Advances in Information Optics and Photonics, eds. A. T. Friberg and R. Dändliker, International Commission for Optics Vol. VI (SPIE, Bellingham, 2008) , Chapter 8, pp. 153-170

# Limits to Optical Components

**David A. B. Miller**
Ginzton Lab, Stanford University, Stanford CA 94305-4088, USA



## 1.1 Introduction

Is there a limit to the performance of linear optical components? Suppose we asked a specific question, such as how much glass would we need to make a device that would split 32 wavelengths in the telecommunications C-band near 1.5 microns wavelength. Intuitively, we would probably agree that 1 cubic micron of glass would not be enough. We could, however, certainly achieve this goal with a room full of optics, and in fact we know we could purchase a commercial arrayed waveguide grating device, with a scale of centimeters, to make such a splitter. So our intuitive experience suggests that there is some limit on performance of such optical components, though historically we have not had a general limit we could use.

The need for a limit is not merely academic. With modern nanophotonic techniques, we can make a very broad variety of devices, some with little or no precedent. In part because of the high refractive index contrast available to us in photonic nanostructures, design of such devices is often quite difficult, and we would at least like to know when to stop trying to improve the performance. Classes of devices of interest to us could include dispersive structures, slow light elements, holograms, or any kind of device that separates different kinds of input beams or pulses to different positions in space or time.

Recently, we have been able to devise quite a general approach to limits for the performance of linear optical components[1]. This approach gives upper limits to performance that are quite independent of the details of the design, being dependent instead only on the overall geometry of the device, and, for example, the largest dielectric constant variation anywhere in the structure at any wavelength. This overall limit has already been applied to calculate limits to dispersive devices[1] and to slow light[2]. Here we will introduce this limit, summarizing its derivation and the applications so far.

This limit is based on the idea of counting possible orthogonal wave functions that can be generated when an optical component acts to "scatter" an incident wave into a receiving volume. This idea in turn is based upon some earlier work[3] that is a generalization of diffraction theory to volumes, in which we can count the orthogonal "communications modes" – the best choices of sources in one volume and the resulting waves in another for communicating between the two.

Below, after summarizing the background to the need for a new limit especially in nanophotonics, we will then introduce the underlying mathematical methods, including a discussion of communications modes and their applications. We will give the proof of a new general theorem for strong and/or multiple scattering, a theorem that underlies our limit to optical components. Then we will summarize two applications of the new limit, one to slow light devices and the other to dispersion of pulses, before indicating future directions and drawing conclusions.

## 1.2 Background

There has been extensive prior work on limits or at least design techniques for optical components that are intended to separate beams or pulses (see references in Ref. 1). That prior work, however, largely deduces limits for devices designed in a specific way, such as a simple resonator, coupled resonators, a periodic structure or a grating. For example, we could deduce limits to group delay for a Fabry-Perot resonator, based on an explicit model of such a resonator. Such a limit would show a trade-off between the magnitude of the group delay and the bandwidth over which such delay would be available. In such structures, we can often have relatively simple and sometimes intuitive models of how adjusting some part of the design will lead to a specific consequence. But there is a class of structures that can be made, and that can have relatively very good performance, for which there are no such models. Such structures can result from purely numerical optimization in design.

For example, we recently had set out to design and test superprism wavelength splitters made from one-dimensional dielectric stacks[4,5,6]. In a superprism, the effective angle of propagation of a beam in a structure can vary very strongly with the wavelength of the incident light. Such superprism effects have been known for some time at least for periodic structures[7], and can be understood in that case in terms of the band structure of the periodic system. Better performance, both in the linearity of the shift with wavelength and in the magnitude of the shift, can, however, be obtained from non-periodic structures in which the thickness of each layer is potentially adjusted during the design process[4,5]. It is easy to understand that such a structure might be able to give better performance – there are simply more engineering degrees of freedom available if we do not merely restrict ourselves to designing the unit cell of a periodic structure. We also found in this work that, for over 600 designs with different starting design concepts, different materials, and different optimizations, the performance of all of these lay near or below a specific line[5], suggesting some underlying limit.

In work on designs of two-dimensional structures, in this case for mode-splitting[8,9], we were able to devise quite effective and very compact designs by numerical optimization based on removing or adding dielectric "rods" in a region of the structure, but in this case in particular we simply do not know "how" it works. We cannot say that a particular rod or group of rods does a specific function in a specific way. All the rods interact with the optical field in performing the final function. Since we do not know how it works, we also cannot say how well it could work, or what the limit to it should be based on some analysis of this specific device type. The "exhaustive search" approach of trying all possible designs to establish the best one is generally computationally very expensive, so we would very much like the guidance of some limit that was independent of device details.

The challenge is to devise a limit to the performance of optical components, completely independent of the design approach, so that we can bound the performance not only of the kinds of devices we have used up to now, but for any future device based on any kind of optical structure, including the many possibilities enabled by nanophotonics.

## 1.3 Mathematical approach

Our approach to this limit[1] is based upon counting the number of distinct available channels or "modes" for communicating with an incident wave through a "scattering" volume (the volume that contains our optical component structure of interest) to a receiving volume. As we will see below, with only simple information about the scattering structure, such as its size and shape and the largest dielectric constant variation within it, we can deduce upper limits to this available number of modes. When we are also able to state the number of channels needed for a given optical function, then we can deduce whether that function could possibly be performed by such a structure regardless of how we design it.

