Static two-dimensional aperture coding for multimodal, multiplex spectroscopy

Michael E. Gehm, Scott T. McCain, Nikos P. Pitsianis, David J. Brady, Prasant Potuluri, and Michael E. Sullivan

We propose a new class of aperture-coded spectrometer that is optimized for the spectral characterization of diffuse sources. The instrument achieves high throughput and high spectral resolution by replacing the slit of conventional dispersive spectrometers with a more complicated spatial filter. We develop a general mathematical framework for deriving the required aperture codes and discuss several appealing code families. Experimental results validate the performance of the instrument. © 2006 Optical Society of America

OCIS codes: 300.6190, 120.6200.

1. Introduction

Traditional slit-based dispersive spectrometers are ill-suited for the task of diffuse source spectroscopy—estimating the mean spectral density of a source that is highly spatially multimodal and therefore large in both spatial and angular extent. In this paper we describe a spectrometer design that is optimized for working with such sources.

Previously, we considered how the constant radiance theorem complicates the characterization of diffuse sources.[1] Entropic considerations require that the modal volume of a source cannot be reduced without a concomitant reduction in power. As such, the brightness of diffuse sources cannot be increased.

This is particularly unfortunate in the case of spectroscopy, as traditional spectrometers utilize narrowband spatial filtering to disambiguate between spatial and spectral modes of the field. Consider the schematic of a slit-based spectrometer shown in Fig. 1. The dispersive element produces a wavelength-dependent shift of the image of the input slit. Since each spectral channel must correspond to a unique shift, the spectral width of a resolution element is directly proportional to the slit width ($\Delta\lambda \propto \Delta x$). It is this relationship that provides the challenge to diffuse-source spectroscopy. Achieving a reasonable spectral resolution requires that the input slit to the spectrometer be narrow. However, because the source is diffuse, the radiation field cannot be focused through the slit. Instead, only a small fraction of the light can enter the instrument. If the source is weak as well as diffuse, then often the instrument is so photon starved that no spectral measurement is possible.

The throughput of an optical instrument, sometimes referred to as the étendue, can be approximated as the product of the area of the input aperture and the solid angle from which the instrument will accept light:

$$G \approx A\Omega. \tag{1}$$

The acceptance solid angle is determined by the internal optics of an instrument. For a given optical arrangement, the only way to increase the étendue of the system is to increase the size of the input aperture. However, as we have seen above, such an approach reduces the resolution of the spectrometer as it increases the throughput.

The two main challenges in diffuse-source spectroscopy are these:

1. Maximizing spectrometer throughput without sacrificing spectral resolution
2. Maximizing the signal-to-noise-ratio (SNR) of the estimated spectrum for a given system throughput and detector noise

M. E. Gehm (mgehm@ee.duke.edu), S. T. McCain, N. P. Pitsianis, and D. J. Brady are with the Duke University Fitzpatrick Center for Photonics, Box 90291, Durham, North Carolina 27708. P. Potuluri and M. E. Sullivan are with Centice Corp., 4020 Stirrup Creek Drive, Suite 115, Research Triangle Park, North Carolina 27703.

Received 10 August 2005; accepted 31 October 2005; posted 21 November 2005 (Doc. ID 64019).

0003-6935/06/132965-10$15.00/0

© 2006 Optical Society of America

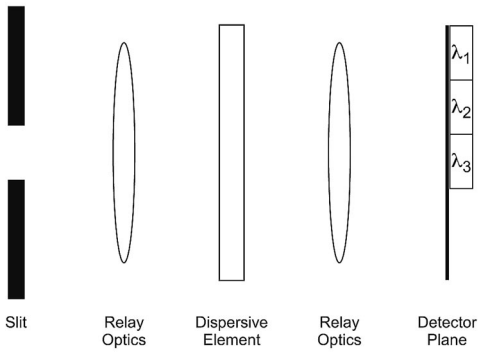

Fig. 1. Schematic of a traditional slit-based spectrometer. Internal optical elements form an image of the slit at wavelength-dependent locations on the detector plane. Spectral resolution comes from the capability for spatially distinguishing these different locations. As such, the width of a spectral channel is proportional to the width of the input slit.

