Preprint of D. A. B. Miller, "Communicating with Waves Between Volumes - How Many Different Spatial Channels Are There?," in *Optics in Computing '98*, Pierre Chavel, David A. B. Miller, Hugo Thienpont, Editors, Proceedings of SPIE Vol. 3490, 111 - 114 (1998)

Communicating with Waves Between Volumes - How Many Different Spatial Channels Are There?

David A. B. Miller
Ginzton Laboratory, Stanford University, Stanford CA 94305-4085, USA

Abstract

We show how to calculate exactly the number and strength of the connections between two arbitrary volumes, and derive and apply a novel sum rule.

It is obviously important to know how many "channels" are available for sending information from one volume to another using waves, and how strongly connected these channels are. This will be one of the factors that will limit our ability to communicate information, providing one of the bounds, for example, on optical interconnection, optical memory access and capacity, and our ability to exploit techniques such as very fine line lithography or near-field microscopy. Despite its basic importance in many areas, relatively little has apparently been known about this limit. Here we show that there is a rigorous and exact approach to this problem that (i) allows us to define uniquely the set of available spatial channels for communicating between arbitrary volumes (the "communications modes"), (ii) gives us a very general "sum rule" for the connection strength and number of such channels, (iii) enables us to deduce the previous approximate answers based on communicating between parallel planar surfaces (e.g., as in diffraction limits to the number of resolvable spots on a surface with a lens), and (iv) gives us new results based on the strength and number of "communications modes" between volumes.

The best-known previous model for the number of possible channels between two parallel surfaces is that due to Gabor [1]. He presumes, essentially, that a beam from one surface can be focused to a spot on another, second surface, where the size of the spot is that corresponding to the diffraction angle of a cone converging from the entire first surface area, and deduces the number of available distinct channels as the number of such Gaussian spots that can be placed on this second surface. This approach is technically informal (the Gaussian spots are not truly orthogonal), but gives results that are useful, correct for the range of circumstances to which they apply, and agree with an intuitive diffraction picture. This approach, however, tells us nothing about situations where the cross-sectional areas are less than a wavelength (as is often the case for a microphone and a loudspeaker, for example). It also raises a paradox. Finite areas result in finite numbers of degrees of freedom in the Gabor approach, but it is also well known that, if we know the amplitude and derivative of a wave over any finite surface, we can deduce the wave field everywhere, and to describe a field everywhere would require an infinite number of degrees of freedom. This paradox was resolved (see, e.g., Toraldo di Francia [2]) by the realization that the actual description of such communications modes, at least between plane-parallel circular or rectangular surfaces, was a kind of eigenmode problem. For the rectangular or circular "apertures", there is an exact set of functions (the prolate spheroidal functions[3]), each corresponding to a specific wave or source pattern on the surfaces in question, that define a set of orthogonal channels. The number of such channels that are "strongly connected" is the number deduced less formally by Gabor [1] previously, with the remaining required infinite number of channels or "degrees of freedom" being so weakly connected as to be negligible in practice while still resolving the formal paradox. This kind of eigenmode approach was also used to analyze problems such as communication through a turbulent atmosphere [4].

Neither of these previous approaches [1][2] tell us anything about consequences of the thickness of the volume. Both are also based on diffraction theory that, though useful, is an approximation to start with, and hence is of dubious value for drawing fundamental conclusions. Even the eigenmode approaches [2] were apparently only able to give conclusions about the number of channels for the specific cases of circular or rectangular apertures, in which cases they relied on specific properties of the eigenvalues associated with prolate spheroidal wavefunctions.

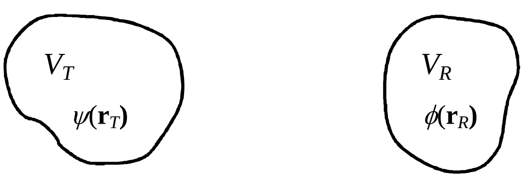

Fig. 1. Illustration of transmitting (V_T) and receiving (V_R) volumes, with associated source ($\Psi(\mathbf{r}_T)$) and wave ($\phi(\mathbf{r}_R)$), functions

Diffraction theory is based on effective, approximate sources in an aperture, whose amplitudes are given by the local wave amplitude or derivative. Here, by contrast, we work directly with the (scalar) (Helmholtz) wave equation and its exact solutions. We will restrict the discussion here to monochromatic waves (of angular frequency $\omega = ck$, where k is the magnitude of the wavevector and c is the wave propagation velocity), though that is not a necessary restriction. We start with a source function, $\Psi(\mathbf{r}_T)$, in a "transmitting" volume, V_T, which gives rise to a wave, $\phi(\mathbf{r}_R)$, in a "receiving" volume, V_R, as illustrated in Fig. 1.