### 1.3.1 Communications modes

Before discussing the full problem of scattering waves from one volume to another, we can look at the simpler problem of communicating from sources in a transmitting volume to generate waves in a receiving volume[3]. If those volumes were simply thin parallel surfaces (see Fig. 1), and we were considering waves of

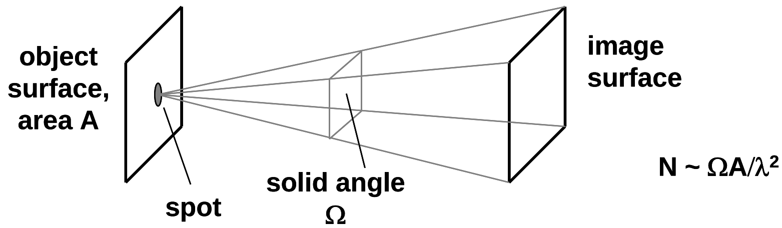


**Figure 1** Illustration of diffraction from an "object" surface to an "image" surface, with the number of distinct spatial channels being given approximately by the number N of resolvable spots as deduced from a simple diffraction model[3].

a specific frequency only, in optics we would fall back on our understanding of diffraction to tell us how many distinct channels there are[10]. We would expect that there would be essentially one distinct channel possible for each resolvable spot, where the size of the resolvable spot is deduced from diffraction theory. Though such an approach is somewhat informal, it is essentially correct for such a problem.

There is a more formal and rigorous approach that was initially understood at least for simple (e.g., square or circular) parallel surfaces[11,12], and that can be extended also to the case of volumes[3]. We sketch this approach here. Consider transmitting and receiving volumes[3], as shown in Fig. 2, which are generalizations of the object surface and image surface respectively. We presume that we have some source function in the transmitting volume $V_T$, and we write this function as $|\psi_T\rangle$. This source function generates a wave, leading specifically to a wave $|\phi_R\rangle = \mathsf{G}_{TR}|\psi_T\rangle$ in the receiving volume $V_R$, where $\mathsf{G}_{TR}$ is the coupling operator (basically the Green's function of the wave equation).

Note, incidentally, that here and below we are using Dirac's "bra-ket" notation as a convenient notation for linear algebra; the reader can think of a "ket" such as $|\psi_T\rangle$ as a column vector whose elements are the values of the function at the various values of the argument of the function, for example. The "bra" vector $\langle\psi_T|$ is then the row vector that is the Hermitian adjoint of $|\psi_T\rangle$. The Hermitian adjoint, which can also be indicated by a superscript dagger "$\dagger$", is the transpose of a vector or matrix in which we also take the complex conjugate of all the elements (so, e.g., $\langle\psi_T| = |\psi_T\rangle^\dagger$). We use "san serif" letters, such as $\mathsf{G}$, to represent operators, which we can think of as matrices.

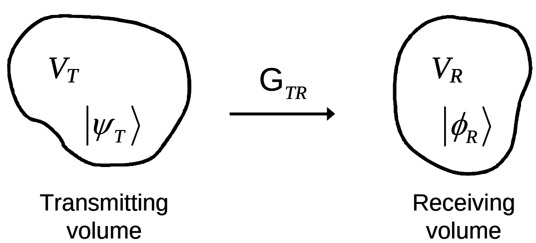


**Figure 2** Illustration of a source function $|\psi_T\rangle$ in a transmitting volume $V_T$, giving rise to a wave $|\phi_R\rangle$ in a receiving volume $V_R$, formally through a coupling operator $\mathsf{G}_{TR}$ (the Green's function of the wave equation).

Suppose then we wish to find the best possible set of distinct (i.e., mathematically orthogonal) source functions in the transmitting volume that would generate the largest possible amplitudes of wave functions in the receiving volume. The general solution to such a problem is known. It is well understood also in the context of imaging between surfaces (see Ref. 13, and references therein). We find the solution by formally performing the singular value decomposition of the coupling operator $\mathsf{G}_{TR}$ between these volumes. The best choices of source functions are the (orthogonal) eigenfunctions $|\psi_{Ti}\rangle$ of the operator $\mathsf{G}_{TR}^{\dagger}\mathsf{G}_{TR}$, with eigenvalues $|s_i|^2$. The corresponding wave functions in the receiving volume are the (orthogonal) eigenfunctions $|\phi_{Ri}\rangle$ of the operator $\mathsf{G}_{TR}\mathsf{G}_{TR}^{\dagger}$, also with eigenvalues $|s_i|^2$. Once we have solved these two eigen problems, we can write $\mathsf{G}_{TR}$ in its singular value decomposition form as

$$\mathsf{G}_{TR} = \sum_i s_i |\phi_{Ri}\rangle\langle\psi_{Ti}| \tag{1}$$

where the $s_i$ are called the singular values.

Specifically, the source function $|\psi_{Tm}\rangle$ leads to a corresponding wave $s_m|\phi_{Rm}\rangle$ in the receiving volume; such a pair of one of these source functions $|\psi_{Tm}\rangle$ in the transmitting volume and its corresponding wave function $|\phi_{Rm}\rangle$ in the receiving volume can be called a "communications mode"[3]. These communications modes represent the best possible set of source and wave function pairs for establishing orthogonal communications channels from one volume to another. If we choose the sets $|\psi_{Ti}\rangle$ and $|\phi_{Ri}\rangle$ also to be normalized, then the singular values represent the coupling strengths of these modes.