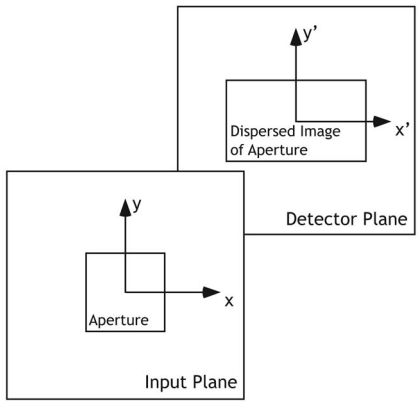

Fig. 2. Definition of the coordinate systems used throughout the paper.

Both problems have been long studied, and a number of ingenious designs have been proposed to address one or both. In the language of the field, a design that solves the first problem is said to have a Jacquinot (or large-area or throughput) advantage.[2] A design that solves the second problem is said to have a Fellgett (or multiplex) advantage.[3]

The earliest approach to solving these problems was through coded-aperture spectroscopy—the replacement of the input slit with a more complicated pattern of openings. Golay created the first coded-aperture spectrometer in the early 1950s,[4,5] and advancements followed rapidly over the next several decades.[6] As the mathematical treatments gained sophistication, the appeal of apertures based on Hadamard matrices[7] became apparent,[8,9] and the majority of coded-aperture spectrometers became Hadamard-transform (HT) spectrometers.[10–14] The study of HT spectrometers remains active to this day.[15–18] Over most of their development, however, HT spectrometers had only single-channel detectors or limited arrays of discrete detectors. As a result, most designs contained at least two coding apertures—at both the input and the output planes. Further, the designs usually required motion of one mask with respect to the other. The majority of the resulting instruments exhibited only the Jacquinot advantage or the Fellgett advantage.

Aperture coding is not the only approach to solving these spectrometer design problems, however. Interferometric spectrometers, such as the Fourier-transform (FT) spectrometers,[19] also can exhibit the Jacquinot and Fellgett advantages. The FT spectrometer, in fact, exhibits both. However, the majority of FT spectrometers contain mechanical scanning elements.

In this paper we propose a new type of coded-aperture spectrometer that exhibits both the Jacquinot and the Fellgett advantages and that does so with a completely static design that is simpler, cheaper, and more robust than the dynamic designs of most HT and FT spectrometers. Our design concept centers around modern multichannel detectors. We call this general design a static, multimodal, multiplex spectrometer (static MMS).

In the remainder of this paper we derive a mathematical model of a dispersive spectrometer and show how [along with two-dimensional (2-D) multichannel detectors] simple aperture codes can result in both the Jacquinot and the Fellgett advantages. Further, we derive several different classes of aperture patterns that are of interest to us. Finally, we present experimental results from one of the many static MMS designs we have constructed.

2. Mathematical Treatment

A. System Model

We begin by considering the following simplified model of a dispersive spectrometer:

$$I(x',\,y') = \int\int\int \mathrm{d}\lambda\mathrm{d}x\mathrm{d}y\; H(x,\,x',\,y,\,y';\,\lambda)T(x,\,y) \times S(x,\,y;\,\lambda). \quad (2)$$

Here $H(x, x', y, y'; \lambda)$ is the kernel describing propagation through the spectrometer, $T(x, y)$ is a transmission function describing the input aperture, and $S(x, y; \lambda)$ is the input spectral density at position (x, y). We use unprimed variables for quantities defined in the input plane, while primed variables are used for the quantities in the detector plane. A simple schematic of these coordinates is shown in Fig. 2.

We take $H(x, y; \lambda) = \delta(y - y')\delta\{x - [x' + \alpha(\lambda - \lambda_c)]\}$ as the propagation kernel. This kernel represents a basic dispersive spectrometer with unity-magnification optics, a linear dispersion α in the x direction, and a center wavelength of λ_c for an aperture at $x = 0$. Inserting this into Eq. (2) and performing the λ and y integrals yield

$$I(x',\,y') = \int \mathrm{d}x\; T(x,\,y')S\left(x,\,y';\,\frac{x - x'}{\alpha} + \lambda_c\right). \quad (3)$$

A traditional slit spectrometer takes the input aperture as $T(x, y) = \delta(x)$ so that

$$I(x', y') = S\left(0, y'; \lambda_c - \frac{x'}{\alpha}\right). \quad (4)$$

Thus the intensity profile in the detector plane is a direct estimate of the spectral density at the slit location.

