# Fundamental limit for optical components

**David A. B. Miller***

*Ginzton Laboratory, Stanford University, 450 Via Palou, Stanford, California 94305-4088, USA*
**Corresponding author: dabm@ee.stanford.edu*

Received November 6, 2006; accepted December 17, 2006;
posted January 18, 2007 (Doc. ID 76778); published July 19, 2007

We show that there is a general limit to the performance of linear optical components, based only on their size, shape, and dielectric constants. The limit is otherwise independent of the design. The mathematics involved applies generally to linear systems with arbitrarily strong multiple scattering. Relevant optical structures include dielectric stacks, photonic crystals, nanometallics, metamaterials, and slow-light structures. The limit also covers acoustic and quantum-mechanical waves, and electromagnetic waves of any frequency. In an example, a one-dimensional glass/air structure, a thickness of at least 41.7 $\mu$m is required for the separation of pulses of 32 different frequencies near 1.55 $\mu$m center wavelength. Larger available dielectric constants would lead to correspondingly shorter limits. © 2007 Optical Society of America

*OCIS codes:* 290.4210, 260.2030, 350.7420, 230.4170.

## 1. INTRODUCTION

It is the function of many linear optical components to separate different kinds of light beams. A filter, such as a Fabry–Perot cavity or a complex dielectric stack structure, for example, or any wavelength dispersive device, separates waves of different frequencies to different output positions. Various dispersive or delay devices exist that could separate pulses of different center frequencies to different points in space or delays in time. A hologram might be designed to send waves of different shapes to different positions. Slow-light devices delay pulses to quite different arrival times.

At least in practice, trade-offs appear to be inherent in many, if not all, such linear optical components. A Fabry–Perot resonator used as a group delay device, for example, near its resonance, can provide a large group delay over a small wavelength range, or a smaller group delay over a larger wavelength range. In general, we appear not to know a simple answer to some key questions about such trade-offs and limits. For example, if we have a piece of material or some combination of materials of a given total volume, and we are allowed to structure it in any way we want, can we or can we not make a device that will, say, separate 32 different wavelengths to distinct output positions? Or, given a particular volume of material in which the dielectric constant can be varied arbitrarily within some range, is there some general limit we can write down on the number of optical pulses of different center wavelengths that can be separated out in time, or the number of incident spatial modes in a multimode light beam that can be separated out to different single-mode light beams? Our experience may suggest to us that there ought to be some quite general limiting answer to such questions, but no such general limit is apparently known.

There is substantial literature devoted to limits, or at least design techniques, in various different optical components whose function is to disperse beams or pulses in space or time, or to perform the inverse operation of dispersion compensation (see [1–35] for representative literature). Dispersive devices based on single or multiple resonators can be used for dispersion compensation or other functions in optical communications [1–4]. Photonic crystals [5–9] and other high-index-contrast photonic nanostructures [10–17] can be used for wavelength or mode [16,17] demultiplexing. In the use of optical holograms for information storage, we want to know how many different "pages" or beams we can generate at the output in response to different input beams [18–23]. Slow-light devices would delay light pulses for optical buffering [24–35]. Usually, however, the calculated limits to the performance of such components are based on analyses specific to the type of design being used, or are based on simplified models that apply only in limited regimes (e.g., small refractive index contrast). Often, the limit is given on the basis of some "distortion" or "cross talk" in the output pulse or beam, which requires that some arbitrary criterion for distortion or cross talk be defined; such limits are again specific to the particular device or pulse type being studied.

The recent interest in nanophotonic structures and the proliferation of new fabrication methods raise practical possibilities for new generations of devices, with many new opportunities for making devices with structures that can be quite arbitrary. Many classes of structures are now possible, including dielectric stacks; one-, two-, and three-dimensional photonic crystals; nanometallics; and metamaterials. Devices based on such structures will also likely involve trade-offs of some kind, however, as has already been observed empirically, e.g., by Gerken and Miller [12]. In this work, over 600 quite different and relatively arbitrary multilayer dielectric stack structures were designed for a specific purpose (linear spatial beam dispersion with wavelength by nanophotonic superprism phenomena), and all were found to lie near or under a specific empirical curve. Again, there is no known general limit to explain this.

A general limit could tell us whether some proposed device could simply not be designed, and could give us scal-

 

ing rules to tell us how hard it could be to make specific classes of devices of given performances. Given the difficulty of design of complex functions in nanophotonics, a guiding limit like this could be very useful. Our goal here is to find such a general limit valid for broad classes of optical devices.

Here, we derive a fundamental limit on the performance of the linear optical components that separate, disperse, or delay light beams. The limit is independent of the details of the device structure; in fact, just a knowledge of the maximum dielectric constant is sufficient for one version of the limit. The limit does not explicitly say how much a pulse can be delayed, or through what angle a beam can be deflected, but rather how many distinct positions in time or space are possible, which may be a more fundamental point.

The limit we derive is very broadly defined. In its most general form, it applies to a very wide class of linear systems, not merely optical ones; it would apply, for example, to scattering of acoustic or quantum-mechanical waves, as well as to electromagnetic waves of all frequencies. In particular, it gives a useful limit to the performance of systems that rely on multiple scattering of linear waves to perform their function.

In Section 2, we derive the central mathematical results, which can be expressed in a general linear algebra notation valid for a broad range of linear waves. In Section 3, we illustrate the limit with the example of one-dimensional dielectric structures in the simplest, scalar-wave approximation. This will expose most of the key features of this approach, as well as providing practically useful results. As a concrete example, we explicitly calculate the minimum thickness of a glass structure to separate pulses of different frequencies in the telecommunications C-band. We draw conclusions in Section 4.

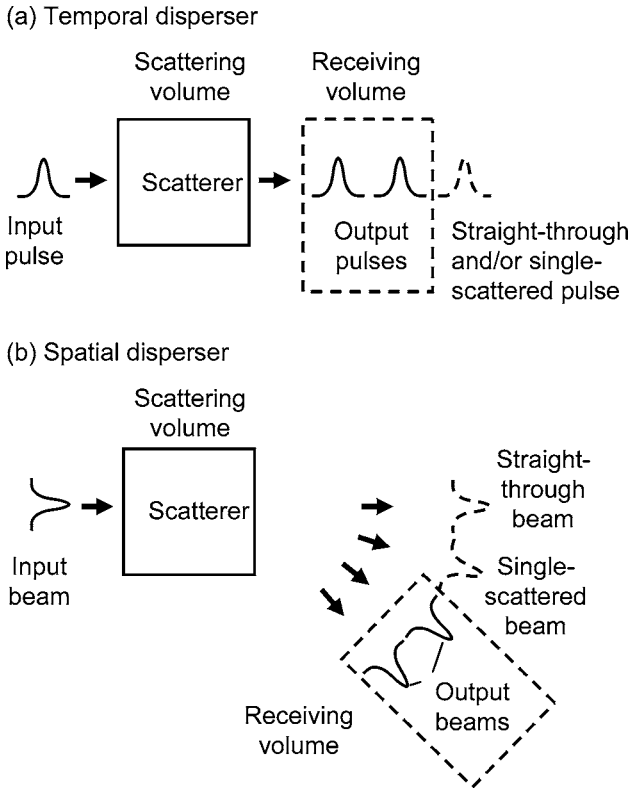


Fig. 1. Illustration of scattered (a) pulses and (b) beams for temporal and spatial dispersers, respectively. The straight-through and single-scattered pulses or beams may not actually be present physically, but the theory first considers outputs that are orthogonal to what both of these would be mathematically. 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.

## 2. DERIVATION OF THE CENTRAL RESULT

We specifically want to derive limits that are useful in the case of arbitrarily strong scattering or interaction, such as could result from optical structures with large contrasts in the dielectric constant between different parts of the structure, or as in some dense atomic vapor. A problem with constructing any such general limit for strong scatterers is that approaches such as summing perturbation series do not work—the series will typically not converge. Hence a different approach is needed. The approach we take here is based on deducing a limit $M$ to the number of orthogonal functions that can be generated in a receiving volume as a result of scattering of a wave from a scattering volume. A key result of this paper is that an upper limit on $M$ can be deduced based on limits to quantities like the range of dielectric constant variation in the volume, irrespective of the details of the design of the structure. Knowing that limit, we will be able to use it to deduce limits for specific optical functions, such as dispersive elements.

We consider two volumes, a scattering volume $V_S$ that contains a scatterer (e.g., 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 (see Fig. 1). We have previously had success [36–38] with the simpler problem of understanding the orthogonal "communications modes" that couple one volume to another. This led to a generalization of diffraction theory and a universal sum rule on the strengths of couplings between source modes in one volume and waves in another. Here we take a related approach, though we need to go substantially further by including the multiple scattering within $V_S$.

Because we want to consider not only physical volume, but also time and frequency, we will generalize to considering mathematical (Hilbert) spaces that include all of these attributes and possibly others. Specifically, for example, we can consider some scattering space $H_S$ that includes all possible functions of interest to us in the scattering volume, possibly restricted to some specific range of times, and possibly also restricted to some range of frequencies. Similarly we can consider some receiving space $H_R$ that includes all possible functions of interest to us in the receiving volume, functions that again might be restricted to some time and/or frequency range. This general mathematical approach can cover more complicated functions with more attributes, such as a full vector electromagnetic field including polarizations or quantum-mechanical fields with attributes such as spin, and the theorems we prove include such cases. We will not, however, include such attributes in the specific one-dimensional example we give at the end of this paper.

### A. Mathematical Notation

For the general mathematics of this linear problem, we will use a Dirac notation. In such a notation, $|\psi\rangle$ represents a function in a particular space, and we can if we wish visualize it as a column vector. $\langle\psi|$ is the Hermitian adjoint of $|\psi\rangle$ and can be visualized as the corresponding transposed row vector with complex conjugated elements. We will write linear operators, which we can visualize as matrices, with sans serif characters, e.g., $\mathsf{B}$. The action of such an operator on a function $|\psi\rangle$ is to generate another function $|\phi\rangle = \mathsf{B}|\psi\rangle$. Use of this notation is typical for quantum mechanics, though here we use it merely as a convenient notation for linear algebra that we can use for any kind of wave. One particularly useful expression for what follows is that, if we have some orthonormal set of functions $|\psi_j\rangle$ that is complete for describing any function in a given space $H$, then we can write the identity matrix or operator for that space $\mathsf{I}_H$ in Dirac notation as

$$\mathsf{I}_H = \sum_j |\psi_j\rangle\langle\psi_j|. \tag{1}$$

### B. Formulation of the Problem

If for the moment we presume that there is a specific net source $|\psi_{Sm}\rangle$ in the scattering space, then that source would lead to a wave $|\phi_{SCm}\rangle$ within the scattering space through some linear operator (a Green's function) $\mathsf{G}_S$, i.e.,

$$|\phi_{SCm}\rangle = \mathsf{G}_S|\psi_{Sm}\rangle. \tag{2}$$

Note that in our approach $|\psi_{Sm}\rangle$ represents *all* sources in the volume; it includes any source that we might regard as being generated as a result of the interaction between, e.g., waves and dielectrics, such as induced polarizations or currents. Hence $\mathsf{G}_S$ is the "free-space" Green's function. In what follows, we presume, indeed, that there are no sources other than those that are a result of the interaction of the wave with the scatterer; we have no other sources in the scattering volume.

We presume that there is some incident wave $|\phi_{Im}\rangle$ that caused all of these sources through its scattering. The total wave $|\phi_{Sm}\rangle$ in the space must be the sum of the incident and scattered waves, i.e.,

$$|\phi_{Sm}\rangle = |\phi_{Im}\rangle + |\phi_{SCm}\rangle. \tag{3}$$

Any wave $|\phi\rangle$ interacting with the scatterer will in turn give rise to sources $|\psi\rangle$ through some other linear operator $\mathsf{C}$, i.e.,

$$|\psi\rangle = \mathsf{C}|\phi\rangle. \tag{4}$$

For example, $\mathsf{C}$ could essentially just represent the dielectric constant of the material, with waves giving rise to effective sources because of their interaction with the dielectric. Such an interaction will often be local, that is, a wave at a given point gives rise only to a source at the same point, though the general mathematics here is not restricted to such a case.