When we take such an approach to the coupling between thin, plane-parallel square or circular transmitting (i.e., "object") and receiving (i.e., "image") volumes relatively far apart (so we can use a paraxial approximation in considering wave propagation), then there are specific so-called "prolate spheroidal" functions that are the eigenfunctions of these problems for each surface[11,12]. These functions are not simple small spots on one or other surface – each of the prolate spheroidal functions covers the whole surface – but these functions are truly orthogonal on a given surface, in contrast to a set of spots that are only approximately orthogonal insofar as their "tails" do not overlap very much. These prolate spheroidal functions then form the communications modes of truly orthogonal channels. With such functions, it is known that the singular values are approximately all the same up to some critical number, after which they fall off drastically. That critical number corresponds in this case to the number *N* we would deduce from the idea of counting the number of resolvable spots.

With the general mathematical formalism above for singular value decomposition, we can also solve for communications modes between volumes[3] rather than merely surfaces, and we need not restrict ourselves to communications modes that can be calculated analytically; the numerical prescription for finding the communications modes for arbitrary volumes is quite clear in principle, merely requiring finding eigenfunctions of some specific operators. In this more general approach, we are also not restricted to paraxial approximations, and we can even find the communications modes in near field problems[3,14].

For more arbitrary volumes, it is not in general true that the singular values are all of similar size up to some critical number; that similarity in size of singular values arguably is a characteristic of volumes of uniform thickness[3]. There is, however, another relation that does bound the singular values. It is quite generally true for a linear operator that the sum of the squared modulus of its matrix elements is independent of the (orthonormal) basis sets used to represent it. One way of representing the coupling operator is in terms of its Green's function (which is technically the expansion of the operator on continuous (delta-function) basis sets, one set for each of the transmitting and receiving volumes). Another representation is in terms of the singular value decomposition sets as in Eq. (1) above. Let us take as a concrete simple example the (retarded) Green's function for a free-space monochromatic scalar wave, which is, for a point source at position $\mathbf{r}_T$ in the transmitting volume and a resulting wave at position $\mathbf{r}_R$ in the receiving volume

$$\mathsf{G}_{TR}\left(\mathbf{r}_T,\mathbf{r}_R\right)=\frac{\exp\left(-ik\left|\mathbf{r}_T-\mathbf{r}_R\right|\right)}{4\pi\left|\mathbf{r}_T-\mathbf{r}_R\right|} \tag{2}$$

We can then equate the sum of the modulus squared of the singular values, which is the sum of the modulus squared of the matrix elements of $\mathsf{G}_{TR}$ in its singular value decomposition form, to the integral of the modulus squared of $\mathsf{G}_{TR}\left(\mathbf{r}_T,\mathbf{r}_R\right)$ over the volumes, i.e.,

$$\sum_i\left|s_i\right|^2=\iint_{V_T,V_R}\left|\mathsf{G}_{TR}\left(\mathbf{r}_T,\mathbf{r}_R\right)\right|^2 d^3\mathbf{r}_R d^3\mathbf{r}_T=\frac{1}{(4\pi)^2}\iint_{V_T,V_R}\frac{1}{\left|\mathbf{r}_T-\mathbf{r}_R\right|^2}d^3\mathbf{r}_R d^3\mathbf{r}_T \tag{3}$$

Hence, performing a simple volume integral over the two volumes gives us an absolute upper bound to the sum of the squares of the coupling strengths between the volumes[3]. Even if the coupling strengths of the communications modes do not have the simple form of being approximately constant up to some value, there is still a limit on the sum of their squares. We can view this statement as being a generalization of the concept of the "diffraction limit", now for arbitrary three-dimensional volumes. This limit does agree with the specific results for planar surfaces above. Note then that if we try to defeat the diffraction limit, we will necessarily have to use communications modes that are very weakly coupled – in other words, we will need relatively very large source amplitudes.

This concept of communications modes has recently seen a number of applications in optics[13,14,15,16,17]. The theory has also been generalized to vector electromagnetic waves[18], and it is helpful in understanding the limits to the synthesis of light fields in three dimensions[19]. It has also proved useful in analyzing wireless communications[20,21,22], where the transmitting and receiving antennas and volumes are not plane surface.

The communications mode concept also illustrates some of the power of considering waves between volumes in terms of orthogonal source and wave functions. The theorem below expands on this approach for a quite different application.

### 1.3.2 New theorem for strong or multiple scattering

As mentioned above, we will think of our optical component in general as a "scatterer" – that is, there is some incident wave that is "scattered" by the optical component, generating a wave in the receiving volume. Our scattering problem here is particularly severe; we want to consider arbitrarily strong and/or multiple scattering by the object, including arbitrarily large and/or abrupt changes in dielectric constants. As a result, we cannot proceed simply by adding up all of the successive scatterings in some series – such a series will typically not converge. Below we will derive the core theorem that we use to deal with the problem. This derivation closely follows that of Ref. 1.