We know directly from the wave equation that the Green's function corresponding to outgoing waves (i.e., the wave resulting from a point source at position $\mathbf{r}_T$) is

$$G(\mathbf{r}, \mathbf{r}_T) = \frac{e^{-ik|\mathbf{r}-\mathbf{r}_T|}}{4\pi|\mathbf{r}-\mathbf{r}_T|} \tag{1}$$

and so we can write the wave $\phi(\mathbf{r}_R)$, as

$$\phi(\mathbf{r}_R) = \int_{V_T} G(\mathbf{r}_R, \mathbf{r}_T)\,\psi(\mathbf{r}_T)\,d^3\mathbf{r}_T \tag{2}$$

We can choose to expand the source and wave in complete orthonormal sets for each volume, $a_{T1}(\mathbf{r}_T)$, $a_{T2}(\mathbf{r}_T)$, $a_{T3}(\mathbf{r}_T)$, ... for V_T, and $a_{R1}(\mathbf{r}_R)$, $a_{R2}(\mathbf{r}_R)$, $a_{R3}(\mathbf{r}_R)$, ... for V_R. For the moment there is no restriction on what these functions are other than that they are complete sets in their respective volumes. Suppose for the moment that the source function is one of these functions, $a_{Ti}(\mathbf{r}_T)$. In the wave generated in the receiving volume, we wish to know what is the resulting amplitude, g_{ji}, of one of the particular receiving modes, $a_{Rj}(\mathbf{r}_R)$. That amplitude will represent a "coupling coefficient" between these particular source and receiving modes. By multiplying both sides of Eq. (2) by $a_R^*(\mathbf{r}_R)$ and integrating, we obtain

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

Now, expanding the Green's function, $G(\mathbf{r}_R,\mathbf{r}_T)$, in the same basis sets gives

$$G(\mathbf{r}_R, \mathbf{r}_T) = \sum_{i,j} g_{ji} a_{Rj}(\mathbf{r}_R)\,a_{Ti}^*(\mathbf{r}_T) \tag{4}$$

We also know trivially from Eq. (1) that

$$|G(\mathbf{r}_R, \mathbf{r}_T)|^2 = \frac{1}{(4\pi)^2|\mathbf{r}_R-\mathbf{r}_T|^2} \tag{5}$$

so, from Eqs. (4) and (5), we find

$$\|\Gamma\| \equiv \sum_{i,j}|g_{ji}|^2 = \frac{1}{(4\pi)^2}\int_{V_R}\int_{V_T}\frac{1}{|\mathbf{r}_R-\mathbf{r}_T|^2}\,d^3\mathbf{r}_T\,d^3\mathbf{r}_R \tag{6}$$

Eq. (6) is quite a remarkable sum rule. It tells us that, for waves obeying the scalar Helmholtz wave equation, the sum of the squares of the "connection strengths" from sources in one volume generating waves in another volume is given by a simple volume integral over the two volumes. There are *no approximations* in this result, and it applies to *any* volumes, and for *any* orthonormal basis sets in the two volumes.

Before proceeding to consequences of this sum rule, we next ask what are the "best" choices of the basis functions. It is immediately clear from Eq. (6) that there is some maximum value of $|g_{ji}|^2$. Hence, there must be some pair of (normalized) "transmitting", $\Psi_1(\mathbf{r}_T)$, and "receiving", $\phi_1(\mathbf{r}_R)$, functions that are the most strongly coupled, i.e., for which the coupling coefficient g_1 has the largest squared modulus, $|g_1|^2$, and we choose $\Psi_1(\mathbf{r}_T)$ and $\phi_1(\mathbf{r}_R)$ to be the first members of the new basis function sets for the two volumes. We could find the functions $\Psi_1(\mathbf{r}_T)$ and $\phi_1(\mathbf{r}_R)$ by some variational procedure, for example. Then we could proceed to find the next members of the sets by finding those functions $\Psi_2(\mathbf{r}_T)$ and $\phi_2(\mathbf{r}_R)$ that are the most strongly coupled, under the constraint that they are orthogonal to $\Psi_1(\mathbf{r}_T)$ and $\phi_1(\mathbf{r}_R)$ respectively, and so on to establish the rest of the member of the basis sets (with each new member orthogonal to all the previous ones). By this means, we could establish all of the orthogonal "communications modes" between the two volumes, and their associated coupling strengths, g_i.