However, as discussed above, the drawback to such an approach is that the throughput of the system is severely curtailed. We wish to consider more complicated aperture patterns so that we can increase the photon-collection efficiency of the system.

Our fundamental goal is to develop an aperture code that allows us to estimate the mean spectrum across an extended aperture, which we define as

$$S_{\text{mean}}(\lambda) \propto \int\int dx dy\, S(x, y; \lambda). \quad (5)$$

B. Coding Approach

In the more general case, to convert the intensity profile of Eq. (3) into an estimate of the mean spectrum, we multiply it by an analysis function $\tilde{T}(x'', y')$ and integrate over the extent of the patterns in y':

$$\begin{aligned} E(x', x'') &= \int_{y'_{\min}}^{y'_{\max}} dy'\, \tilde{T}(x'', y') I(x', y') \\ &= \int_{y'_{\min}}^{y'_{\max}} dy' \int dx\, \tilde{T}(x'', y') T(x, y') \\ &\quad \times S\left(x, y'; \frac{x - x'}{\alpha} + \lambda_c\right). \end{aligned} \quad (6)$$

We then make the assumption that $S(x, y'; \lambda)$ is constant, or slowly varying in y'. We can write this as

$$S(x, y'; \lambda) \approx I(y') S(x; \lambda). \quad (7)$$

Inserting this into Eq. (6), we find

$$\begin{aligned} E(x', x'') &\approx \int_{y'_{\min}}^{y'_{\max}} dy' \int dx\, \tilde{T}(x'', y') T(x, y') I(y') \\ &\quad \times S\left(x; \frac{x - x'}{\alpha} + \lambda_c\right). \end{aligned} \quad (8)$$

If $T(x, y')$ and $\tilde{T}(x'', y')$ are constructed such that

$$\int_{y'_{\min}}^{y'_{\max}} dy'\, \tilde{T}(x'', y') T(x, y') I(y') = \beta \delta(x - x''), \quad (9)$$

then our estimate becomes

$$\begin{aligned} E(x', x'') &\approx \beta \int dx\, \delta(x - x'') S\left(x; \frac{x - x'}{\alpha} + \lambda_c\right) \\ &\approx \beta S\left(x''; \frac{x'' - x'}{\alpha} + \lambda_c\right). \end{aligned} \quad (10)$$

This result can be interpreted as a 2-D function containing estimates of the input spectrum at different input locations. A slice through this function at a constant value of x'' corresponds to the input spectrum at a particular value of x. In other words, if we halt our analysis at this point, we have created a one-dimensional (1-D) imaging spectrometer. We will discuss the implications of this imaging capability in a future publication. In this paper we wish to proceed further and convert $E(x', x'')$ into an estimate of $S_{\text{mean}}(\lambda)$.

Since the spectral estimates of Eq. (10) are shifted with respect to each other, to calculate the mean spectrum, we must integrate along the line $x' = \lambda\alpha + x''$:

$$\begin{aligned} S_{\text{mean}}(\lambda_c - \lambda) &\propto \int\int dx'' dx'\, \delta[x' - (\lambda\alpha + x'')] E(x', x'') \\ &\propto \int dx''\, S(x''; \lambda_c - \lambda). \end{aligned} \quad (11)$$

Thus with appropriately designed input apertures and analysis functions, we can convert an intensity profile at the detector plane into an estimate of the input spectrum. But how does one perform this design subject to the constraint of Eq. (9)?

C. Orthogonal and Independent Column Codes

If we rewrite Eq. (9) in a form where x and x'' are not coordinates but instead parameters

$$\int_{y'_{\min}}^{y'_{\max}} dy'\, \tilde{T}_{x''}(y') T_x(y') I(y') = \beta \delta(x - x''), \quad (12)$$

we arrive at an equation that is identical to the orthogonality constraint for eigenfunctions in Sturm–Liouville theory,[20] where $I(y')$ is the weighting function and β is the norm. Therefore we can meet the design requirement by basing the input aperture pattern on any family of orthogonal functions.