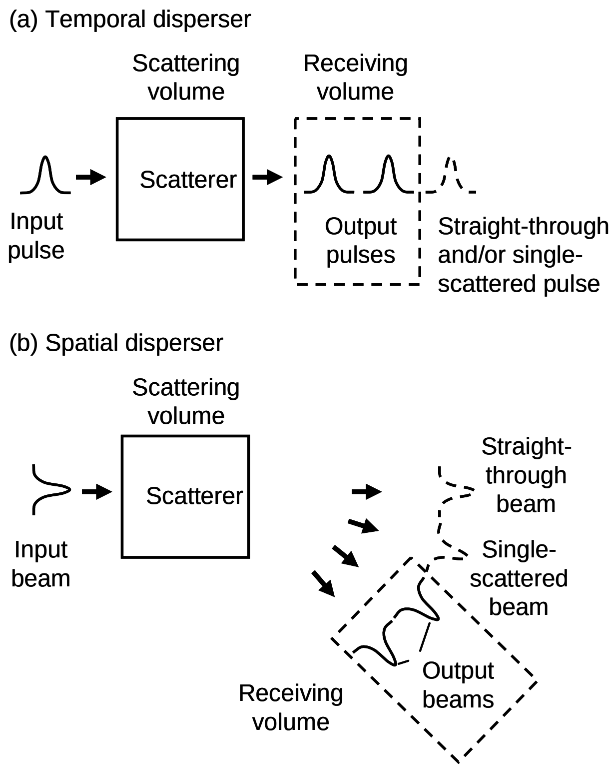


**Figure 3** Illustration of scattered (a) pulses, and (b) beams for temporal and spatial dispersers respectively. The "straight-through" and single-scattered pulses or beams may not actually be present physically, but the theory first considers outputs that are orthogonal to what both of these would be mathematically. Here we show the case where the "straight-through" and single-scattered pulses or beams miss the receiving volume, though in general they may not[1].

We consider (Fig. 3) two volumes, a scattering volume $V_S$ that contains a scatterer that is our optical device (such as a dielectric with dielectric constant that can vary strongly within the volume), and a receiving volume $V_R$ in which we want to generate waves.

An incident wave (the input pulse or input beam in Fig. 3) will lead to some net wave within the scattering volume. That net wave that will interact with the scatterer to generate some resulting effective source in the scatterer. A specific net source $|\psi_{Sm}\rangle$ in the scattering space will generate a wave $|\phi_{SCm}\rangle$ within the scattering space through the Green's function operator $\mathsf{G}_S$ within the scattering space, i.e.,

$$|\phi_{SCm}\rangle = \mathsf{G}_S |\psi_{Sm}\rangle \tag{4}$$

Here $|\psi_{Sm}\rangle$ represents *all* sources in the volume. Hence $\mathsf{G}_S$ is the "free-space" Green's function. We presume here that the only sources are the induced polarizations or currents generated as a result of the interaction between the net wave and the scattering material; we have no other sources in the scattering volume. We presume some incident wave $|\phi_{Im}\rangle$ caused all of these sources through its scattering. The net wave $|\phi_{Sm}\rangle$ in the space must be the sum of the incident and scattered waves, i.e.,

$$|\phi_{Sm}\rangle = |\phi_{Im}\rangle + |\phi_{SCm}\rangle \tag{5}$$

Any wave $|\phi\rangle$, interacting with the scatterer, will in turn give rise to sources $|\psi\rangle$ through some other linear operator $\mathsf{C}$, an operator that we can think of simply as representing the dielectric constant of the material, for example. So

$$|\psi\rangle = \mathsf{C}|\phi\rangle \tag{6}$$

We must have self-consistency, and so we require that the net source in the scattering volume, $|\psi_{Sm}\rangle$, is the one that would be generated by the net wave $|\phi_{Sm}\rangle$ interacting with the scatterer. Hence

$$\begin{aligned} |\psi_{Sm}\rangle &= \mathsf{C}|\phi_{Sm}\rangle = \mathsf{C}|\phi_{Im}\rangle + \mathsf{C}|\phi_{SCm}\rangle \\ &= \mathsf{C}|\phi_{Im}\rangle + \mathsf{C}\mathsf{G}_S|\psi_{Sm}\rangle = \mathsf{C}|\phi_{Im}\rangle + \mathsf{A}_S|\psi_{Sm}\rangle \end{aligned} \tag{7}$$

where

$$\mathsf{A}_S = \mathsf{C}\mathsf{G}_S \tag{8}$$

If we presume now that we have some specific source function $|\psi_{Sm}\rangle$ in the scattering space, then there also be some linear operator $\mathsf{G}_{SR}$ (again a "free-space" Green's function, like the $\mathsf{G}_{TR}$ considered above for the communications mode problem) that we could use to deduce the resulting wave $|\phi_{Rm}\rangle$ in the receiving space, i.e.,

$$|\phi_{Rm}\rangle = \mathsf{G}_{SR}|\psi_{Sm}\rangle \tag{9}$$

Given that we want to separate out light beams or pulses into some receiving space, we ask that the various waves we generate in the receiving space are mathematically orthogonal, just as we were considering above for the communications modes between two volumes. We will try to deduce some limit on the number of such orthogonal functions $|\phi_{Rm}\rangle$ that can be generated in the receiving space. In particular here, we will use the sets of functions $|\psi_{Sm}\rangle$ and $|\phi_{Rm}\rangle$ that are the communications modes between the scattering and receiving spaces, i.e., the singular value decomposition of $\mathsf{G}_{SR}$, and specifically we will restrict these sets to those elements corresponding to non-zero singular values, i.e., we only want to consider source functions that give rise to non-zero wave amplitudes in the receiving space.

We will separate the counting of orthogonal waves into two parts. Specifically, we will come back later and consider the waves that correspond to "straight-through" or "single-scattered" waves. "Straight-through" waves are the waves that would exist in the receiving space in the absence of any scatterer; they correspond to propagation of the incident wave straight through the scattering volume. "Single-scattered" waves are the waves that would hypothetically arise from the scattering of the incident wave if it were imagined to be completely unchanged by its interaction with the scatterer, i.e. formally a wave $\mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle$. We will be interested for the moment only in waves in the receiving space that are formally orthogonal to both the "straight-through" and "single-scattered" hypothetical waves. These concepts are sketched in Fig. 3. This neglect of "straight-through" and "single-scattered" means we are only counting those orthogonal waves in the receiving volume that are the result of strong and/or multiple scattering in the scattering volume.