In fact, this kind of problem is known mathematically, and some aspects of it have already been applied (in the two dimensional case, and based on diffraction theory approximations) in the theory of turbulence [4]. If we were to rewrite this linear algebra problem in matrix form, the solutions we seek would reduce to the results of a singular value decomposition of

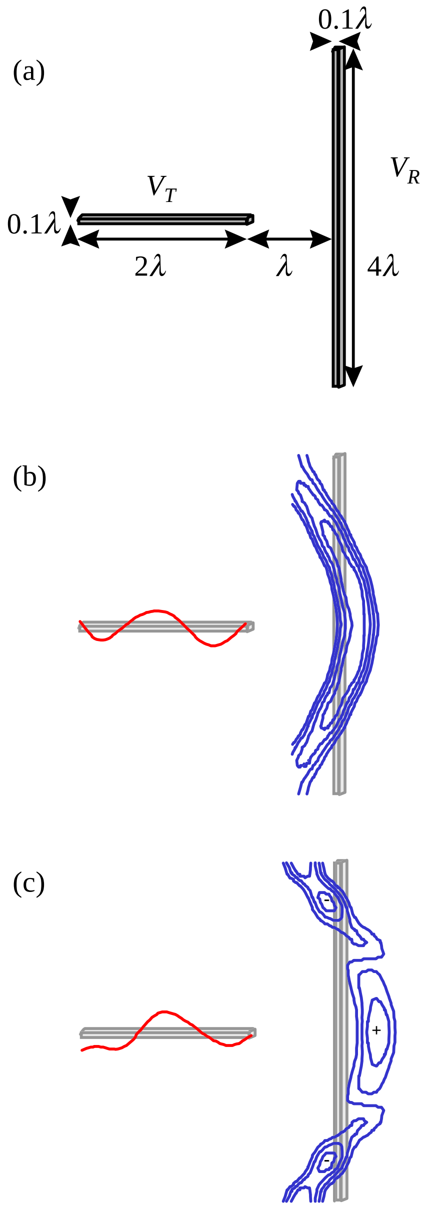

Fig. 2. Illustration (a) of two thin volumes considered in this example, (b) the strongest communications mode, and (c) the second communications mode. For the transmitting volume, the real part of the wave amplitude along the length of the volume is shown for a particular arbitrary phase. For the receiving volume, the real part of the wave is shown in a contour plot illustrating approximately half a period of the wave, and with horizontal scale such that 2π of phase is the same size as one wavelength on the diagram. With this choice of scale, the curvature of the phase fronts corresponds 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 the upper and 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% in the second.

the matrix of coefficients g_{ji}. In integral form, the results are that the set of pairs of functions, $\Psi_i(\mathbf{r}_T)$ and $\phi_i(\mathbf{r}_R)$, that correspond to our desired "communications modes" are the solutions of the integral eigen equations

$$|g_i|^2 \psi(\mathbf{r}'_T) = \int_{V_T} K(\mathbf{r}'_T, \mathbf{r}_T)\psi(\mathbf{r}_T) d^3\mathbf{r}_T \tag{7}$$

$$|g_i|^2 \phi(\mathbf{r}'_R) = \int_{V_T} J(\mathbf{r}'_R, \mathbf{r}_R)\phi(\mathbf{r}_R) d^3\mathbf{r}_R \tag{8}$$

where

$$K(\mathbf{r}'_T, \mathbf{r}_T) = \int_{V_R} G^*(\mathbf{r}_R, \mathbf{r}'_T) G(\mathbf{r}_R, \mathbf{r}_T) d^3\mathbf{r}_R \tag{9}$$

and

$$J(\mathbf{r}'_R, \mathbf{r}_R) = \int_{V_T} G^*(\mathbf{r}_T, \mathbf{r}'_R) G(\mathbf{r}_R, \mathbf{r}_T) d^3\mathbf{r}_T \tag{10}$$

Note that the eigenvalues $|g_i|^2$ are the same for both equations (7) and (8).