Using the language of Sturm–Liouville theory, if T and $\tilde{T}$ are the same set of codes, we say that the system is self-adjoint. In this case the complete set of codes in T can be viewed as abstract vectors defining an orthogonal basis on a Hilbert space, and we refer to a family of this type as an orthogonal column code.

If T and $\tilde{T}$ are not the same set of codes, the system is said to be non-self-adjoint. Here the complete set of codes in T can be viewed as abstract vectors defining

a nonorthogonal basis on a Hilbert space. We refer to a family of this type as an independent column code.

In Eq. (12) x and x'' can be either continuous or discrete parameters, depending on the eigenvalue spectrum of the chosen family of functions. In the discrete case the Dirac delta function $\delta(x - x'')$ is properly replaced with the Kronecker delta $\delta_{x,x''}$. Further, in this case the input mask and analysis pattern will be pixelated in the x and x'' directions, respectively.

D. Heuristic Treatment

Considerable insight can be gained from a heuristic view of orthogonal and independent column coding. From Eq. (3), we see that, for the case of uniform input intensity, the output intensity distribution is a convolution of the input aperture and the input spectrum:

$$I(x',\, y') = \int \mathrm{d}x\; T(x,\, y')S\left(\frac{x - x'}{\alpha}\right). \qquad (13)$$

Thus the light falling at a given value of x' in the detector plane arises from a combination of different wavelengths passing through different locations on the input aperture. A well-designed code allows us to break this ambiguity and determine the spectral content of the light. By choosing a family of functions as our transmission mask, we provide a unique code to each possible x location in the input plane. We can view the transmission pattern at position x as an abstract vector $|T_x\rangle$. The full family of transmission patterns then forms a basis $\{|T_x\rangle\}$. If we consider the light distribution falling at a given x' location in the detector plane as the abstract vector $|I_{x'}\rangle$, the contribution from position x on the input aperture is given simply by $\langle T_x|I_{x'}\rangle$, the projection of $|I_{x'}\rangle$ onto the adjoint of the corresponding vector $|T_x\rangle$ ($\langle T_x| \equiv |T_x\rangle^{\dagger}$). Because only light of wavelength $\lambda_{x,x'} = (x - x')/\alpha + \lambda_c$ can propagate from x to x', this inner product also represents an estimate of $S(x, \lambda_{x,x'})$. Forming the set of all inner products of the form $\langle T_x|I_{x'}\rangle$, yields the 2-D spectral estimate function E.

3. Specific Mask Families

In Section 2 we demonstrated the appeal of using orthogonal or independent column codes as aperture mask patterns in dispersive spectroscopy. The number of possible families is, of course, infinite. In this section we describe certain specific families of interest.

A. Harmonic Masks

Above we allude to the fact that the intensity profile $I(y)$ takes the role of the weighting function in Sturm–Liouville theory and, in conjunction with the integration limits, controls the nature of the orthogonal functions. If we consider a uniform input intensity, symmetric integration limits ($y_{\min} = -Y$, $y_{\max} = Y$), and a discrete eigenvalue spectrum, we get the constraint (for the remainder of this section we use y in place of y', as there is no chance of confusion)

$$\int_{-Y}^{Y} \mathrm{d}y\; \widetilde{T}_{x''}(y)T_x(y) = \beta\delta_{x,x''}, \qquad (14)$$

which is satisfied by the well-known harmonic functions. For example (using $\mathbb{Z}^*$ to represent the nonnegative integers),

$$T_x,\; \widetilde{T}_{x''} \in \left\{\cos\left(m\,\frac{y\pi}{Y}\right)\right\}, \qquad m \in \mathbb{Z}^* \qquad (15)$$

is an obviously self-adjoint solution to Eq. (14). However, there is a problem with this set of functions. Because we are working with incoherent illumination, T_x can modulate only the light intensity, not the field. As a result, we are forced to consider only functions with values in the interval [0, 1].

This has a significant effect on the nature of the solutions that we may find. It is not possible to find a self-adjoint set of continuous functions that meets this requirement. Since negative values are not allowed, the inner product between any two such functions is positive definite. Hence the functions in T_x cannot also be the functions in $\widetilde{T}_{x''}$. We are forced to consider an independent column code.