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

Hence

$$\begin{aligned}|\psi_{Sm}\rangle &= \mathsf{C}|\phi_{Sm}\rangle = \mathsf{C}|\phi_{Im}\rangle + \mathsf{C}|\phi_{SCm}\rangle \\ &= \mathsf{C}|\phi_{Im}\rangle + \mathsf{C}\mathsf{G}_S|\psi_{Sm}\rangle = \mathsf{C}|\phi_{Im}\rangle + \mathsf{A}_S|\psi_{Sm}\rangle,\end{aligned} \tag{5}$$

where

$$\mathsf{A}_S = \mathsf{C}\mathsf{G}_S. \tag{6}$$

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

$$|\phi_{Rm}\rangle = \mathsf{G}_{SR}|\psi_{Sm}\rangle. \tag{7}$$

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 be mathematically orthogonal; such orthogonality gives us a clear definition of what we mean by "separate." We will therefore be trying to deduce some limit on the number of such orthogonal functions $|\phi_{Rm}\rangle$ that can be generated in the receiving space.

In a broad range of problems in electromagnetism and elsewhere (including all those for which operator $\mathsf{G}_{SR}$ can be written as a Hilbert–Schmidt kernel [39]), the operator $\mathsf{G}_{SR}$ is compact. (In practice, compact operators are essentially those that can be approximated to an arbitrary accuracy by a sufficiently large matrix.) Because it is compact, we can perform a singular value decomposition [40] of $\mathsf{G}_{SR}$, which yields an orthogonal set of source functions $|\psi_{Sm}\rangle$, associated singular values $s_m$, and a corresponding orthogonal set of waves in the receiving space $|\phi_{Rm}\rangle$. These pairs of functions are the optimum "communications modes" for coupling between these two spaces, with coupling strengths given by the singular values, and they lead to results that are generalizations of diffraction theory for volumes [36,37].

For our actual problem, we are interested only in sources within the scattering space that will generate nonzero waves in the receiving space, and we therefore restrict further interest only to that subset of the $|\psi_{SM}\rangle$ with nonzero singular values $s_m$. Note, incidentally, that the results of this singular value decomposition are unique (except for the usual arbitrariness of the eigenfunctions of degenerate eigenvalues, which here becomes the arbitrary choice of orthogonal linear combinations of functions corresponding to identical singular values). There is essentially only one set of orthogonal functions $|\psi_{Sm}\rangle$ in the scattering space that gives rise to essentially only one set of corresponding orthogonal functions $|\phi_{Rm}\rangle$ in the receiving space.

We will separate the counting of orthogonal waves into two parts. Specifically, we will presume that we can 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 a supposed unperturbed propagation of the incident wave straight through the scattering volume. It could be, anyway, that the receiving space is chosen in such a way that such straight-through waves completely miss the receiving space, as in the case of a receiving

space set up to see only reflected or backscattered waves, though such a restriction is not necessary for this argument. 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 $\mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle$. The single-scattered waves may not always be as easy to understand and dismiss as the straight-through waves, and the arguments for dismissing or dealing with them will vary depending on the specific situation being considered. Nevertheless, 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. 1. This will mean essentially that we are interested for the moment only in 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. (7) gives the total wave in the receiving space, and so that wave becomes

$$|\phi_{Rm}\rangle = \mathsf{G}_{SR}|\psi_{Sm}\rangle = \mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle + \mathsf{G}_{SR}\mathsf{A}_S|\psi_{Sm}\rangle. \quad (8)$$

Because we presume we are interested only 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

$$\langle\phi_{Rm}|\mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle = 0. \quad (9)$$

Hence, from Eqs. (8) and (9)

$$\langle\phi_{Rm}|\phi_{Rm}\rangle = \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle = 0 + \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}\mathsf{A}_S|\psi_{Sm}\rangle. \quad (10)$$

### C. Proof of Core Sum Rule

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 as in Eq. (1), i.e., $\mathsf{I}_{HS} = \sum_j |\psi_{Sj}\rangle\langle\psi_{Sj}|$, to obtain from Eq. (10)

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle = \sum_j \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sj}\rangle\langle\psi_{Sj}|\mathsf{A}_S|\psi_{Sm}\rangle = \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle, \quad (11)$$

where in the last step we have used the fact that

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sj}\rangle = \langle\phi_{Rm}|\phi_{Rj}\rangle = 0, \text{ unless } m = j, \quad (12)$$

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. (11) that, for each $m$ for which $|\psi_{Sm}\rangle$ gives rise to a nonzero wave in the receiving space, i.e., for which

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle \neq 0, \quad (13)$$

then

$$\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle = 1, \quad (14)$$

and hence, trivially,

$$|\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle|^2 = 1. \quad (15)$$

If we could separately establish a result of the form

$$\sum_i |\langle\psi_{Si}|\mathsf{A}_S|\psi_{Si}\rangle|^2 \leq S_A \quad (16)$$

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

$$M \leq S_A, \quad (17)$$

i.e., we would have established a maximum dimensionality of the space of functions that can be generated in the receiving space by (strong) scattering.

The requirement of strict orthogonality between the single-scattered wave and the multiple-scattered waves can be relaxed somewhat, as discussed in Appendix A. As long as the overlap between the single-scattered wave and the multiple-scattered waves is bounded, a sum rule can still be obtained, and so some small overlap between the single-scattered wave and the multiple-scattered waves does not invalidate the essence of the result, Eq. (17).

### D. Interpretation of Bound

The number $M$ is the maximum number of members of the orthogonal set $|\phi_{Rm}\rangle$ of waves in the receiving volume that can be created as a result of multiple scattering, where such waves are orthogonal to the straight-through wave and also orthogonal (or approximately so—see Appendix A) to the single-scattered wave. The limit does not tell us which of these members are accessible and which are not, i.e., it does not tell us what waves we can and cannot generate in this way. But it does tell us for any given design, with the constraints of orthogonality to straight-through and single-scattered waves, that there is an upper bound for the number of degrees of freedom we can access in the wave we can generate in the receiving volume regardless of what initial wave we use for scattering. For any given operator $\mathsf{A}$, and hence for any given design of scatterer, there is a specific maximum number of elements in the orthogonal set of waves $|\phi_{Rm}\rangle$ that can possibly be accessed in this way. That subset of elements is a property of the operator $\mathsf{A}$ (and hence of the design), and not of the incident fields. The subset is the set $|\phi_{Rm}\rangle$ for which the corresponding $|\psi_{Sm}\rangle$ satisfies Eq. (15) (in the case of exact orthogonality to the single-scattered wave). Remember that $|\phi_{Rm}\rangle$ and $|\psi_{Sm}\rangle$ are uniquely defined (within the degeneracy associated with degenerate singular values) by the singular value decomposition of the operator $\mathsf{G}_{SR}$ that couples sources in $H_S$ to resulting waves in $H_R$.

Suppose then we are interested in designing a device to perform some function, and suppose, for simplicity, that we have already concluded that the single-scattered and straight-through waves are not of interest or do not fall within the receiving space being considered. Then, for any possible choices of input waves, $M$ is the maximum number of degrees of freedom we have in designing the form of the waves generated in the receiving volume. It is the maximum number of coefficients we can choose for lin-

early independent quantities in the generated wave amplitudes in the receiving volume. We will return below to consider such arguments explicitly in example one-dimensional cases. There we will also consider the straight-through and single-scattered waves.

### E. Evaluation of Bound

For scattering operators in scalar waves or electromagnetism, at least if we restrict our interest to finite frequency ranges, times, volumes, and dielectric constants (or other equivalent quantities), we will generally find, as we discuss below, that there is indeed such a bound $S_A$ that can be evaluated. Furthermore, we find we can evaluate this limit based on two separate parts. One of those parts requires knowledge only of the shape and extent of the spaces; the other part requires knowledge only of the scatterer material, and in particular allows the calculation of an upper bound based only on the limits on the variation of the scattering object (e.g., its dielectric constant) within the structure. Importantly, we do not need to know the actual form of the scattering object (other than knowing some volume that bounds it), nor solve the actual problem of scattering from this specific object, in order to calculate a limit.

Now we proceed to evaluate the limit of interest. By definition, we have, from Eq. (6),

$$\langle \psi_{Sm} | \mathsf{A}_S | \psi_{Sm} \rangle = \langle \psi_{Sm} | \mathsf{C}\mathsf{G}_S | \psi_{Sm} \rangle. \tag{18}$$

Now, we can in general write

$$\mathsf{C}^\dagger | \psi_{Sm} \rangle = c_{Sm}^* | \phi_{Cm} \rangle \tag{19}$$

for some normalized $|\phi_{Cm}\rangle$ and some complex number $c_{Sm}$, or equivalently

$$\langle \psi_{Sm} | \mathsf{C} = c_{Sm} \langle \phi_{Cm} |, \tag{20}$$

similarly,

$$\mathsf{G}_S | \psi_{Sm} \rangle = g_{Sm} | \phi_{Gm} \rangle \tag{21}$$

for some normalized $|\phi_{Gm}\rangle$ and some complex number $g_{Sm}$. Eqs. (19)–(21) are just statements that the result of operating on a normalized function with an operator is another function with generally some different magnitude.

Hence

$$\langle \psi_{Sm} | \mathsf{A}_S | \psi_{Sm} \rangle = c_{Sm} g_{Sm} \langle \phi_{Cm} | \phi_{Gm} \rangle, \tag{22}$$

and consequently

$$|\langle \psi_{Sm} | \mathsf{A}_S | \psi_{Sm} \rangle|^2 = |c_{Sm}|^2 |g_{Sm}|^2 |\langle \phi_{Cm} | \phi_{Gm} \rangle|^2. \tag{23}$$

Now, because $|\langle \phi_{Cm} | \phi_{Gm} \rangle|^2$ is simply the magnitude squared of the overlap between two normalized functions, in general $|\langle \phi_{Cm} | \phi_{Gm} \rangle|^2 \le 1$. Hence—also using Eq. (15)—for all of the $|\psi_{Sm}\rangle$ meeting our criterion of generating nonzero waves in the receiving volume [Eq. (13)],

$$|c_{Sm}|^2 |g_{Sm}|^2 \ge |\langle \psi_{Sm} | \mathsf{A}_S | \psi_{Sm} \rangle|^2 = 1. \tag{24}$$

Now we note, from the definition Eq. (19), and from the property that a real quantity is equal to its own Hermitian adjoint,

$$|c_{Sm}|^2 = \langle \psi_{Sm} | \mathsf{C}\mathsf{C}^\dagger | \psi_{Sm} \rangle = \langle \psi_{Sm} | \mathsf{C}^\dagger \mathsf{C} | \psi_{Sm} \rangle. \tag{25}$$

We also have

$$|g_{Sm}|^2 = \langle \psi_{Sm} | \mathsf{G}_S^\dagger \mathsf{G}_S | \psi_{Sm} \rangle. \tag{26}$$

We can write some specific values for the sums of each of these terms $|c_{Sm}|^2$ and $|g_{Sm}|^2$; we can formally write those sums in terms of the trace operator Tr (which is by definition the sum of the diagonal matrix elements of an operator):

$$\sum_m |c_{Sm}|^2 = N_c \equiv \mathrm{Tr}(\mathsf{C}^\dagger \mathsf{C}) \tag{27}$$

and

$$\sum_m |g_{Sm}|^2 = N_{GS} \equiv \mathrm{Tr}(\mathsf{G}_S^\dagger \mathsf{G}_S), \tag{28}$$

where these sums are over all the $m$ for which the $|\psi_{Sm}\rangle$ meet the required criterion. (Since we will be evaluating upper bounds, in fact these sums can be over a larger set of functions. Such a larger set could only increase the resulting $N_C$ and $N_{GS}$, because we would only be adding further positive quantities to the sum, and hence could lead only to an overestimate of the bound. The resulting final inequality would still be valid. We will discuss this point explicitly for the case of $N_C$ below.)

Let us note at this point that there will be such a maximum of the sum of the $|\langle \psi_{Sm} | \mathsf{A}_S | \psi_{Sm} \rangle|^2$ for reasonable physical problems. If we presume that the source generated by a finite wave is itself always finite, then we can write $|c_{Sm}|^2 \le |c_{S\,\max}|^2$ for some finite number $c_{S\,\max}$. Then we have from Eq. (24)

$$\sum_m |\langle \psi_{Sm} | \mathsf{A}_S | \psi_{Sm} \rangle|^2 \le |c_{S\,\max}|^2 \sum_m |g_{Sm}|^2 = |c_{S\,\max}|^2 N_{GS}. \tag{29}$$

Hence as long as the Green's function is square integrable over the scattering space [i.e., the trace $N_{GS} = \mathrm{Tr}(\mathsf{G}_S^\dagger \mathsf{G}_S)$ is finite] and the response of the medium is finite, there will be a maximum to the sum, and hence there will be some maximum $M$ from Eq. (17).