Our neglect of "straight-through" waves means that Eq. (9) gives the total wave in the receiving space, and so that wave becomes

$$|\phi_{Rm}\rangle = \mathsf{G}_{SR}|\psi_{Sm}\rangle = \mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle + \mathsf{G}_{SR}\mathsf{A}_S|\psi_{Sm}\rangle \tag{10}$$

Because we presume we are only interested in scattered waves $|\phi_{Rm}\rangle$ that are orthogonal to the "single-scattered" wave $|\phi_{RIm}\rangle = \mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle$, by definition we have

$$\langle\phi_{Rm}|\mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle = 0 \tag{11}$$

Hence, from Eqs. (10) and (11)

$$\begin{aligned}\langle\phi_{Rm}|\phi_{Rm}\rangle &= \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle \\ &= 0 + \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}\mathsf{A}_S|\psi_{Sm}\rangle\end{aligned} \tag{12}$$

Now, since the $|\psi_{Sm}\rangle$ are by definition complete and orthonormal for the source space of interest, we can introduce the identity operator for that space, which we can write as $\mathsf{I}_{HS} = \sum_j |\psi_{Sj}\rangle\langle\psi_{Sj}|$, to obtain from Eq. (12)

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle=\sum_j\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sj}\rangle\langle\psi_{Sj}|\mathsf{A}_S|\psi_{Sm}\rangle$$
$$=\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle \quad (13)$$

where in the last step we have used the fact that

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sj}\rangle=\langle\phi_{Rm}|\phi_{Rj}\rangle=0 \text{ unless } m=j \quad (14)$$

because of the orthogonality we are enforcing for the waves $|\phi_{Rm}\rangle$ we want to generate in the receiving volume. Hence we come to the surprising conclusion from Eq. (13) that, for each $m$ for which $|\psi_{Sm}\rangle$ gives rise to a non-zero wave in the receiving space, i.e., for which

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle\neq 0 \quad (15)$$

then

$$\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle=1 \quad (16)$$

and hence, trivially,

$$\left|\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle\right|^2=1 \quad (17)$$

If we could separately establish a result of the form

$$\sum_i\left|\langle\psi_{Si}|\mathsf{A}_S|\psi_{Si}\rangle\right|^2\leq S_A \quad (18)$$

for some finite number $S_A$, then we would conclude that the maximum number $M$ of possible orthogonal waves that could be generated in the receiving space by our strong scattering is

$$M\leq S_A \quad (19)$$

An important additional result[1], which we will not prove here, is that we can quite generally evaluate such a limit, and we can split it into two parts. Specifically, we can write

$$M\leq\sqrt{N_C N_{GS}} \quad (20)$$

Here

$$N_C\equiv Tr\left(\mathsf{C}^{\dagger}\mathsf{C}\right) \quad (21)$$

is essentially the integral of the modulus squared of the dielectric constant variation in the structure, and

$$N_{GS}\equiv Tr\left(\mathsf{G}_S^{\dagger}\mathsf{G}_S\right) \quad (22)$$

is essentially the integral of the modulus squared of the wave equation's Green's function within the scattering volume.

Hence we can see the core of a remarkable bound on the possible performance of optical components. First, there is a number $M$ we can evaluate, based only on average properties of the dielectric medium and the shape of the scattering volume. Second, that number tells us an upper bound on the number of orthogonal functions that can be generated as a result of multiple and/or strong scattering into the receiving volume (technically, we are only counting those functions that are orthogonal to the "single-scattered" or "straight-through" waves, though for many strong scattering problems we will be able to deal with these separately or discount them altogether). For each such orthogonal function we want to be able to generate, we use up one unit of "strength" of the scatterer. Since the scatterer only overall has a finite "strength" $M$, there is a bound to the number of such orthogonal functions we can generate by this scattering into the receiving volume.

Note a key difference between this limit and the limit calculated above for the communications modes in Eq. (3). The communications modes limit was a bound on the sum of the strengths of the couplings between the transmitting and receiving volumes. It did not tell us directly the number of the communications modes that would be strongly coupled. We needed some other criterion to determine the actual coupling strengths of such strongly coupled modes. The new scattering limit we have proved here, Eq. (20), gives us an upper bound on the number of possible orthogonal functions that can be generated as a result of scattering from the scattering volume into the receiving volume, independent of the strengths of those scatterings (as long as they are non-zero). Once we can say how many functions we need to be able to control in the receiving volume for our optical component (scatterer) to have performed our desired optical operation, then we can use our new limit to tell us if that operation is impossible with an optical component of a given volume and given range of dielectric constants.

## 1.4 Limit to the performance of linear optical components

The use of this limit involves two steps. First, for some class of optical structures and problems of interest, we need to evaluate $\sqrt{N_C N_{GS}}$ . Second, we need to think of how to express the optical operation we wish to perform in terms of the number of orthogonal functions we will need to be able to control in the receiving space. So far, in our work we have considered both of these steps for the simple case of one-dimensional systems, i.e., systems like dielectric stack structures, a beam within a medium (such as an atomic vapor) or, with an appropriate renormalization to allow for mode overlap, single-mode guided wave structures, and we will summarize some results below.