One possible independent column code based on harmonic functions is

$$T_x \in \left\{\frac{1}{2}\left[1 + \cos\left(m\,\frac{y\pi}{Y}\right)\right]\right\}, \qquad m \in \mathbb{Z}^*. \qquad (16)$$

The corresponding analysis codes are then

$$\widetilde{T}_{x''} \in \left\{2\,\cos\left(m\,\frac{y\pi}{Y}\right)\right\}, \qquad m \in \mathbb{Z}^*. \qquad (17)$$

An aperture mask based on this independent column code with $m = 1 - 64$ is shown in Fig. 3.

B. Legendre Masks

There is another set of famous orthogonal functions that satisfies the constraint of Eq. (14)—the Legendre polynomials

$$P_n(y) = \frac{1}{2^n}\sum_{m=0}^{\lfloor n/2\rfloor}(-1)^m\binom{n}{m}\binom{2n-2m}{n}y^{n-2m}, \qquad (18)$$

where

$$\binom{a}{b} = \frac{a!}{(a-b)!\,b!}. \qquad (19)$$

As was the case with the harmonic masks, the functions form a self-adjoint set of codes:

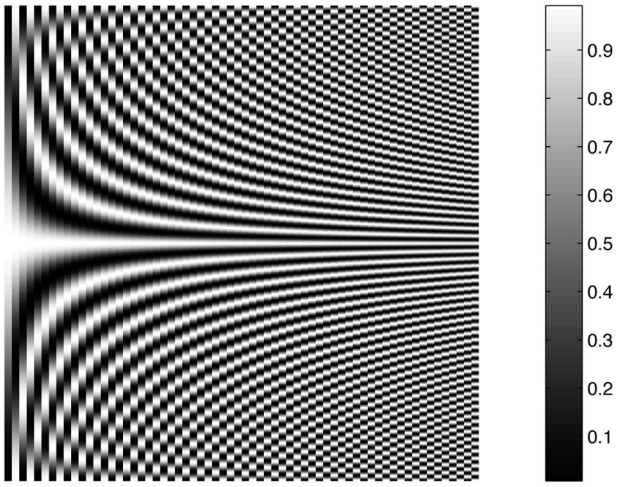

Fig. 3. Aperture pattern for an independent column code based on harmonic functions. The codes were chosen such that the transmission has physically realizable values in the interval [0, 1]. Note that the pattern is continuous vertically but discrete horizontally.

$$T_x,\ \widetilde{T}_{x''} \in \left\{ P_m\left(\frac{y}{\overline{Y}}\right) \right\}, \qquad m \in \mathbb{Z}^*. \tag{20}$$

However, as above, these codes involve modulation values that are not physically possible in an incoherent system. Scaling to produce physically realizable values results in an independent column code. One possible version is

$$T_x \in \left\{ \frac{1}{2}\left[1 + P_m\left(\frac{y}{\overline{Y}}\right) \right] \right\}, \qquad m \in \mathbb{Z}^*. \tag{21}$$

The corresponding analysis codes are then

$$\widetilde{T}_{x''} \in \left\{ 2P_m\left(\frac{y}{\overline{Y}}\right) \right\}, \qquad m \in \mathbb{Z}^*. \tag{22}$$

An aperture mask based on this independent column code with $m = 1 - 64$ is shown in Fig. 4.

C. Hadamard Masks

In the previous sections we consider only continuous functions of y as possible code families. Based on the heuristic insights of Subsection 2.D, it seems reasonable to also consider discrete functions of y, particularly the pixelated functions based on Hadamard matrices.[7] We define H_n as an order-n Hadamard matrix and use the symbols $H_n(:, m)$ and $H_n(m, :)$ to refer to the mth column and row of H_n, respectively. Then

$$T_x,\ \widetilde{T}_{x''} \in \{H_n(:,\ m)\}, \qquad m \in \mathbb{Z}^*;\ m \le n \tag{23}$$

is a self-adjoint set of codes. Given that the elements of a Hadamard matrix are either 1 or -1, this is again not realizable with incoherent illumination. Shifting and scaling the code values results in a non-