Now we could attempt to find a maximum as in Eq. (17), but we can be more specific than that, and ask instead directly for the maximum number $M$ of orthogonal $|\psi_{Sm}\rangle$ that obey the condition of Eq. (24). Now that we know there is a specific maximum number $M$ of such possible functions, to make this number as large as possible, we presume that all of the available resource for the $|c_{Sm}|^2$, i.e., the $N_C$ in Eq. (27), and similarly all of the available resource for the $|g_{Sm}|^2$, i.e., the $N_{GS}$ in Eq. (28), are made available for only these $M$ possibilities, i.e., we now restrict the sums in Eqs. (27) and (28) to be from 1 to $M$. Given that, we know therefore that the average values, $\overline{|c_{Sm}|^2}$ and $\overline{|g_{Sm}|^2}$, respectively, of $|c_{Sm}|^2$ and $|g_{Sm}|^2$ are $\overline{|c_{Sm}|^2} = N_C/M$ and $\overline{|g_{Sm}|^2} = N_{GS}/M$, respectively. Hence the product of these average values is given by

$$P = \overline{|c_{Sm}|^2 |g_{Sm}|^2} = \frac{N_C N_{GS}}{M^2}. \tag{30}$$

If this average value $P$ falls below 1, then we know there is at least one product $|c_{Sm}|^2|g_{Sm}|^2<1$ since there is no other way of obtaining an average of positive quantities that is less than 1. But we know from Eq. (24) that, for all $m$ that meet our criteria, $|c_{Sm}|^2|g_{Sm}|^2 \geq 1$. Hence we must have $P \geq 1$, and so $N_C N_{GS}/M^2 \geq 1$, and so, finally, we have

$$M \leq \sqrt{N_C N_{GS}}. \tag{31}$$

Hence we have the following theorem.

Suppose we have

1. a scattering space $H_S$ and a receiving space $H_R$;
2. a compact, linear, "free-space" Green's function operator $\mathsf{G}_{SR}$ that generates waves in $H_R$ in response to sources in $H_S$;
3. a linear, square-integrable, "free-space" Green's function operator $\mathsf{G}_S$ that generates waves in $H_S$ in response to sources in $H_S$;
4. a bounded linear operator $\mathsf{C}$ that generates sources in $H_S$ in response to waves in $H_S$;

and that we therefore calculate

5. the orthonormal set $|\psi_{Sm}\rangle$ of functions in $H_S$ that are the result of the singular value decomposition of the operator $\mathsf{G}_{SR}$, including only functions with nonzero singular values in this set;
6. the traces $N_C = \mathrm{Tr}(\mathsf{C}^\dagger \mathsf{C})$ and $N_{GS} = \mathrm{Tr}(\mathsf{G}_S^\dagger \mathsf{G}_S)$, taken over the set $|\psi_{Sm}\rangle$;

then

the maximum number $M$ of orthogonal waves that can be generated in $H_R$ that are orthogonal to the single-scattered and straight-through waves is $M \leq \sqrt{N_C N_{GS}}$.

Note here that $N_{GS}$ is a property only of $\mathsf{G}_S$, which in turn is determined only by the Green's function of the underlying wave equation and by the geometry of the scattering and receiving spaces. $N_{GS}$ need contain no properties of the scattering material or structure. $N_C$ contains only the properties of the scattering material and structure. This split into these two separate parts is particularly important in devising simple sum rules.

Note also that this result is very general. We have only had to require that the operators $\mathsf{G}_S$ and $\mathsf{C}$ be linear, that $\mathsf{C}$ be bounded, and that $\mathsf{G}_S$ be square integrable over the scattering space and $\mathsf{G}_{SR}$ be compact. Hence the result can apply to a very broad range of linear wave problems, including linear classical vector and scalar waves and quantum-mechanical waves (which are fundamentally always linear by hypothesis).

## 3. ONE-DIMENSIONAL EXAMPLE

As a concrete illustration, we consider a one-dimensional problem. Essentially, we can imagine that we want to use a one-dimensional structure to separate pulses of different center wavelengths to different delays in time. The scatterer could be, for example, a dielectric stack, which therefore has no variation in the lateral directions; the various waves could be plane waves of some kind; and we could be considering a "receiving" volume that similarly has effectively infinite lateral extent, but some finite thickness. Single-mode waveguide scattering problems, such as fiber gratings, could also be written as one-dimensional problems by suitable transformation of the problem.

### A. Choice of Spaces

For simplicity, we restrict the problem. First, we presume we are interested only in a relatively small frequency bandwidth $\Delta\omega$ around some (positive) center frequency $\omega_c$, so $\Delta\omega \ll \omega_c$, or equivalently, in terms of wave vectors, a wave vector bandwidth of $\Delta k = \Delta\omega/\nu_o$ around a center wave vector $k_c = \omega_c/\nu_o$, where $\nu_o$ is the wave velocity in the background medium, e.g., the medium outside the structure.

Second, we choose the scattering space to have length $\Delta z_S$, much longer than the wavelength in the background material, centered around $z=0$, and not overlapping with the receiving volume (see Fig. 2).

Third, we will presume an arbitrarily long receiving volume, formally of length $\Delta z_R$ and centered around some position $z_{RC}$ (see Fig. 2). For definiteness, for the moment we presume that the receiving volume is at larger $z$ (i.e., we are working in transmission), though the final results for the $N_C$ and $N_{GS}$ sums would be the same if we chose it at smaller $z$ (i.e., working in reflection).

Fourth, we presume that we are interested only in some very small time interval $\Delta t_R \ll 1/\omega_c$ in the receiving volume, centered about some time $t_{Ro} = z_{RC}/\nu_o$. The idea here is that we are taking a "snapshot" of the wave in the receiving volume, at a time $t_{Ro}$ when the scattered waves that left the scattering volume at time zero are approxi-

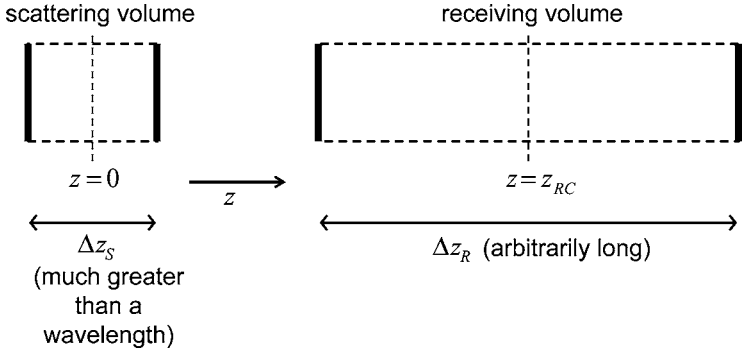


Fig. 2. Illustration of the physical volumes used for the one-dimensional example.

mately in the middle of the receiving volume. We have chosen that receiving volume long enough to capture any possible scattering response of interest. We will be interested in waves that are orthogonal as seen in that snapshot. (We could equally well have chosen here to take a very thin slice for the spatial volume and to look over a very long time at the wave in that thin slice—the results from our analysis would be identical.) We will want to know how many orthogonal waves can be generated in this receiving space as a result of scattering from the scattering space.

For this illustration we will take a simple scalar wave equation, which becomes, for waves of a given frequency $\omega_o$,

$$\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} + k_o^2\phi = -k_o^2\eta(z,\omega_o)\phi, \tag{32}$$

where $k_o = \omega_o/v_o$. Such an equation is valid for electromagnetic waves in isotropic, nonmagnetic materials with no free charge or free currents in such one-dimensional problems. In that electromagnetic case, $\eta$ gives the fractional variation in the relative dielectric constant in the scatterer, i.e.,

$$\eta(z,\omega_o) \equiv \frac{\Delta\varepsilon(z,\omega_o)}{\varepsilon_{ro}}, \tag{33}$$

where $\varepsilon_{ro}$ is the background relative dielectric constant, and the wave velocity in the background medium is $v_o = c/\sqrt{\varepsilon_{ro}}$, where $c$ is the velocity of light.

The detailed evaluation of the sums is given in Appendix B. Note that to evaluate these sums it is not actually necessary to solve the singular value decomposition of the operator $\mathsf{G}_{SR}$. We prove in Appendix C that other, more convenient sets of functions (e.g., Fourier wave basis functions) may be used provided they satisfy a sum rule condition. We obtain the results

$$N_{GS} = \frac{\Delta k \Delta z_R}{\pi}\frac{k_c^2(\Delta z_S)^2}{12} \tag{34}$$

and

$$N_C \le \frac{\Delta k \Delta z_R}{\pi}\eta_{rms}^2 \le \frac{\Delta k \Delta z_R}{\pi}\eta_{\max}^2, \tag{35}$$

where

$$\eta_{rms}^2 = \frac{1}{\Delta z_S \Delta\omega}\int_{\omega=v_o\left(k_c-\frac{\Delta k}{2}\right)}^{v_o\left(k_c+\frac{\Delta k}{2}\right)}\int_{z=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}} |\eta(z,\omega)|^2 \mathrm{d}z\,\mathrm{d}\omega \tag{36}$$

is the frequency and spatial average of the mean (modulus) square variation in $\eta$ (i.e., $\eta_{rms}$ is the root mean square average over frequency and space of the relative variation in dielectric constant compared with the background), and $\eta_{\max}$ is the maximum value of $|\eta|$ at any frequency and position.

Note, incidentally, that this sum rule includes the possibility that $\eta$ can take on completely different spatial forms at different frequencies, or different spectral forms at different positions. Hence, for example, it is not restricted to structures such as dielectric stacks, which will have the same form of spatial distribution at all frequencies, nor to spatial patterns of one kind of atom. The sum does include the possibility therefore of two different spatial patterns of two (or more) different kinds of atoms or other interlaced materials.

### B. Limit for *M*

Hence, finally, putting together Eqs. (34) and (35), we have

$$M \le \sqrt{N_C N_{GS}} \le n_{tot} N_o, \tag{37}$$

where

$$N_o = \frac{\pi}{\sqrt{3}}N_{S\lambda}\eta_{rms} \le \frac{\pi}{\sqrt{3}}N_{S\lambda}\eta_{\max}, \tag{38}$$

$$n_{tot} = \frac{\Delta k \Delta z_R}{\pi}, \tag{39}$$

$$N_{S\lambda} = \frac{k_c \Delta z_S}{2\pi}. \tag{40}$$

These terms $n_{tot}$ and $N_{S\lambda}$ have the following physical meanings. $n_{tot}$ is the number of degrees of freedom required to define a function of (spatial) bandwidth $\Delta k$ in a distance $\Delta z_R$ (or equivalently a function of angular frequency bandwidth $\Delta\omega$ over a time $\Delta t = v_o \Delta z_R$) according, for example, to the sampling theorem; this is the maximum number of degrees of freedom for any wave in the receiving volume given its length and the frequency bandwidth chosen. $N_{S\lambda}$ is the length of the scattering volume measured in wavelengths $\lambda_c = 2\pi/k_c$ at the center frequency in the background medium. Note, incidentally, that if we know $\eta_{\max}$ we can write particularly general and simple upper-bound limits without other assumptions about the spatial and frequency variation of the scattering material.

We have calculated this here for forward-scattered waves. For backscattered ("reflected") waves, we would instead have chosen the receiving volume to be at smaller $z$, and would have chosen backwards-propagating Fourier source basis functions. The derivation would otherwise proceed identically, and the result for such a backscattered case would be identical to Eq. (37) and the equations following.

### C. Final Limit

To obtain the final bound for a useful limit, we need to consider adding additional possible orthogonal functions that are associated with the straight-through and/or single-scattered waves. In two or three dimensions, we could have a situation where we reasonably believed both of these mathematical waves would miss the receiving volume because they were going in different directions. In one dimension, however, in transmission, both the straight-through and single-scattered waves are likely to pass through the receiving volume at some time (as shown in Fig. 2), and in reflection at least the single-scattered wave could pass through the receiving volume at some time. Since we have chosen the receiving volume to be very long, the single-scattered wave (and, in trans-

mission, also the straight-through wave) will fall somewhere within the volume at the observation time $t_{Ro}$. Hence we cannot neglect these mathematical waves in constructing our final limit.

