, A. Consortini, L. Ronchi, and B. R. Frieden, "Degrees of freedom and eigenfunctions for the noisy image," *J. Opt. Soc. Am.* **64**, 1498–1502 (1974).
 15. D. Slepian and H. O. Pollak, "Prolate spheroidal wave function, Fourier analysis and uncertainty—I," *Bell Syst. Tech. J.* **40**, 43–63 (1961).
 16. J. H. Shapiro, "Optimal power transfer through atmospheric turbulence using state knowledge," *IEEE Trans. Commun. Technol.* **COM-19**, 410–414 (1971).
 17. J. H. Shapiro, "Normal-mode approach to wave propagation in the turbulent atmosphere," *Appl. Opt.* **13**, 2614–2619 (1974).
 18. J. H. Shapiro, "Imaging and optical communication through atmospheric turbulence," in *Laser Beam Propagation in the Atmosphere*, J. W. Strohbehn, ed. (Springer-Verlag, Berlin, 1978), pp. 171–222.
 19. D. A. B. Miller, "Huygens's wave propagation principle corrected," *Opt. Lett.* **16**, 1370–1372 (1991).
 20. See, for example, D. Porter and D. S. G. Stirling, *Integral Equations* (Cambridge U. Press, Cambridge, 1990), for a discussion of these issues.
 21. R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1989).
 22. C. Flammer, *Spheroidal Wave Functions* (Stanford U. Press, Stanford, Calif., 1957).
 23. T. A. Beu and R. I. Câmpeanu, "Prolate radial spheroidal wave functions," *Comp. Phys. Commun.* **30**, 177–185 (1983).
 24. T. A. Beu and R. I. Câmpeanu, "Prolate angular spheroidal wave functions," *Comp. Phys. Commun.* **30**, 187–192 (1983).
 25. The phase-conjugate cavity discussed here has two phase-conjugating mirrors. This configuration differs from the situation with one phase-conjugating mirror and one conventional mirror that is more extensively discussed in the literature. The case with two phase-conjugate mirrors is briefly discussed in, for example, J. F. Lam and W. P. Brown, "Optical resonators with phase-conjugate mirrors," *Opt. Lett.* **2**, 61–63 (1980).