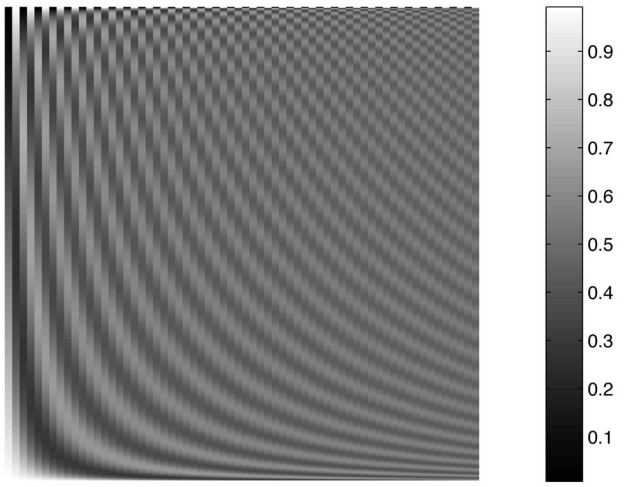

Fig. 4. Aperture pattern for an independent column code based on Legendre polynomials. The codes were chosen such that the transmission has physically realizable values in the interval [0, 1]. Note that the pattern is continuous vertically but discrete horizontally.

self-adjoint independent column code

$$T_x \in \left\{ \frac{1}{2}\left[1 - H_n(:,\ m) \right] \right\}, \qquad m \in \mathbb{Z}^*;\ m \le n. \tag{24}$$

With the corresponding analysis code

$$\widetilde{T}_{x''} \in \{2H_n(:,\ m)\}, \qquad m \in \mathbb{Z}^*;\ m \le n. \tag{25}$$

This particular choice is known as an S matrix in the traditional Hadamard literature. An aperture based on an S-matrix code is shown in Fig. 5.

In all the aperture masks discussed so far, we shift and scale the code values to achieve a physically realizable modulation. In every case the application of

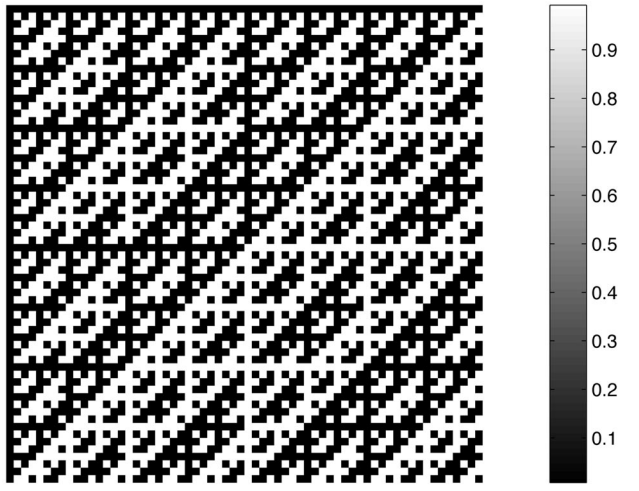

Fig. 5. Aperture pattern for an independent column code based on a Hadamard S matrix. The codes were chosen such that the transmission has physically realizable values in the interval [0, 1]. Note that the pattern is discrete both horizontally and vertically.

a shift turns an orthogonal column code into an independent column code. However, if we had a method for identifying the sign of a code value, then we could apply the sign in the software (by multiplying the measured value by -1 where appropriate). By adding this extra computational step, we could achieve a physically realizable aperture, avoid the need for a shift, and have a self-adjoint set of codes.

Unfortunately, any row of the code contains both positive and negative values. The multiplex nature of the system then ensures that light from these different regions are combined on the detector plane, making it impossible to apply the appropriate weighting in software. However, if we could segregate positive and negative regions of the code onto separate rows, then we could apply a weighting to entire rows in the detector plane and achieve our goal. We refer to codes that have been modified in this manner as row doubled.

To row double a Hadamard matrix, we replace each original row $H_n(m, :)$ with two rows:

$$H_n(m, :) \rightarrow \begin{bmatrix} 1/2\,[1 + H_n(m, :)] \\ 1/2\,[1 - H_n(m, :)] \end{bmatrix}. \tag{26}$$

If we denote a row-doubled version of H_n as $\hat{H}_n$, then

$$T_x,\ \widetilde{T}_{x''} \in \{\hat{H}_n(:, m)\}, \qquad m \in \mathbb{Z}^*;\ m \le n \tag{27}$$

is a physically realizable orthogonal column code when it is combined with the now-possible computational step of weighting the appropriate rows in the measurement by -1. An aperture based on a row-doubled Hadamard matrix is shown in Fig. 6.