No matter what incident waves we are considering, the maximum number of degrees of freedom we could possibly need to describe any possible single-scattered waves in the receiving volume is simply $n_{tot}$. After all, that number is sufficient to describe any possible wave in our frequency band of interest in the receiving volume. Similarly, in the case where the receiving volume is on the right, we can allocate a further number $n_{tot}$ of degrees of freedom to describe the straight-through waves. Hence we can write, after inclusion of the straight-through and single-scattered waves, for the reflection case, a maximum number of degrees of freedom available for describing waves that could physically exist in the receiving volume,

$$M_{reftot} \leq n_{tot}(1+N_o), \tag{41}$$

and for the transmission case,

$$M_{transtot} \leq n_{tot}(2+N_o). \tag{42}$$

### D. Interpretation of Limit for Separating Pulses

Suppose that we want a scattering device that can separate pulses in time based on their center wavelength. Hence we might imagine that these various possible input pulses would all be centered at the same time, and even have the same shape; all that would distinguish them is their center wavelength or frequency, as sketched in Fig. 3.

To use the limit to analyze this problem, we need three steps. First, given our restricted set of inputs, namely, those corresponding to pulses of different wavelengths but all at the same time, we will need to understand how to scale the limit results we have. Second, we will need to understand how many orthogonal functions (or degrees of freedom) we need to be accessible in the receiving volume if we are to separate the pulses as we desire. Third, we put these two together to deduce the final practical limit.

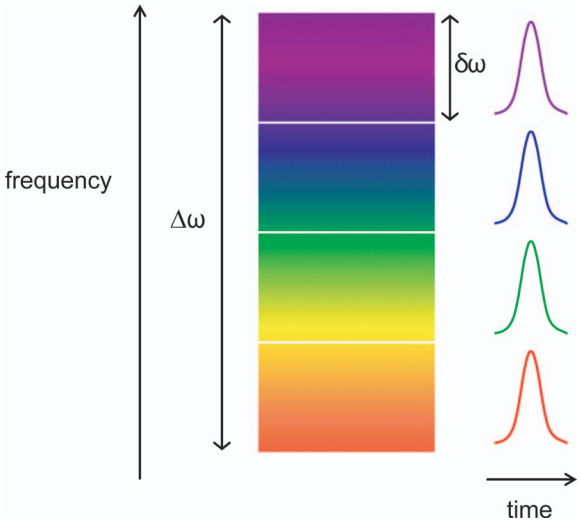


Fig. 3. Sketch of four input pulses of identical envelope shapes in time but different center frequencies, with their corresponding spectral ranges.

#### 1. Scaling the Number of Degrees of Freedom

*Informal argument*. We presume that these pulses of interest have some characteristic time duration $\Delta t_p$, or, equivalently, some characteristic length $\Delta z_p = v_o \Delta t_p$. These various pulses will each generate some corresponding scattered pulse in the receiving volume.

However, another similar set of input pulses at some time $\Delta t_p$ later will generate an exactly similar set of corresponding pulses in the receiving volume, but delayed by an amount $\Delta t_p$, or equivalently, displaced by a distance $\Delta z_p$ (see Fig. 4). This replica set of scatterings is, as far as our problem is concerned, essentially identical to the set we have already considered, and we do not want to count them again. In fact, with our very long receiving volume, we are essentially counting the same set of scatterings $\Delta z_R/\Delta z_p$ times. Hence, to avoid such multiple counting, we should derate from $n_{tot}$ by a factor $\Delta z_R/\Delta z_p$ in the expressions (41) and (42).

*Rigorous argument*. We can take a more rigorous approach to this argument. Suppose now that we say we are interested in a set of frequency bands, each of bandwidth $\delta\omega \equiv v_o \delta k$, that together make up the entire bandwidth $\Delta\omega = v_o \Delta k$ (Fig. 3). There will be altogether

$$N_b = \frac{\Delta\omega}{\delta\omega} = \frac{\Delta k}{\delta k} \tag{43}$$

such frequency bands. Each such band has, from the sampling theorem,

$$n_b = \frac{\delta k \Delta z_R}{\pi} \tag{44}$$

possible orthogonal functions required to specify all possible fields in this bandwidth in the receiving volume. As mentioned above, we pick an identical pulse shape for each band, but with different center frequency. Hence, because we are restricting the dimensionality of the input space by a factor $1/n_b$ (we are choosing only one out of each $n_b$ input functions), we might expect that we could derate the dimensionality of the output space by the same factor. This heuristic statement requires some deeper discussion.

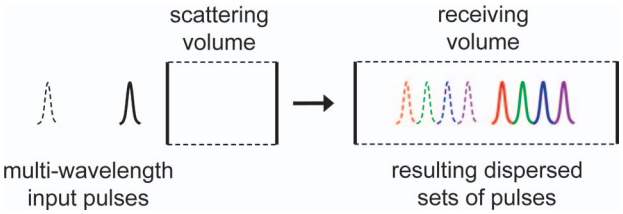


Fig. 4. Sketch of dispersion of two different pulses. An input pulse (solid black curve) consisting of a superposition of four pulses of the same shape but different center wavelengths is dispersed by the scatterer to give the resulting set of pulses (solid colored curves) in the receiving volume. A similar but delayed input pulse (dashed black curve) would, however, similarly be dispersed to give a delayed set of dispersed pulses (dashed colored curves). The simple limit formulas [Eqs. (41) and (42)] count both of these as distinct results, an overcounting that must be corrected for a useful limit.

As far as counting the number of straight-through or single-scattered modes is concerned, this statement is straightforward and obvious; if we restrict the input space to some smaller number of functions, then we similarly restrict the number of possible straight-through or single-scattered wave functions we need to consider in the receiving volume.

For the $N_{GS}$ sum, because we are restricting the number of input functions by this factor, the corresponding number of source functions that can be of interest has been reduced by the same factor. Each orthogonal source basis function is associated with one linearly independent input function, and so the dimensionalities of these spaces remain the same as each other and are hence reduced by the same factor.

In the $N_{GS}$ sum in Appendix B, the contributions to the sum from each of the different Fourier basis functions are all equal ($\gamma_{cn}=\langle\psi_{Acn}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{Acn}\rangle=k_c^2(\Delta z_S)^2/12$ [Eq. (B37)] for the cosine basis functions, and similarly for the sine basis functions $|\psi_{Asn}\rangle$). We can also show (Appendix D) that, for any other complete orthonormal basis we might choose that spans the same space $H_S$, the contributions of each of those functions to the $N_{GS}$ sum would also be the same number as for each of the Fourier basis functions. Hence we can scale the $N_{GS}$ sum as we reduce the dimensionality of the space we are considering, regardless of the basis set we want to use for that space.

If we are using the form $N_C \leq (\Delta k \Delta z_R/\pi)\eta_{\max}^2$ in Eq. (35) for the $N_C$ sum, then this bound scales directly with no further assumptions, because this upper bound implicitly assumes equal contributions from every frequency and position in the scatterer. This form also allows us to assume nothing in particular about the frequency spectrum of the pulses we are using (other than that they fit within our bands of width $\delta\omega$). For the form $N_C \leq (\Delta k \Delta z_R/\pi)\eta_{rms}^2$, whether or not we can simply scale down the $N_C$ sum depends on whether our new restricted set effectively averages $\eta$ over position and frequency in the same way as the original set. At least: (i) if the scattering space has a similar range of variation of $\eta$ over the entire scattering volume; and (ii) if there are no resonances in the dielectric response of the materials that are sharp compared to our frequency band of interest, then we can simply scale down the $N_C$ sum. If we presume here, for example, that we are dealing with a dielectric stack made from materials that are themselves relatively nondispersive, and that the materials are relatively uniformly distributed over the volume, we should have no problems with this scaling of the $N_C$ sum. [In any case, for simplicity below in our calculated example, we take the simpler $N_C \leq (\Delta k \Delta z_R/\pi)\eta_{\max}^2$ form.]

Presuming that we are obeying conditions that allow us to derate both $N_{GS}$ and $N_C$ by $1/n_b$, then the quantity $\sqrt{N_{GS}N_C}$ will also derate by $1/n_b$. Since, from Eqs. (39), (43), and (44), $n_{tot}/n_b=N_b$, on this argument we would obtain, for the reflection case,

$$M_{reftot} \leq N_b(1+N_o) \tag{45}$$

and for the transmission case

$$M_{transtot} \leq N_b(2+N_o) \tag{46}$$

instead of Eqs. (41) and (42), where we remember that $N_b$ is the number of frequency bands we are considering for our pulses.

*2. Representation of Pulses*

Having thus eliminated multiple counting, we return to the problem of separating pulses of different center frequencies, to understand how many orthogonal functions or degrees of freedom we need to be able to control in the receiving space in order to do so.

Presume for the moment that we wish to work with pulses that are approximately the minimum duration in time for the given bandwidths $\delta\omega$ for each of the frequency bands. Such pulses are formed in a Fourier basis essentially by taking equal amounts of each Fourier basis function with the same phase or time delay. (More rigorously, in the sine and cosine basis of Appendix B, we take equal amounts $a$ of each sine function and equal amounts $b$ of each cosine function, though $a$ and $b$ may be different numbers, with the ratio of $a$ and $b$ determining the phase of the resulting pulse within a given half-cycle of the carrier.) Different precise forms of pulse could have slightly different coefficients, but any pulse of approximately this minimum duration will essentially involve similar amplitudes from all the basis functions within the bandwidth (subject to the discussion of $a$ and $b$ above).

Pulses formed with exactly equal amounts of the Fourier basis functions (subject to the $a$ and $b$ coefficient distinction) will form an exactly orthogonal set, with pulse centers spaced by integer multiples of a time $\Delta t_p=2\pi/\delta\omega$. There are two such sets, one corresponding to the combinations of the sine basis functions only ($a$ finite, $b=0$), and the other with cosine basis functions only ($b$ finite, $a=0$), or equivalently corresponding to the two different orthogonal phases of the carrier. The two such sets taken together form a complete basis for the receiving space of interest, just as does the Fourier set in Appendix B (the functions in this new set are all orthogonal, and there is the same total number of them as there is of the Fourier functions from which they are formed). Such a pulse basis does have a small "ripple" at times outside the main peak, a ripple that progressively disappears as we consider larger numbers of Fourier functions, but it does give us a rigorously orthogonal pulselike basis.

*3. Number of Required Basis Functions*

Consider the first pulse, say the one from the first frequency band, and for the moment in this argument, let us neglect the phase of the carrier. To represent this first pulse in the receiving volume, obviously we need a finite amplitude in one of the pulse basis functions. But, if we are trying to separate $N_b$ such pulses of different center frequencies to different times, we need also to specify that this first pulse is not present at any of the other $N_b-1$ times at which the other pulses are to appear in the receiving volume. Hence we need to set the coefficients of those other $N_b-1$ basis functions to zero. Hence, overall, we need to control the amplitudes of $N_b$ orthogonal functions in the receiving volume to position this first pulse in the receiving volume. (Equivalently, we need to set the

generally nonzero coefficients of $N_b$ Fourier functions to define this pulse at this time, or we need to set the $N_b$ coefficients of the singular value decomposition basis, which in this case would be prolate spheroidal functions [37]). This is sketched in Fig. 5. The pulse in the second frequency band already has a different center frequency, and so all the pulse basis functions associated with it are already orthogonal to those in the first band. To position the pulse at some other place in the receiving volume, we need therefore another $N_b$ orthogonal functions in the receiving volume. Hence to position all $N_b$ pulses at distinct positions in the receiving volume, we need $N_b^2$ available orthogonal functions in the receiving volume.

Now let us return to consider carrier phase (i.e., the phase of the underlying oscillation at the center frequency $\omega_c$. The simplest case to consider is that in which we explicitly care about the carrier phase of each pulse. Then, instead of requiring $N_b$ orthogonal functions or coefficients to describe one pulse, we require $2N_b$ functions. Hence we would require altogether $N_{tot}=2N_b^2$ available orthogonal functions in the receiving volume instead of the $N_b^2$ deduced above. Just to separate pulses out from one another in time, we might well not care about the phase of each specific pulse, and hence we could let that one degree of freedom "float". However, we would still need to specify that the functions corresponding to both carrier phases were of zero amplitude for all of the other pulse positions, and so we would require control over $2N_b-1$ amplitudes for each pulse, for a total of

$$N_{tot}=N_b(2N_b-1) \tag{47}$$

degrees of freedom to be controlled. We could also argue that the absolute timing of the first pulse is not important if we require only that all the pulses be at different times, so we could subtract 1 from this result. For what follows, however, we will take the $N_b(2N_b-1)$ result for simplicity.