### 1.4.1 Explicit limit for one-dimensional systems

As shown in Ref. 1, we can obtain simple explicit results for one-dimensional systems, i.e., any systems that can be described by a wave equation for a wave of frequency $f_o$ that can be written

$$\frac{d^2\phi}{dz^2}+k_o^2\phi=-k_o^2\eta\left(z,f_o\right)\phi \tag{23}$$

Here $k_o=2\pi/\lambda_o=2\pi f_o/v_o$, where $v_o$ is the wave velocity and $\lambda_o$ is the wavelength, both in the background medium. This is an appropriate equation for electromagnetic waves in one-dimensional problems in isotropic, non-magnetic materials with no free charge or free currents. Then $\eta$ is the fractional variation in the relative dielectric constant in the structure, i.e.,

$$\eta\left(z,f_o\right)\equiv\frac{\Delta\varepsilon\left(z,f_o\right)}{\varepsilon_{ro}} \tag{24}$$

where $\varepsilon_{ro}$ is the background relative dielectric constant, the wave velocity in the background medium is $v_o=c/\sqrt{\varepsilon_{ro}}$, where $c$ is the velocity of light, and for a relative dielectric constant $\varepsilon\left(z,f_o\right)$, we define $\Delta\varepsilon=\varepsilon\left(z,f_o\right)-\varepsilon_{ro}$. (Note that $\varepsilon$ may be complex.) With appropriate re-scaling of the dielectric constant variation to include mode overlap, such an approach can also be taken for any single-mode system, such as a single-mode waveguide.

We will restrict ourselves here to situations where the frequency bandwidth $\delta f$ of interest is much less than the center frequency $f_c$, and where the thickness $L$ of the scattering medium is much larger than the wavelength $\lambda_c=v_o/f_c$ in the background material at the center frequency. Here we presume the scattering structure outputs pulses by transmission into a receiving space that is "behind" the slow light structure, as in Fig. 3 and Fig. 4. (The case of reflection rather than transmission is also easily handled, and gives similar results[1,2].) We also allow the receiving space thickness, $\Delta z_R$, to be arbitrarily long so that it will capture any possible orthogonal function that results from scattering. With these simplifying restrictions, we can evaluate the quantities $Tr\left(\mathsf{C}^\dagger\mathsf{C}\right)\le n_{tot}\eta_{\max}^2$, where $\eta_{\max}$ is the maximum value of $|\eta|$ at any frequency within the band of interest at any position within the scatterer, and $Tr\left(\mathsf{G}_S^\dagger\mathsf{G}_S\right)=n_{tot}(\pi^2/3)\left(L/\lambda_c\right)^2$. Here $n_{tot}=2\delta f\Delta z_R/v_o$ is the number of degrees of freedom required to define a function of bandwidth $\delta f$ over a time $\tau=\Delta z_R/v_o$. The resulting $M$ from Eq. (20) becomes

$$M\le n_{tot}\frac{\pi}{\sqrt{3}}\frac{L}{\lambda_c}\eta_{\max} \tag{25}$$

If the scatterer has a similar range of variation of $\eta$ over the entire scattering volume and if there are no dielectric constant resonances that are sharp compared to the frequency band of interest, then we can use the root mean square variation, $\eta_{rms}$, of the magnitude of $\eta$, averaged over position and frequency[1] instead of $\eta_{\max}$ in Eq. (25) and the expressions that follow below.

### 1.4.2 Slow light limit

A particularly clear and simple example of the application of this limit[2] is to the problem of slow light. Our approach here follows that of Ref. 2. We would like

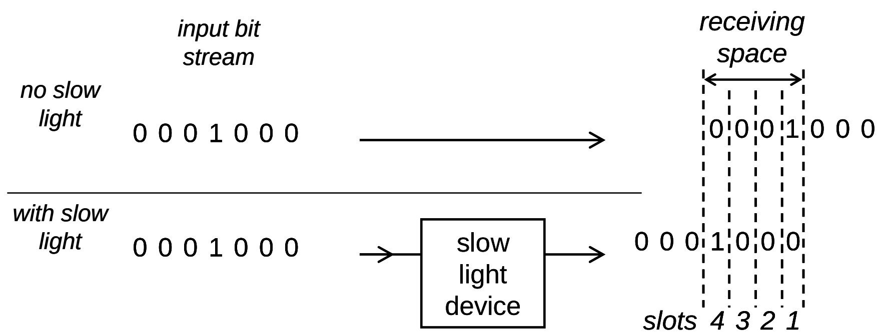


**Figure 4**. Illustration of a bit pattern delayed by 3 bit periods by scattering in a slow light device. In the design of the device, the scattering into a total of 4 bit slots has to be controlled so that the "1" appears in slot 4, and "0"s appear in each of slots 3, 2, and 1, hence requiring the control of 4 orthogonal functions in the receiving space[2].

to understand for some linear optical system just what are the limits to the amount of delay we can get of some bit stream, in particular the number of bits of delay. We obtain quite a simple and general upper bound answer, an answer that does not even depend on the bandwidth of the slow light system. This approach does not require that we assume any particular pulse shape, and, unlike other approaches to slow light limits[23,24,25], it does not rely on the concept of group velocity, a concept that has limited meaning within anything other than a uniform or periodic structure.

We note that we need $N$ linearly independent functions to represent an arbitrary $N$ bit binary number in a receiving space, and so the basis set of physical wave functions used to represent the number in the receiving space must have at least $N$ orthogonal elements. Suppose we have a incident bit stream with a logical "1" surrounded by logical "0"s (see Fig. 4). Without a slow light device, the bit stream propagates through to the receiving space. With the slow light device, however, we want the bit stream to be shifted, so that the "1" appears in a later bit period. To obtain a delay of $S$ bit periods, in the scattering in the slow light device we need to be able to control the amplitudes of at least $S$+1 orthogonal physical functions in the receiving space, so that we can center a function representing a "1" in the ($S$+1)th bit period, and functions representing "0"s centered in the other $S$ bit periods. We presumably control the amplitudes of these various functions by the design of the slow light device (the "scatterer").