Note also that the sets $\Psi_i(\mathbf{r}_T)$ and $\phi_i(\mathbf{r}_R)$ are complete orthogonal sets within their respective volumes, and that these "communications modes" are orthogonal to one another, in that source function $\Psi_i(\mathbf{r}_T)$ is only coupled to the wave function $\phi_i(\mathbf{r}_R)$, and not to any other wave functions, i.e., mathematically, g_{ji} is zero unless $i = j$. (This also means that, with these basis function sets, the sum in Eq. (6) reduces to a sum over i only.) Hence, we have identified the distinct communications channels between two volumes, with the best possible couplings from source in V_T to wave in V_R.

It is mathematically straightforward now to deal with the situation of communication modes between two plane parallel rectangular surfaces by formally considering very thin volumes and taking a paraxial approximation, and we omit the details here. The result of applying the analysis above to this case is to recreate the previous results of Toraldo di Francia [2], who concluded that the optimum functions for both $\Psi_i(\mathbf{r}_T)$ and $\phi_i(\mathbf{r}_R)$ were prolate spheroidal functions. Applying the sum rule, Eq. (6), here tells us not only that the connection strengths drop off abruptly once we pass a specific number of modes but that their squares sum to a finite number.

A more substantial use of this analysis is to examine what happens in situations other than thin volumes. One simple question, relevant, for example, to reading out volume optical memories, is whether the use of a thick volumes gives us more spatial communication channels in or out of the volume; the answer, in most cases of optical interest, is that it does not. The number of channels we would deduce from the Gabor approach from the sizes of the end face and of the lens addressing it remains correct even as we make the volume thicker. To start to get more usable spatial communications channels in and out of the volume, we conclude that we would have to have a volume whose thickness was comparable to the separation between the volume and the lens (or other optical volume) addressing it, or go to a system with very high numerical aperture (>> 1), or both. We omit the details of this analysis here.

We can use this method to deal numerically with extreme situations, or cases that simply cannot be approached by the previous methods. An example is shown in Fig. 2. We have two very thin (1/10th wavelength) volumes at right angles to each other and only 1 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 very large "numerical aperture" in the vertical direction, especially as seen from the nearer end of the transmitting volume. Solving numerically for the communications modes using Eqs. (7) and (8) gives the functions illustrated in Fig. 2 for the strongest two modes. (In each case, the functions do not vary significantly along the "thin" directions.) The first mode (Fig. 2(a)) takes at least ~86% of the available communications strength (i.e., $|g_1|^2 \cong 0.86\,\|\Gamma\|$) and the second mode (Fig. 2(b)) takes at least ~11% (i.e., $|g_2|^2 \cong 0.11\,\|\Gamma\|$). There are apparently no other modes of significant strength (the 3% unaccounted for may be from limitations of the numerical technique). Incidentally, this problem is symmetric, giving essentially the same solutions if the roles of transmitting and receiving volume and of the source and 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, though it is necessary to look at the entire complex function to see this clearly. We note that the second mode has more intensity in the wave at the edges, and a somewhat stronger contribution from the "far end" of the source, as we might expect intuitively.

In a specific application more relevant to use in information technology, this approach can also be used, for example, to derive an absolute upper bound on the sum of the squares of the diffraction efficiencies of a volume hologram. We find this bound is proportional to the square of the maximum refractive index change in the medium (for small index changes) and a simple volume integral. Again, we omit the details of this analysis here.

In summary, we have (a) proposed a rigorous way of defining the orthogonal spatial channels for communication between two volumes, (b) shown a very basic and general sum rule for the squares of the interconnection strengths, (c) used this approach to reproduce previous approximate results for the number of "degrees of freedom" in communicating between two surfaces, (d) drawn clear conclusions about the effects of finite thickness on the number of communications channels, (e) illustrated by example the extreme cases in which new communications modes appear for very closely spaced volumes and (f) sketched results for general limits on diffraction efficiencies of volume holograms. This novel approach may have a variety of applications in information processing and elsewhere.

1. D. Gabor, "Light and Information", Progress in Optics, Vol. 1, ed. E. Wolf (North-Holland, Amsterdam, 1961), 109 - 153
2. G. Toraldo di Francia, "Degrees of Freedom of an Image", J. Opt. Soc. Am. **59**, 799 - 804 (1969)
3. B. R. Frieden, "Evaluation, Design and Extrapolation Methods for Optical Signals, Based on the Use of the Prolate Functions", in "Progress in Optics, IX", ed. E. Wolf, (North-Holland, Amsterdam, 1971), pp. 311-407
4. J. H. Shapiro, "Imaging and Optical Communication Through Atmospheric Turbulence", in "Laser Beam Propagation in the Atmosphere", ed. J. W. Strohbehn (Springer-Verlag, Berlin, 1978)