We had presumed above that, for the purposes of argument, we were working with pulses of essentially minimum length for the available bandwidth. If our pulses were longer than the minimum but still of the same bandwidth, we would need more available "time slots" from our basis set to be able to separate them, and so we would need more degrees of freedom. Therefore, our counting here is establishing a minimum number of required available basis functions to separate a number of pulses of given bandwidth.

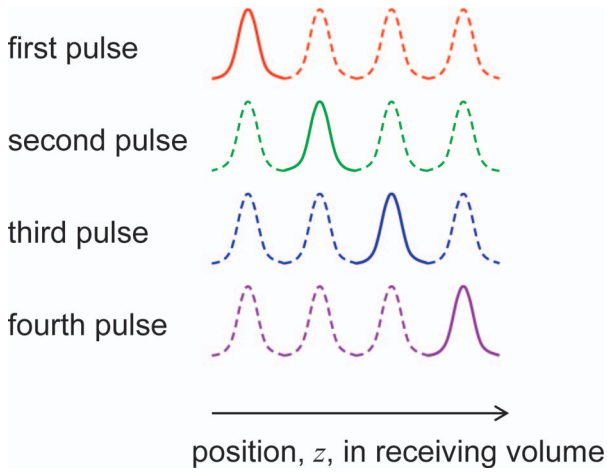


Fig. 5. Sketch of number of orthogonal basis functions (or degrees of freedom) required for separating pulses of different center frequencies in the receiving volume, shown for separation of $N_b=4$ pulses of different center frequencies. Pulse basis functions sketched in solid parts of curves are set to finite amplitudes, while those in dashed parts are set to zero amplitude, hence requiring $N_b^2=16$ basis functions altogether (neglecting carrier phase).

### E. Limit to Separating Pulses

Now we know both the number of required accessible orthogonal modes in the receiving volume [Eq. (47)] and the number of available orthogonal modes in the receiving volume [Eq. (45) for the reflection case or Eq. (46) for transmission]. The number of available modes must equal or exceed the number of required modes. Hence, taking the transmission case as an explicit example, we must have $N_b(2+N_o)\geq N_b(2N_b-1)$; so, for the transmission case

$$N_b \leq \frac{3}{2}+\frac{N_o}{2}=\frac{3}{2}+\frac{\pi}{2\sqrt{3}}N_{S\lambda}\eta_{\max}, \tag{48}$$

or, for the reflection case with similar algebra,

$$N_b \leq 1+\frac{N_o}{2}=1+\frac{\pi}{2\sqrt{3}}N_{S\lambda}\eta_{\max}. \tag{49}$$

*Explicit calculation example*. Suppose then that we wish to separate pulses of 32 different equally spaced frequencies, all within a relatively narrow band (e.g., the telecommunications C-band) about a free-space wavelength of $\lambda=1.55\,\mu$m, and that we will do this using a one-dimensional structure (a dielectric stack) consisting of glass and air. Taking glass to have a dielectric constant of 2.25 (as corresponds to a refractive index of 1.5), and presuming the background dielectric medium is air, the maximum possible value of $\eta_{\max}=\sqrt{(2.25-1)^2/1}=1.25$. Hence, from Eq. (48) [which gives the more favorable result of the pair of Eqs. (48) and (49)], with $N_b=32$ and $k_c=2\pi/\lambda$, the length of the scattering volume (i.e., of the dielectric stack) must be at least 41.7 $\mu$m.

Note that this result:

1. is independent of how well we design this structure or what method we use;
2. is independent of the actual frequency bandwidth around the carrier, as long as that bandwidth is relatively narrow compared with the carrier frequency; and
3. has required knowledge only of the maximum dielectric constant in the material.

And note incidentally that the above result for this structure, in which the relative variation in the dielectric constant is of order unity, corresponds to of the order of one wavelength of structure thickness for each frequency of pulse to be separated in this way $[\pi/(2\sqrt{3})\simeq 0.907]$.

We could go on here to somewhat more sophisticated limits were we to make some more assumptions about the structure. We could for example presume approximately equal amounts of two materials in the structure, which would give an $\eta_{rms}<\eta_{\max}$, or we could believe we were designing only structures whose pulse-splitting ability was

not fundamentally dependent on the background dielectric constant outside the scatterer, in which case we could evaluate a limit based on the "worst" possible choice of background dielectric constant compared with the dielectric constants in the materials of the layered structure. Such different assumptions would generally lead to longer minimum sizes for the structure. The point in this calculation is not so much the specific answers we get from this problem, however, but to demonstrate that we can write down a bound based on such limited and general information. Nonetheless, the answers are physically reasonable numbers; any such high-performance dispersive elements would need to have significant thickness, and for such a one-dimensional approach, no amount of design would allow structures shorter than those calculated using this limit.

## 4. CONCLUSIONS

We have proved here an upper-bound limit to the performance of optical components. The core of this limit is a very general bound to linear multiple-scattering processes [Eq. (31)]. We have shown that there is a fundamental limit to the number of orthogonal waves that can be generated by multiple scattering from any linear scatterer. To write down such a limit, we need to know only (1) the outside shape of the scatterer and the volume into which we wish to scatter, (2) some bounding properties of the material used for the scatterer, and (3) the frequency and time ranges of interest. We have also been able to separate the evaluation of the limit into two distinct parts [in Eq. (31)], one involving only the shape of the volumes, the other involving only the properties of the scattering material. The general results are true for multiple scattering of any linear wave, including scalar waves (such as acoustics and simple quantum mechanics) and vector waves (such as in electromagnetics or more complicated quantum-mechanical systems), with only very broad requirements on the mathematics of those waves, and for arbitrary complex dielectric constants, including absorption and gain.

We have illustrated the limit with an example for one-dimensional structures, as found, for example, in multilayer dielectric stacks, and discussed its impact on optical components for applications in pulse dispersion. The resulting calculated limit requires knowledge only of the limits to the dielectric constant. Though not calculated explicitly here, the approach could also be extended to apply to atomic vapor systems and any other system with highly dispersive material behavior.

Note that these results prove that, if one wants very small dispersive devices, then one has to have large dielectric constants. The limit to the size of one-dimensional devices would scale down approximately in proportion as the available dielectric constant increased. This therefore favors, for example, metallic structures for small components, hence giving another basic reason for exploring nanometallic and plasmonic structures. While that conclusion might seem intuitively obvious, here we have proved it explicitly and rigorously.

Further work here could include an application of these results to calculate a limit for slow-light propagation, explicitly for the number of pulse widths by which a pulse could be delayed for a specific one-dimensional structure. The approach should also be applicable to situations with very strong spectral variation of dielectric response, as in electromagnetically induced transparency. A recasting of the one-dimensional example using the slightly different effective one-dimensional wave equation required for waves with transverse magnetic polarization would allow an explicit comparison with the empirical results and semi-empirical formulas of Gerken and Miller [12] for the related problem of spatial separation of beams of different wavelengths using tilted dielectric stacks.

Finally, the approach here could be extended to two-dimensional and three-dimensional structures so as to obtain explicit limits for the functions that could be performed by photonic nanostructures and other structures with large contrast in their dielectric constants.

## APPENDIX A: RELAXATION OF STRICT ORTHOGONALITY REQUIREMENT

We could regard the requirement of absolute orthogonality between the single-scattered wave $|\phi_{RIm}\rangle=\mathsf{G}_{SR}\mathsf{C}|\phi_{Im}\rangle$ and $|\phi_{Rm}\rangle$ as being somewhat extreme. The single-scattered wave might be weakly distributed over a large volume of space and/or time, making it practically impossible to restrict our interest to those functions that do not overlap with it at all. In practice we might be content to say that $|\phi_{RIm}\rangle$ and $|\phi_{Rm}\rangle$ are approximately orthogonal. In that case, we would have $\langle\phi_{Rm}|\mathsf{G}_{SR}|\psi_{Im}\rangle=f_m$ for some presumably small (generally complex) number $f_m$. Then, instead of Eq. (11), we would have

$$\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle=f_m+\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle; \tag{A1}$$

so, again for all $m$ such that Eq. (13) is obeyed, instead of Eq. (14) we would have

$$1=h_m+\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle, \tag{A2}$$

where $h_m=f_m/\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle$.

Now, using the triangle inequality ($|a+b|\le|a|+|b|$), we can say that $1\le|h_m|+|\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle|$. Also, noting from Eq. (A2) that $\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle=1-h_m$, then, using the triangle inequality again,

$$|\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle|\le 1+|-h_m|=1+|h_m|, \tag{A3}$$

so

$$1-|h_m|\le|\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle|\le 1+|h_m|. \tag{A4}$$

Hence a more practical and realistic criterion is that we consider all $m$ for which

$$|h_m|=\left|\frac{\langle\phi_{Rm}|\mathsf{G}_{SR}|\psi_{Im}\rangle}{\langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle}\right|\le u \tag{A5}$$

for some (real and positive) $u\ll 1$ that we choose, in which case $|\langle\psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle|$ is within $\pm u$ of unity. In other words, we merely ask that the overlap of the single-scattered wave with the multiple-scattered wave basis function in the receiving volume should be much smaller than the squared amplitude of the multiple-scattered wave itself.

Then we would have, from the triangle inequality,

$$(1-u)^2 \le |\langle \psi_{Sm}|\mathsf{A}_S|\psi_{Sm}\rangle|^2 \le (1+u)^2, \tag{A6}$$

and hence, instead of Eq. (17), we would have

$$M \le \frac{S_A}{(1-u)^2}, \tag{A7}$$

which still is a meaningful and useful limit. We can also continue this result through the rest of the analysis in Section 2 if we wish.

## APPENDIX B: DETAILED EVALUATION OF SUMS FOR ONE-DIMENSIONAL CASE

### 1. Time Ranges

We are interested only in sources in the scattering space that can generate waves in the receiving space. Hence, we restrict the time ranges of interest for sources in the scattering space to those that, given the $\nu_o$ in the background medium, give a wave in the receiving volume at time $t_{Ro}$. Hence, the latest emission time we consider for a point $z_S$ in the scattering volume is

$$t_{\max}(z_S) = \frac{1}{\nu_o}\left(\frac{\Delta z_R}{2} + z_S\right), \tag{B1}$$

which is the time at which a wave leaving point $z_S$ would just arrive at the left of the receiving volume (the point $z_{RC}-\Delta z_{R/2}$) at time $t_{Ro}$. Similarly, the earliest time we consider is

$$t_{\min}(z_S) = \frac{1}{\nu_o}\left(-\frac{\Delta z_R}{2} + z_S\right), \tag{B2}$$

by which time the resulting wave would be just about to leave the receiving volume at the right.

### 2. Green's Functions

We rewrite the wave equation Eq. (32) as

$$\frac{1}{k_o^2}\frac{\mathrm{d}^2\phi}{\mathrm{d}z^2} + \phi = -\eta(z,\omega_o)\phi, \tag{B3}$$

which has the associated retarded Green's function [i.e., the wave at $z$ generated by a source amplitude $\delta(z-z_o)\exp(i\omega_o t)$]:

$$\mathsf{G}_{\omega_o}(z,z_o) = (ik_o/2)\exp(-ik_o|z-z_o|). \tag{B4}$$

Using the standard result $\delta(t-t_o)=(1/2\pi)\int_{-\infty}^{\infty}\exp[i\omega_o(t-t_o)]\mathrm{d}\omega_o$, the Green's function corresponding to the source $\delta(z-z_o)\delta(t-t_o)$ in the full time-dependent wave equation would be

$$\mathsf{G}_\delta(z,t;z_o,t_o) \equiv \frac{i\nu_o}{4\pi}\int_{-\infty}^{\infty} k_o \exp\{i[\nu_o(t-t_o)-|z-z_o|]\}\mathrm{d}k_o. \tag{B5}$$

To handle our restriction to a frequency band $\Delta\omega$, we define a Green's function that operates only on functions in this range. Formally including positive and negative frequency bands and using $\Delta k \ll k_c$, we obtain the new Green's function

$$\mathsf{G}(z,t;z_o,t_o) \simeq \frac{-\nu_o k_c}{2\pi}\int_{k_c-\frac{\Delta k}{2}}^{k_c+\frac{\Delta k}{2}} \sin\{k_o[\nu_o(t-t_o)-|z-z_o|]\}\mathrm{d}k_o. \tag{B6}$$

Note that this Green's function operator works for waves seen at any point $z$ in space at any time $t$, either inside or outside the scattering space.