In optics we typically consider pulses on a carrier frequency $f_c$. Then we note that there will be two different but almost identical pulses that have essentially the same amplitude envelope, but that are formally orthogonal only because they have a carrier phase that differs by 90 degrees. Since typically we look only at the amplitude envelope, we then need to double the number of amplitudes we control in the bit periods containing logic "0"s, so that both of these pulses are "low" (i.e., logic "0"s). Likely we do not care about the carrier phase of the pulse in the desired slot, so we need not add in another degree of freedom to control that. In this case, therefore, we need to control 2$S$+1 orthogonal functions in the receiving space to delay a pulse or bit stream by $S$ bit periods.

One subtlety that we have to deal with in applying this limit is in the counting of available orthogonal functions as given by Eq. (25). Because we chose the receiving space to be arbitrarily long, we have included in $M$ as separate possibilities the scattering not only of the pulse of interest, but also of every distinct delayed version of it. There are $n_{tot}$ such different delayed versions of the same scattering that fit in the receiving space. Since we need consider only one of these, we can remove the factor $n_{tot}$ below.

We also previously noted that this number $M$ is the number of orthogonal waves possible in the receiving volume that are also orthogonal to the "straight-through" and "single-scattered" waves. At best, for a transmission device, considering these other two waves could at most add in two more available controllable degrees of freedom. Hence, we finally obtain for the upper limit to the number of accessible orthogonal functions in the receiving space for the scattering of a single pulse

$$M_{tot} \leq 2 + \frac{\pi}{\sqrt{3}} \frac{L}{\lambda_c} \eta_{\max} \qquad (26)$$

Given that we need to control at least $2S+1$ amplitudes to delay by $S$ bits, we must therefore have $M_{tot} \geq 2S+1$. Hence, the maximum delay $S_{\max}$ in bit periods that we can have is

$$S_{\max} \leq \frac{1}{2} + \frac{\pi}{2\sqrt{3}} \frac{L}{\lambda_c} \eta_{\max} \qquad (27)$$

Suppose, for example, we want to delay by 32 bits (i.e., $S = 32$), we work with a layered structure of glass ($\varepsilon_r = 2.25$ ) and air ($\varepsilon_r = 1$ ) so $\eta_{\max} = 1.25$, and we choose a center wavelength of 1.55 microns. Then

$$L \geq \left( S - \frac{1}{2} \right) \frac{2\sqrt{3}}{\pi \eta_{\max}} \lambda_c = 43\ \mu\text{m} \qquad (28)$$

Note that this limit cannot be exceeded for a linear one-dimensional fixed glass/air optical structure, no matter how we design it; we would always need at least 43 μm of thickness.

### 1.4.3 Limit to dispersion of pulses

We can also calculate in a similar way an upper bound to a structure that we wish to have separate out pulses of different center wavelengths to different delays in time. The counting of the required number of orthogonal functions here is somewhat more complicated, so we omit the details in this summary. In the end, we achieve a rather similar result to that above for the slow light result. In this case, the bound for the number $N_B$ of different wavelengths of pulses that can be separated in time is given by

$$N_B \le \frac{3}{2} + \frac{\pi}{2\sqrt{3}} \frac{L}{\lambda_c} \eta_{\max} \tag{29}$$

for a transmissive device. Hence, with a similar calculation to that above for the slow light structure, we would have, for a device to separate pulses of 32 different center wavelengths using a structure made of air and glass, that we must have the length $L \ge 41.7\ \mu\text{m}$. Again this limit is independent of how we design this glass/air layered structure.

## 1.5 Future directions

The underlying limit to the performance of optical components, Eq. (20), is expressed in very general terms. It applies to any kind of linear wave interacting with a fixed medium, including scalar waves such as sound waves, vector electromagnetic waves, and even quantum mechanical waves. So far here we have only evaluated specific limits for one-dimensional structures, but again the general form here should also give limits for two-dimensional and three-dimensional structures once the appropriate evaluation is performed for the quantity $N_{GS} \equiv Tr\left(\mathsf{G}_S^\dagger \mathsf{G}_S\right)$ of Eq. (22) for the two- and three-dimensional Green's functions for the relevant wave equations.

We can therefore expect additional specific limits for two- and three-dimensional optical structures. Since we have only considered one-dimensional structures so far, we have also only considered problems for optical pulses, because there is no different "position" other than a temporal one to which we can scatter waves when there is only one beam possible. With results for two- and three-dimensional situations, we could also examine limits to static scattering problems, such as limits to high-contrast holograms. Three-dimensional results for the vector electromagnetic case could also be particularly interesting for analyzing wireless communications in strongly scattering environments. In general, there is a broad range of additional wave problems to which we could apply this kind of approach to limits.

## 1.6 Conclusions

The body of work summarized here shows that there are many applications in optics for the idea of considering optics in terms of the number of orthogonal waves or communications modes that can be supported or generated in an optical system. Not only does this approach lead to a precise notion of the idea of diffraction limits, even beyond the simple plane-to-plane case of optical imaging or focusing to the case of volumes and near-field problems, but it also leads to well-defined and very general limits to the performance of optical systems and components. The new limit recently proposed[1] to the performance of optical devices, even those made with arbitrary structures of high index contrast, gives us for the first time a bounding limit to the performance of linear optical devices, completely independent of how we design them. We look forward to many novel applications and implications of this approach and the resulting bounds in optics.