### 3. Choosing the Source Function Set for the Scattering Space

To choose the set of functions that will form the basis set in the scattering space, we first find a convenient way of writing the operator $\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR}$. The eigenfunctions of this operator are the set of functions in the scattering space that are deduced from the singular value decomposition of the operator $\mathsf{G}_{SR}$ and that lead to orthogonal waves in the receiving space. These functions therefore span the space $H_S$.

#### *a. Form of the Operators Mapping from Source to Scattering Spaces*

The operator $\mathsf{G}_{SR}$ that maps from the source function in the scattering space to the resulting wave in the receiving space is of the form of Eq. (B6), but it always has $|z_R-z_S| = z_R-z_S$ because by choice made above the scattering and receiving volumes are physically disjoint, and we chose the receiving volume to be at larger $z$. Using this form, we can evaluate $\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR}$. Changing to centered variables of the form $k=k_o-k_c$, dropping trigonometric terms whose arguments are highly oscillatory—$2k_c z_R$ in the internal integral over the position $z_R$ in the receiving space—and using the fact that $\Delta z_R$ is arbitrarily large, we obtain

$$\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR} = \frac{\nu_o^2 k_c^2 \Delta t_R}{4\pi} \times \int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} \cos\{(k_c+k)[(\nu_o t_1 - z_{S1}) - (\nu_o t_2 - z_{S2})]\}\mathrm{d}k. \tag{B7}$$

In general, in using the operator $\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR}$, we will be integrating over time and position variables in the space $H_S$. Noting the time limits we deduced above in Eqs. (B1) and (B2), those integrals will be of the form

$$\int_{z_S=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}} \int_{t_S=\frac{1}{\nu_o}\left(-\frac{\Delta z_R}{2}+z_S\right)}^{\frac{1}{\nu_o}\left(\frac{\Delta z_R}{2}+z_S\right)} \mathrm{d}t_S \mathrm{d}z_S. \tag{B8}$$

It will be more convenient to change to a time variable of the form $\tau_S = t_S - z_S/\nu_o$. Then the integrations will be of the form

$$\int_{z_S=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}} \int_{\tau_S=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}} \mathrm{d}\tau_S \mathrm{d}z_S, \tag{B9}$$

and we can rewrite Eq. (B7) in terms of these new variables as

$$\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR} = \frac{\nu_o^2 k_c^2 \Delta t_R}{4\pi} \times \int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} \cos[(k_c+k)\nu_o(\tau_{S1}-\tau_{S2})]\mathrm{d}k. \tag{B10}$$

*b. Basis Set*

Exact eigenfunctions of this operator $\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}$ do exist, expressible in terms of prolate spheroidal functions (see, e.g., [37]), but those are not convenient for some manipulations below. Instead, we find another more convenient set of functions that spans the same space. Specifically, we prove that we can use a Fourier basis. Hence we end up with a particular number of Fourier functions that we can use as the basis for the scattering space $H_S$. We can use these basis functions later to evaluate the sums $N_{GS}=\mathrm{Tr}(\mathsf{G}_S^{\dagger}\mathsf{G}_S)$ and $N_C=\mathrm{Tr}(\mathsf{C}^{\dagger}\mathsf{C})$ that we need for the overall sum rule we seek. We first set up the results to prove the validity of our convenient set.

*Condition for completeness of a basis set*. We want to work with some convenient basis set $|\psi_{An}\rangle$ for the space $H_S$. We know already that there is some set $|\psi_{Sm}\rangle$ found from the singular value decomposition of the operator $\mathsf{G}_{SR}$ that spans this space, but these functions may not be convenient to use. We can define a quantity

$$S_{SR}=\sum_m \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle. \tag{B11}$$

Note that we have restricted the set $|\psi_{Sm}\rangle$ to include only functions that give finite waves in the receiving volume, i.e., for which

$$|s_m|^2 \equiv \langle\psi_{Sm}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Sm}\rangle > 0. \tag{B12}$$

Now suppose that we find some other set $|\psi_{An}\rangle$ such that

$$\sum_n \langle\psi_{An}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{An}\rangle = S_{SR}. \tag{B13}$$

We prove in Appendix C that this new set $|\psi_{An}\rangle$ is then also complete for the same space.

*Evaluation of $S_{SR}$*. The sum $S_{SR}$ can be evaluated by using some set of functions we know for certain will be complete for the scattering space, namely the delta function set $|\psi_p\rangle=\delta(\tau_S-\tau_{Sp})\delta(z_S-z_{Sp})$. Any conceivable function of position $z_S$ and time $\tau_S$ can be represented by this basis. Possibly this set also can represent functions that lie outside the space $H_S$, but such other functions cannot by definition make any contribution to the sum $S_{SR}$ because they do not generate any waves in $H_R$, so the result $S_{SR}$ will still be correct even with this over-complete orthogonal set. With such a delta function basis, the sum over the elements of this set reduces to an integral over $\tau_{Sp}$ and over $z_{Sp}$. Formally, we have

$$\langle\psi_p|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_p\rangle = \mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}(z_p,\tau_p;z_p,\tau_p). \tag{B14}$$

Hence

$$\begin{aligned} S_{SR} &= \sum_p \langle\psi_p|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_p\rangle \\ &= \frac{\nu_0^2 k_c^2 \Delta t_R}{4\pi}\int_{z_S=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\int_{\tau_S=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} \mathrm{d}k\,\mathrm{d}\tau_S\mathrm{d}z_S \\ &\quad\times\cos[(k_c+k)\nu_o(\tau_s-\tau_s)] = \frac{\nu_o k_c^2 \Delta t_R \Delta z_S \Delta z_R \Delta k}{4\pi}. \end{aligned} \tag{B15}$$

*Proposal of Fourier basis set*. The convenient set we propose to work with is the Fourier basis for $H_s$ of the form

$$\begin{aligned} &\psi_A(z_S,\tau_S) \\ &= \begin{cases} \psi_{Asn}(z_S,\tau_S) = \sqrt{\dfrac{2\nu_o}{\Delta z_S \Delta z_R}}\sin\left[\left(k_c+\dfrac{2n\pi}{\Delta z_R}\right)\nu_o\tau_S\right] \\ \psi_{Acn}(z_S,\tau_S) = \sqrt{\dfrac{2\nu_o}{\Delta z_S \Delta z_R}}\cos\left[\left(k_c+\dfrac{2n\pi}{\Delta z_R}\right)\nu_o\tau_S\right] \end{cases}, \end{aligned} \tag{B16}$$

where the integer $n$ can take on any of the $\sim\Delta k\Delta z_R/2\pi$ values

$$-\frac{\Delta k\Delta z_R}{4\pi} \le n \le \frac{\Delta k\Delta z_R}{4\pi}. \tag{B17}$$

Note that we have proposed a set in two parts, one with sines and the other with cosines. These functions are all orthogonal to one another, at least formally if we take $\Delta z_R$ equal to an integer number of half-wavelengths of the center frequency, i.e., $\Delta z_R=p\pi/k_c$ for some integer $p$. There will altogether be $n_{tot}=\Delta k\Delta z_R/\pi$ [Eq. (39)] basis functions, counting both the sine and cosine functions. {Note, as mentioned above [Eq. (39)], that this is the number of degrees of freedom required to define a function of (spatial) bandwidth $\Delta k$ in a distance $\Delta z_R$ according, for example, to the sampling theorem.}

Note also that these functions are "propagating source" functions, i.e., functions of $\tau_S$ only, and not formally dependent on $z$. Specifically, these are forward propagating source functions that we expect would lead to waves at larger $z_S$ in the receiving volume.

Now let us evaluate the quantity $\langle\psi_A|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_A\rangle$ for one such function, using a cosine function as an explicit example:

$$\begin{aligned} \mu_{cn} &= \langle\psi_{Acn}|\mathsf{G}_{SR}^{\dagger}\mathsf{G}_{SR}|\psi_{Acn}\rangle \\ &= \frac{\nu_o^2 k_c^2 \Delta t_R}{4\pi}\frac{2\nu_o}{\Delta z_S\Delta z_R}\int_{z_{S2}=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\int_{\tau_{S2}=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\int_{z_{S1}=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\int_{\tau_{S1}=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} \mathrm{d}k\,\mathrm{d}\tau_{S1}\mathrm{d}z_{S1}\mathrm{d}\tau_{S2}\mathrm{d}z_{S2} \\ &\quad\times\cos\left[\left(k_c+\frac{2m\pi}{\Delta z_R}\right)\nu_o\tau_{S2}\right]\cos[(k_c+k)\nu_o(\tau_{S1}-\tau_{S2})]\cos\left[\left(k_c+\frac{2m\pi}{\Delta z_R}\right)\nu_o\tau_{S1}\right]. \end{aligned} \tag{B18}$$

Using trigonometric identities, and dropping trigonometric terms that are highly oscillatory in the $\tau_S$ variables and that would therefore average to zero in integrals over $\tau_S$ variables, we obtain

$$\mu_{cn} = \frac{\nu_o k_c^2 \Delta t_R \Delta z_S}{4} \tag{B19}$$

and an identical result for the case of the sine basis functions.

Since we have $n_{tot}$ [Eq. (39)] identical elements in our sum, all with the same $\langle\psi_A|\mathsf{G}_{SR}^\dagger \mathsf{G}_{SR}|\psi_A\rangle$ we obtain a total

$$n_{tot}\langle\psi_A|\mathsf{G}_{SR}^\dagger \mathsf{G}_{SR}|\psi_A\rangle = \frac{\Delta k \Delta z_R}{\pi}\frac{\nu_o k_c^2 \Delta t_R \Delta z_S}{4} = S_{SR}, \tag{B20}$$

as in Eq. (B15). So this Fourier basis set also spans the space $H_S$, and we can use it as a basis for future evaluations.

#### 4. Evaluation of $N_{GS}$

Next we will evaluate

$$N_{GS} \equiv \mathrm{Tr}(\mathsf{G}_S^\dagger \mathsf{G}_S) = \sum_m \gamma_m, \tag{B21}$$

where $\gamma_m = \langle\psi_m|\mathsf{G}_S^\dagger \mathsf{G}_S|\psi_m\rangle$, and $|\psi_m\rangle$ is any convenient set that spans the space $H_S$. For the operator $\mathsf{G}_S$ we have to use the full form as in Eq. (B6), i.e., including the modulus $|z-z_o|$ because we can be interested in waves generated at points $z$ from sources at points of smaller or larger $z_o$ within the scattering volume.

First we note that we can write $k_o = k_c + k$, so that Eq. (B6) becomes

$$\mathsf{G}_S \equiv \mathsf{G}_S(z,t;z_o,t_o) \simeq \frac{-\nu_o k_c}{2\pi}\int_{-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} \sin\{(k_c+k)[\nu_o(t-t_o) - |z - z_o|]\}\mathrm{d}k. \tag{B22}$$

We rewrite the integrand of Eq. (B22) using trigonometric identities, as

$$I_S = \sin\{k_c[\nu_o(t-t_o) - |z-z_o|]\} \times \cos\{k[\nu_o(t-t_o) - |z-z_o|]\} + \cos\{k_c[\nu_o(t-t_o) - |z-z_o|]\} \times \sin\{k[\nu_o(t-t_o) - |z-z_o|]\}. \tag{B23}$$

The second term in Eq. (B23) will disappear by symmetry in the $k$ integral, so it can be dropped. We can now split the first term to eliminate the modulus. We obtain

$$I_S = \Theta(z-z_o)\sin\{k_c[\nu_o(t-t_o) - (z-z_o)]\} \times \cos\{k[\nu_o(t-t_o) - (z-z_o)]\} + \Theta(z_o - z) \times \sin\{k_c[\nu_o(t-t_o) - (z_o - z)]\}\cos\{k[\nu_o(t-t_o) - (z_o - z)]\}, \tag{B24}$$

where $\Theta(x)$ is the Heaviside (step) function. Rewriting in terms of the $\tau$ variables, we have

$$I_S = \Theta(z-z_o)\sin[k_c\nu_o(\tau-\tau_o)]\cos[k\nu_o(\tau-\tau_o)] + \Theta(z_o - z)\sin[k_c\nu_o(\tau-\tau_o) + 2k_c(z-z_o)] \times \cos[k\nu_o(\tau-\tau_o) + 2k(z-z_o)]. \tag{B25}$$

Now the second term in Eq. (B25) is rapidly varying in $z_o$. The chosen Fourier basis functions [Eq. (B16)] do not depend on $z_o$. Hence, this second term will tend to average to zero in the integration over $z_o$ implicit in an expression of the form $\mathsf{G}_S|\psi_A\rangle$ for any of our Fourier basis functions; certainly, the contribution from such a term will be less than $\sim 1/k_c$ in an integral over $z_o$, a contribution that will usually be relatively negligible provided the overall length of the scattering space is $\gg 1/k_c$, i.e., this approximation will be reasonably valid as long as the scattering volume is many wavelengths long (at the center frequency). Hence with this approximation, we have (in the $z$ and $\tau$ variables)

$$\mathsf{G}_S \equiv \mathsf{G}_S(z,\tau;z_o,\tau_o) \simeq \frac{-k_c\nu_o}{2\pi}\Theta(z-z_o)\sin[k_c\nu_o(\tau-\tau_o)] \times \int_{-\Delta k/2}^{\Delta k/2}\cos[k\nu_o(\tau-\tau_o)]\mathrm{d}k. \tag{B26}$$

To evaluate $\mathsf{G}_S^\dagger \mathsf{G}_S$, we use trigonometric identities and drop trigonometric terms that are rapidly varying in the internal integral over $\tau$, giving

$$\mathsf{G}_S^\dagger \mathsf{G}_S(z_2,\tau_2;z_1,\tau_1) = \frac{k_c^2\nu_o^2}{(2\pi)^2}\frac{1}{4}\cos[k_c\nu_o(\tau_1-\tau_2)] \times \int_{-\Delta z_S/2}^{\Delta z_S/2}\Theta(z-z_2)\Theta(z-z_1)\mathrm{d}z I_D, \tag{B27}$$

where

$$I_D = \int_{\tau=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\int_{k_2=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}}\int_{k_1=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}}\mathrm{d}k_1\mathrm{d}k_2\mathrm{d}\tau \times\{\cos[(k_2-k_1)\nu_o\tau]\cos[k_1\nu_o\tau_1 - k_2\nu_o\tau_2] + \cos[(k_2+k_1)\nu_o\tau]\cos[-k_1\nu_o\tau_1 - k_2\nu_o\tau_2]\}.$$

Now

$$\int_{\tau=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\cos[(k_2-k_1)\nu_o\tau]\mathrm{d}\tau = \frac{2\sin\left[\frac{\Delta z_R}{2}(k_2-k_1)\right]}{(k_1-k_1)\nu_o} \to \frac{2\pi}{\nu_o}\delta(k_2-k_1), \tag{B28}$$

and similarly the $\cos[(k_2+k_1)\nu_o\tau]$ integral leads to $(2\pi/\nu_o)\delta(k_2+k_1)$. So

$$I_D \to \frac{2\pi}{\nu_o}\int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}} 2\cos[k\nu_o(\tau_1-\tau_2)]\mathrm{d}k, \tag{B29}$$

and hence

$$\begin{aligned}&\mathsf{G}_S^\dagger\mathsf{G}_S(z_2,\tau_2;z_1,\tau_1)\\&=\frac{k_c^2\nu_o}{4\pi}\cos[k_c\nu_o(\tau_1-\tau_2)]\int_{-\Delta z_S/2}^{\Delta z_S/2}\Theta(z-z_2)\Theta(z-z_1)\mathrm{d}z\\&\times\int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}}\cos[k\nu_o(\tau_1-\tau_2)]\mathrm{d}k.\end{aligned} \tag{B30}$$

Now for one of our cosine basis functions as an example, let us evaluate

$$\gamma_{cn}=\langle\psi_{Acn}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{Acn}\rangle=\frac{k_c^2\nu_o}{4\pi}\frac{2\nu_o}{\Delta z_S\Delta z_R}I_zI_C, \tag{B31}$$

where

$$I_z=\int_{z=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\int_{z_1=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\int_{z_1=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\mathrm{d}z_1\mathrm{d}z_2\mathrm{d}z\Theta(z-z_2)\Theta(z-z_1), \tag{B32}$$

$$\begin{aligned}I_C=&\int_{\tau_2=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\int_{\tau_1=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\int_{k=-\frac{\Delta k}{2}}^{\frac{\Delta k}{2}}\mathrm{d}k\mathrm{d}\tau_1\mathrm{d}\tau_2\\&\times\cos\left[\left(k_c+\frac{2\pi n}{\nu_o}\right)\nu_o\tau_1\right]\cos[k_c\nu_o(\tau_1-\tau_2)]\\&\times\cos[k\nu_o(\tau_1-\tau_2)]\cos\left[\left(k_c+\frac{2\pi n}{\nu_o}\right)\nu_o\tau_2\right].\end{aligned} \tag{B33}$$

Consider the $I_z$ integral. First we note that

$$\int_{z_1=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\Theta(z-z_1)\mathrm{d}z_1=\int_{-\frac{\Delta z_S}{2}}^{z}\mathrm{d}z_1=z+\frac{\Delta z_S}{2}, \tag{B34}$$

and similarly for the integral over $z_2$. Hence

$$I_z=\int_{z=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\left(z+\frac{\Delta z_s}{2}\right)^2\mathrm{d}z=\frac{(\Delta z_S)^2}{3}. \tag{B35}$$

For $I_C$, using trigonometric identities and dropping terms that are very oscillatory in the $\tau$ variables, we obtain

$$I_C=\frac{\pi\Delta z_R}{2\nu_o^2}. \tag{B36}$$

Hence,

$$\begin{aligned}\gamma_{cn}&=\langle\psi_{Acn}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{Acn}\rangle\\&=\frac{k_c^2\nu_o}{4\pi}\frac{2\nu_o}{\Delta z_S\Delta z_R}\frac{(\Delta z_S)^3}{3}\frac{\pi\Delta z_R}{2\nu_o^2}=\frac{k_c^2(\Delta z_S)^2}{12}.\end{aligned} \tag{B37}$$

We will obtain an identical result if we consider the sine basis functions. Hence, with a total of $n_{tot}=(\Delta k\Delta z_R)/\pi$ basis functions altogether, we have the result of Eq. (34).

### 5. Evaluation of $N_C$

*a. Inequality for $N_C$*

From Eqs. (25) and (27), we know that

$$N_c\equiv\mathrm{Tr}(\mathsf{C}^\dagger\mathsf{C})=\sum_m\langle\psi_{Sm}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Sm}\rangle. \tag{B38}$$

We want a result instead with our substitute Fourier basis $|\psi_{An}\rangle$ (here we mean the index $n$ to include both the cosine and sine functions). We know, because of the sum rule Eq. (B20), that $|\psi_{An}\rangle$ spans at least all the space $H_S$ spanned by $|\psi_{Sm}\rangle$. Possibly $|\psi_{An}\rangle$ spans some larger space. We could formally write that some subset $|\psi_{Ap}\rangle$ of the $|\psi_{An}\rangle$ exactly spans the space $H_S$, and the remaining functions $|\psi_{Aq}\rangle$ lie outside of $H_S$. In that case, because the trace of an operator is the same regardless of the basis used for the space of interest,

$$N_C=\sum_p\langle\psi_{Ap}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Ap}\rangle. \tag{B39}$$

Hence

$$\begin{aligned}\sum_n\langle\psi_{An}|\mathsf{C}^\dagger\mathsf{C}|\psi_{An}\rangle&=\sum_p\langle\psi_{Ap}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Ap}\rangle+\sum_q\langle\psi_{Aq}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Aq}\rangle\\&=N_C+\sum_q\langle\psi_{Aq}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Aq}\rangle.\end{aligned} \tag{B40}$$

But any term of the form $\langle\psi|\mathsf{C}^\dagger\mathsf{C}|\psi\rangle\geq 0$, because, by introducing the identity operator based on some complete set $|\alpha\rangle$, such a term can always be written

$$\langle\psi|\mathsf{C}^\dagger\mathsf{C}|\psi\rangle=\sum_j\langle\psi|\mathsf{C}^\dagger|\alpha_j\rangle\langle\alpha_j|\mathsf{C}|\psi\rangle=\sum_j|\langle\alpha_j|\mathsf{C}|\psi\rangle|^2, \tag{B41}$$

which is a sum of positive quantities. Hence

$$N_c\leq\sum_n\langle\psi_{An}|\mathsf{C}^\dagger\mathsf{C}|\psi_{An}\rangle. \tag{B42}$$

So, we now have a way of using our known basis $|\psi_{Am}\rangle$ to evaluate an upper bound for $N_C$.

*b. Evaluation of Bound on $N_C$*

Now we can evaluate a specific $\langle\psi_{Am}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Am}\rangle$, for example using one of the cosine basis functions $|\psi_{Acn}\rangle$. This evaluation is particularly simple because we have chosen here a totally local response function, i.e., the source at point $z$ inside the source volume is dependent only on the wave at that precise point $z$. Hence we have

$$\begin{aligned}&\langle\psi_{Acn}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Acn}\rangle\\&=\frac{2\nu_o}{\Delta z_S\Delta z_R}\int_{z=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\int_{\tau=-\frac{\Delta z_R}{2\nu_o}}^{\frac{\Delta z_R}{2\nu_o}}\mathrm{d}\tau\mathrm{d}z\\&\times\cos^2\left[\left(k_c+\frac{2n\pi}{\Delta z_R}\right)\nu_o\tau\right]\left|\eta\left(z,\left[k_c+\frac{2n\pi}{\Delta z_R}\right]\nu_o\right)\right|^2\\&\simeq\frac{1}{\Delta z_S}\int_{z=-\frac{\Delta z_S}{2}}^{\frac{\Delta z_S}{2}}\left|\eta\left[z,\left(k_c+\frac{2n\pi}{\Delta z_R}\right)\nu_o\right]\right|^2\mathrm{d}z.\end{aligned} \tag{B43}$$

Here we have replaced the $\cos^2$ term by its average value

because we are integrating a rapidly varying term over an arbitrarily large distance (the length of the receiving space).

The result for a sine basis function $|\psi_{Asn}\rangle$ is identical. Hence

$$N_C \leq \sum_m \langle \psi_{Am}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Am}\rangle = 2\sum_n \langle \psi_{Acn}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Acn}\rangle. \tag{B44}$$

Note, incidentally, that if we know nothing about $\eta$ other than that $|\eta|$ is less than some maximum value $\eta_{\max}$, then we can always say that

$$\langle \psi_{Acn}|\mathsf{C}^\dagger\mathsf{C}|\psi_{Acn}\rangle \leq \eta_{\max}^2, \tag{B45}$$

and use that instead of the result of Eq. (B43) to obtain an upper bound for the sum.

Now, because $\Delta z_R$ is chosen to be very large, we can approximate the sum by an integral, i.e.,

$$\sum_{n=n_{\min}}^{n_{\max}} \simeq \frac{dn}{d\omega}\int_{\omega_{\min}}^{\omega_{\max}} d\omega. \tag{B46}$$

We have $\omega_n = (k_c + 2n\pi/\Delta z_R)\nu_o$ so

$$\frac{d\omega}{dn} = \frac{2\pi\nu_o}{\Delta z_R}, \tag{B47}$$

and we use Eq. (B17) for the bounds on $n$. Hence we obtain the results of Eqs. (35) and (36).

## APPENDIX C: CRITERION FOR COMPLETENESS FOR A BASIS SET IN $H_S$

Here we prove by *reductio ad absurdum* that a set of functions $|\psi_{An}\rangle$ is also complete for the space $H_S$ if it satisfies Eq. (B13). Suppose that this set $|\psi_{An}\rangle$ is not complete for the space. Then there is at least one $|\psi_{Sq}\rangle$ of the original basis functions that cannot be expressed as a linear combination of the $|\psi_{An}\rangle$, and hence can be written as

$$|\psi_{Sq}\rangle = \left(\sum_n a_n|\psi_{An}\rangle\right) + |\psi_{extra}\rangle, \tag{C1}$$

where $a_n = \langle\psi_{An}|\psi_{Sq}\rangle$ and

$$\langle \psi_{An}|\psi_{extra}\rangle = 0, \quad \text{for all } n, \tag{C2}$$

i.e., we have had to introduce some additional function $|\psi_{extra}\rangle$, orthogonal to all the $|\psi_{An}\rangle$, to express $|\psi_{Sq}\rangle$.