## References


1. Miller, D. A. B., "Fundamental limit for optical components," *J. Opt. Soc. Am. B* **24**, A1–A18 (2007).
2. Miller, D. A. B., “Fundamental Limit to Linear One-Dimensional Slow Light Structures,” Phys. Rev. Lett. **99**, 203903 (2007) [doi:10.1103/PhysRevLett.99.203903]
3. Miller, D. A. B., “Communicating with waves between volumes – evaluating orthogonal spatial channels and limits on coupling strengths,” *Appl. Opt.* **39**, 1681–1699 (2000)
4. Gerken, M. and D. A. B. Miller “Multilayer thin-film structures with high spatial dispersion,” *Appl. Opt.* **42**, 1330–1345 (2003)
5. Gerken, M. and D. A. B. Miller, “Limits to the performance of dispersive thin-film stacks,” *Appl. Opt.* **44**, 3349–3357 (2005)
6. Gerken, M. and D. A. B. Miller, “The Relationship between the superprism effect in one-dimensional photonic crystals and spatial dispersion in non-periodic thin-film stacks,” *Optics Lett.* **30**, 2475–2477 (2005).
7. Zengerle, R., “Light propagation in singly and doubly periodic planar waveguides,” *J. Mod. Opt.* **34**, 1589–1617 (1987)
8. Jiao, Y., S.-H. Fan, and D. A. B. Miller, “Demonstration of systematic photonic crystal device design and optimization by low rank adjustments: an extremely compact mode separator,” *Optics Lett.* **30**, Issue 2, 141–143 (2005)
9. Jiao, Y., S.-H. Fan, and D. A. B. Miller, “Systematic photonic crystal device design: global and local optimization and sensitivity analysis,” *IEEE J. Quantum Electron.* **42**, No. 3. 266–279 (2006)
10. D. Gabor, "Light and information", *Progress in Optics I*, ed. E. Wolf (North-Holland, Amsterdam, 1961), pp. 109–153
11. G. Toraldo di Francia, "Degrees of freedom of an image", *J. Opt. Soc. Am.* **59**, 799–804 (1969)
12. B. R. Frieden, "Evaluation, design and extrapolation methods for optical signals, based on the use of the prolate functions", *Progress in Optics IX*, ed. E. Wolf, (North-Holland, Amsterdam, 1971), pp. 311–407
13. Burvall, A., H. H. Barrett, C. Dainty, and K. J. Myers, "Singular-value decomposition for through-focus imaging systems," *J. Opt. Soc. Am. A* **23**, 2440–2448 (2006)
14. Martinsson, P., H. Lajunen, and A. T. Friberg, "Scanning optical near-field resolution analyzed in terms of communication modes," *Opt. Express* **14**, 11392–11401 (2006)
15. Thaning, A., P. Martinsson, M. Karelin, and A. T. Friberg, “Limits of diffractive optics by communications modes,” *J. Opt. A: Pure Appl. Opt.* **5**, 153–158 (2003)

16. Burvall, A., P. Martinsson, and A. Friberg, “Communication modes applied to axicons,” *Opt. Express* **12**, 377–383 (2004)
17. Burvall, A., P. Martinsson, and A. T. Friberg, "Communication modes in large-aperture approximation," *Opt. Lett.* **32**, 611–613 (2007)
18. Piestun, R. and D. A. B. Miller, “Electromagnetic degrees of freedom of an optical system,” *J. Opt. Soc. Am. A* **17**, 892–902 (2000)
19. Piestun, R. and J. Shamir, “Synthesis of three-dimensional light fields and applications,” *Proc. IEEE* **90**, 222–224 (2002)
20. Godavarti M. and A. O. Hero, III, “Diversity and degrees of freedom in wireless communications,” *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 02)*, pp. III-2861–III-2864 (2002)
21. Xu J. and R. Janaswamy, “Electromagnetic degrees of freedom in 2-D scattering environments,” *IEEE Trans. Antennas and Propagation* **54**, 3882–3894 (2006)
22. Hanlen L. and M. Fu, “Wireless communication systems with spatial diversity: a volumetric model,” *IEEE Trans. Wireless Commun.* **5**, 133–142 (2006)
23. R. S. Tucker, P.-C. Ku, and C. J. Chang-Hasnain, “Slow-Light Optical Buffers: Capabilities and Fundamental Limitations,” *J. Lightwave Technol.* **23**, 4046–4066 (2005)
24. R. W. Boyd, D. J. Gauthier, A. L. Gaeta, and A. E. Willner, “Maximum time delay achievable on propagation through a slow-light medium,” *Phys. Rev. A* **71**, 023801 (2005)
25. J. B. Khurgin, "Performance limits of delay lines based on optical amplifiers," *Opt. Lett.* **31**, 948–950 (2006)

David A. B. Miller received the B.Sc. degree from St Andrews University, and, in 1979, the Ph.D. degree from Heriot-Watt University, both in Physics. He was with Bell Laboratories from 1981 to 1996, as a department head from 1987, latterly of the Advanced Photonics Research Department. He is currently the W. M. Keck Professor of Electrical Engineering, the Director of the Solid State and Photonics Laboratory, and a Co-Director of the Stanford Photonics Research Center at Stanford University, Stanford, California, USA. His research interests include physics and devices in nanophotonics, nanometallics, and quantum-well optoelectronics, and fundamentals and applications of optics in information sensing, switching, and processing. He has published more than 200 scientific papers, holds 62 patents, is a Fellow of OSA, IEEE, APS, and the Royal Societies of Edinburgh and London, holds honorary degrees from the Vrije Universiteit Brussel and Heriot-Watt University, and has received numerous awards.