Now, we can define a sum

$$S_A = \sum_n \langle \psi_{An}|\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR}|\psi_{An}\rangle = \sum_{n,m,p} \langle \psi_{An}|\psi_{Sm}\rangle\langle\psi_{Sm}|\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR}|\psi_{Sp}\rangle\langle\psi_{Sp}|\psi_{An}\rangle, \tag{C3}$$

where we have introduced identity operators of the form $\mathsf{I}_{HS} = \sum_m |\psi_{Sm}\rangle\langle\psi_{Sm}|$. Since

$$\langle \psi_{Sm}|\mathsf{G}_{SR}^\dagger\mathsf{G}_{SR}|\psi_{Sp}\rangle = s_m^* s_p \langle\phi_{Rm}|\phi_{Rp}\rangle = s_m^* s_p \delta_{mp} = |s_m|^2\delta_{mp}, \tag{C4}$$

then

$$S_A = \sum_m |s_m|^2 \langle\psi_{Sm}|\sum_n |\psi_{An}\rangle\langle\psi_{An}|\psi_{Sm}\rangle. \tag{C5}$$

Now, for our at least one basis function $|\psi_{Sq}\rangle$, we know from Eqs. (C1) and (C2)

$$\sum_n |\psi_{An}\rangle\langle\psi_{An}|\psi_{Sq}\rangle = \sum_n a_n|\psi_{An}\rangle = |\psi_{Sq}\rangle - |\psi_{extra}\rangle, \tag{C6}$$

and we also can note that

$$\langle\psi_{Sq}|\psi_{extra}\rangle = \left[\left(\sum_n a_n^*\langle\psi_{An}|\right) + \langle\psi_{extra}|\right]|\psi_{extra}\rangle = \langle\psi_{extra}|\psi_{extra}\rangle > 0. \tag{C7}$$

Hence

$$\langle\psi_{Sq}|\sum_n |\psi_{An}\rangle\langle\psi_{An}|\psi_{Sq}\rangle = \langle\psi_{Sq}|\psi_{Sq}\rangle - \langle\psi_{extra}|\psi_{extra}\rangle = 1 - \langle\psi_{extra}|\psi_{extra}\rangle < 1. \tag{C8}$$

Now, for all other basis functions $|\psi_{Sj}\rangle$ that can be expanded on the basis set $|\psi_{An}\rangle$, i.e., for which $|\psi_{Sj}\rangle = \sum_n b_n|\psi_{An}\rangle$ for some coefficients $b_n$, then

$$\langle\psi_{Sj}|\sum_n |\psi_{An}\rangle\langle\psi_{An}|\psi_{Sj}\rangle = \sum_n |b_n|^2 = 1. \tag{C9}$$

Since for at least one $m=q$ in the summation in Eq. (C5), $|s_m|^2$ (which is definitely $>0$) is multiplied by a number $<1$, and for all others it is multiplied by a number never greater than 1, then $S_A < \sum_m |s_m|^2 = S_{SR}$. But this result contradicts the original presumption Eq. (B13). Hence this new set, because it satisfies the condition Eq. (B13), is complete for this space $H_S$.

We do have to bear in mind at least in principle that the set $|\psi_{An}\rangle$ could also span some larger space; such a set could still satisfy Eq. (B13). Since we are looking for an upper limit in our final sum rule, this will not invalidate the upper limit. (In fact, we will typically choose a set that does not span a larger space, but the condition Eq. (B13) is not sufficient to prove that it does not.)

## APPENDIX D: SCALING OF $N_{GS}$ SUM INDEPENDENT OF BASIS

For all the members of the Fourier basis set discussed in Appendix B above, as shown in Eq. (B37), the contributions to the $N_{GS}$ sum are equal, i.e., writing $|\psi_{An}\rangle$ to represent all of the basis set elements (including both the sine and cosine members)

$$\langle\psi_{An}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{An}\rangle = k_c^2(\Delta z_S)^2 = \beta. \tag{D1}$$

It is also true that for different members of the basis set

$$\langle\psi_{Am}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{An}\rangle = 0 \quad m \neq n. \tag{D2}$$

(For basis set members of different frequencies, the result averages to zero in the integral over time, and for sine and cosine members at the same frequency, the result is also zero because of the $\pi/2$ phase difference between $\mathsf{G}_S|\psi_{An}\rangle$ and $\mathsf{G}_S|\psi_{Am}\rangle$ in the integral over time.)

Suppose now we are interested in some other complete orthonormal basis set $|\psi_{Pq}\rangle$, such as a pulsed basis, for the waves in the scattering volume, or more formally the

functions that span the space $H_S$. Then we can expand those functions in the Fourier basis, i.e., $|\psi_{Pq}\rangle = \Sigma_n a_{qn}|\psi_{An}\rangle$, where by orthonormality we know that

$$\langle\psi_{Pq}|\psi_{Pq}\rangle = \sum_{n,p}\langle\psi_{Ap}|a^*_{qp}a_{qn}|\psi_{An}\rangle = \sum_n |a_{qn}|^2 = 1. \quad \text{(D3)}$$

Then

$$\langle\psi_{Pq}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{Pq}\rangle = \sum_{n,p} a^*_{qp}a_{qn}\langle\psi_{Ap}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{An}\rangle$$
$$= \sum_n |a_{qn}|^2\langle\psi_{An}|\mathsf{G}_S^\dagger\mathsf{G}_S|\psi_{An}\rangle = \beta\sum_n |a_{qn}|^2 = \beta. \quad \text{(D4)}$$

Hence the contributions to $N_{GS}$ associated with each of the new basis functions are also all identical, and hence $N_{GS}$ simple scales down as we reduce the number of these new basis functions as we restrict the input space.

## ACKNOWLEDGMENTS

This work was supported by the Defence Advanced Research Projects Agency/Army Research Office Chip Scale Wavelength Division Multiplexing program, the DARPA Optocenters Program, the DARPA Cognitively Augmented Design for Quantum Technology program, and the Air Force Office of Scientific Research "Plasmon Enabled Nanophotonic Circuits" Multidisciplinary University Research Initiative. I am also pleased to acknowledge the hospitality of Prof. Anna Vinattieri and the Department of Physics at the Universitá degli Studi di Firenze, where this work was completed.

## REFERENCES

1. G. Lenz and C. K. Madsen, "General optical all-pass filter structures for dispersion control in WDM systems," J. Lightwave Technol. **17**, 1248–1254 (1999).
2. M. Sumetsky and B. Eggleton, "Modeling and optimization of complex photonic resonant cavity circuits," Opt. Express **11**, 381–391 (2003).
3. O. Schwelb, "Transmission, group delay, and dispersion in single-ring optical resonators and add/drop filters-a tutorial overview," J. Lightwave Technol. **22**, 1380–1394 (2004).
4. K. Yu and O. Solgaard, "Tunable optical transversal filters based on a Gires-Tournois interferometer with MEMS phase shifters," IEEE J. Sel. Top. Quantum Electron. **10**, 588–597 (2004).
5. T. Baba and T. Matsumoto, "Resolution of photonic crystal superprism," Appl. Phys. Lett. **81**, 2325–2327 (2002).
6. B. Momeni and A. Abidi, "Optimization of photonic crystal demultiplexers based on the superprism effect," Appl. Phys. B **77**, 555–560 (2003).
7. C. Y. Luo, M. Soljacic, and J. D. Joannopoulos, "Superprism effect based on phase velocities," Opt. Lett. **29**, 745–747 (2004).
8. L. Wu, M. Mazilu, J.-F. Gallet, and T. F. Krauss, "Dual lattice photonic-crystal beam splitters," Appl. Phys. Lett. **86**, 211106 (2005).
9. B. Momeni and A. Abidi, "Systematic design of superprism-based photonic crystal demultiplexers," IEEE J. Sel. Areas Commun. **23**, 1355–1364 (2005).
10. M. Gerken and D. A. B. Miller, "Multilayer thin-film structures with high spatial dispersion," Appl. Opt. **42**, 1330–1345 (2003).
11. M. Gerken and D. A. B. Miller, "Photonic nanostructures for wavelength division multiplexing," Proc. SPIE **5597**, 82–96 (2004).
12. M. Gerken and D. A. B. Miller, "Limits on the performance of dispersive thin-film stacks," Appl. Opt. **44**, 3349–3357 (2005).
13. M. Gerken and D. A. B. Miller, "The relationship between the superprism effect in one-dimensional photonic crystals and spatial dispersion in nonperiodic thin-film stacks," Opt. Lett. **30**, 2475–2477 (2005).
14. R. E. Klinger, C. A. Hulse, C. K. Carniglia, and R. B. Sargent, "Beam displacement and distortion effects in narrowband optical thin-film filters," Appl. Opt. **45**, 3237–3242 (2006).
15. Y. Jiao, S. H. Fan, and D. A. B. Miller, "Designing for beam propagation in periodic and nonperiodic photonic nanostructures: Extended Hamiltonian method," Phys. Rev. E **70**, 036612-1–036612-9 (2004).
16. Y. Jiao, 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," Opt. Lett. **30**, 141–143 (2005).
17. Y. Jiao, 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**, 266–279 (2006).
18. P. J. van Heerden, "Theory of optical information storage in solids," Appl. Opt. **2**, 393–400 (1963).
19. H. Kogelnik, "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. **48**, 2909–2947 (1969).
20. H. Lee, X. G. Gu, and D. Psaltis, "Volume holographic interconnections with maximal capacity and minimal cross talk," J. Appl. Phys. **65**, 2191–2194 (1989).
21. X. M. Yi, P. Yeh, C. Gu, and S. Campbell, "Crosstalk in volume holographic memory," Proc. IEEE **87**, 1912–1930 (1999).
22. K. Tian and G. Barbastathis, "Cross talk in resonant holographic memories," J. Opt. Soc. Am. A **21**, 751–756 (2004).
23. J. Shamir, "Analysis of volume holographic storage allowing large-angle illumination," J. Opt. Soc. Am. B **22**, 975–986 (2005).
24. A. Yariv, Y. Xu, R. K. Lee, and A. Scherer, "Coupled-resonator optical waveguide: a proposal and analysis," Opt. Lett. **24**, 711–713 (1999).
25. G. Lenz, B. J. Eggleton, C. K. Madsen, and R. E. Slusher, "Optical delay lines based on optical fibers," IEEE J. Quantum Electron. **37**, 525–532 (2001).
26. 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).
27. M. D. Stenner, M. A. Neifeld, Z. Zhu, A. M. C. Dawes, and D. J. Gauthier, "Distortion management in slow-light pulse delay," Opt. Express **13**, 9995–10002 (2005).
28. M. Povinelli, S. Johnson, and J. Joannopoulos, "Slow-light, band-edge waveguides for tunable time delays," Opt. Express **13**, 7145–7159 (2005).
29. Z. S. Yang, N. H. Kwong, R. Binder, and A. L. Smirl, "Distortionless light pulse delay in quantum-well Bragg structures," Opt. Lett. **30**, 2790–2792 (2005).
30. M. R. Fisher and S.-L. Chuang, "Variable group delay and pulse reshaping of high bandwidth optical signals," IEEE J. Quantum Electron. **41**, 885–891 (2005).
31. J. Sharping, Y. Okawachi, J. van Howe, C. Xu, Y. Wang, A. Willner, and A. Gaeta, "All-optical, wavelength and bandwidth preserving, pulse delay based on parametric wavelength conversion and dispersion," Opt. Express **13**, 7872–7877 (2005).
32. M. S. Bigelow, N. N. Lepeshkin, H. Shin, and R. W. Boyd, "Propagation of smooth and discontinuous pulses through materials with very large or very small group velocities," J. Phys.: Condens. Matter **18**, 3117–3126 (2006).
33. A. V. Uskov, F. G. Sedgwick, and C. J. Chang-Hasnain, "Delay limit of slow light in semiconductor optical amplifiers," IEEE Photon. Technol. Lett. **18**, 731–733 (2006).
34. 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).
35. J. B. Khurgin, "Performance limits of delay lines based on optical amplifiers," Opt. Lett. **31**, 948–950 (2006).
36. D. A. B. Miller, "Spatial channels for communicating with waves between volumes," Opt. Lett. **23**, 1645–1647 (1998).
37. D. A. B. Miller, "Communicating with waves between volumes – evaluating orthogonal spatial channels and limits on coupling strengths," Appl. Opt. **39**, 1681–1699 (2000).
38. R. Piestun and D. A. B. Miller, "Electromagnetic degrees of freedom of an optical system," J. Opt. Soc. Am. A **17**, 892–902 (2000).
39. G. W. Hanson and A. B. Yakovlev, *Operator Theory for Electromagnetics* (Springer-Verlag, 2002), p. 172.
40. G. W. Hanson and A. B. Yakovlev, *Operator Theory for Electromagnetics* (Springer-Verlag, 2002), p. 259.