D. Continuous Versus Discrete Codes

There is an important difference between the continuous mask codes (harmonic and Legendre) and the discrete codes (S matrix and row-doubled Hadamard). In the case of the continuous code families, there is an infinite number of possible codes ($m \in \mathbb{Z}^*$). This means that the underlying Hilbert space is infinite dimensional. Any physical aperture based on these codes must choose only a subset of the possible code patterns. As a result, the implemented basis is not complete, and Parseval's relation will not hold. In short, in the presence of noise the total power associated with the different apertures after processing will not necessarily equal the total power measured on the detector plane.

For the discrete codes, however, there is only a finite number of code patterns in any given family ($m \le n$). The underlying Hilbert space is then n dimensional, and an aperture can be designed that contains all of the codes. In this case Parseval's relation will hold, and power is necessarily conserved during the processing.

4. Implementation Issues

There are a variety of implementation issues in which the performance of the real-world system must deviate from the idealizations we consider above. In this section we address the most important of these issues.

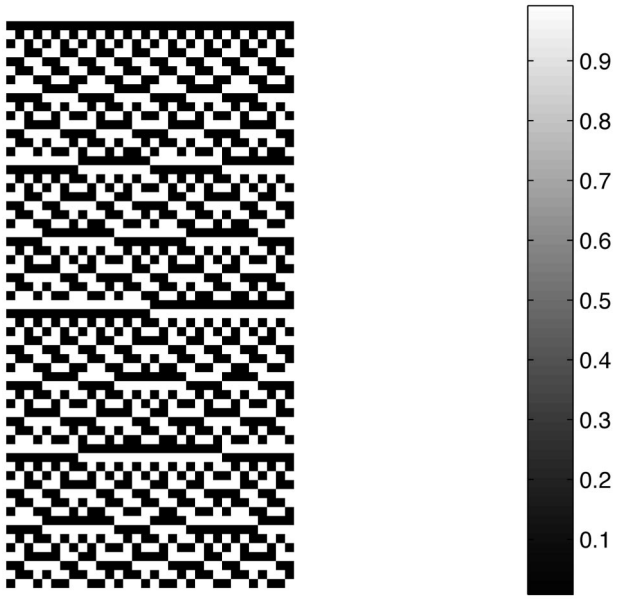

Fig. 6. Aperture pattern for an orthogonal column code (in conjunction with processing of the measured intensity) based on a row-doubled Hadamard matrix. The codes were chosen such that the transmission has physically realizable values in the interval [0, 1]. Note that the pattern is discrete both horizontally and vertically.

A. Pixelization of the Detector Plane

In Section 2 we assume that we have access to the detector-plane intensity distribution $I(x', y')$. In reality, we do not. Our measurement of the intensity profile is downsampled by the pixel size on the detector array. This has several important implications for the system. First, for the continuous codes, Eq. (14) is no longer strictly true. It remains approximately true, however, as long as we include only codes that contain spatial frequencies below the Nyquist limit defined by the pixel size.

Second, for the discrete codes, the aperture must be designed so that when imaged onto the detector, the features involve integral numbers of pixels in the y' direction. This places performance requirements on the manufacturing accuracy of the aperture and on the magnification of the imaging optics in the spectrometer. Additionally, an aperture involving a discrete code must be aligned with respect to the detector plane such that the vertical divisions between features align with divisions between pixels. This requires subpixel positioning capability on the input aperture during construction and alignment.

B. Manufacturing Gray-Scale Patterns

Physical realities in Section 3 limit us to coding patterns with values in the interval [0, 1]. However, the fact that a given modulation pattern can be physically imprinted

on the input intensity has no bearing on the manufacturability of the required input aperture.

Arbitrarily patterned, continuous-tone masks with transmissions ranging from 0% to 100% are indeed possible. However, given the complexity of most orthogonal column code patterns, the cost to manufacture transmission masks to the required precision is prohibitive. One alternative is to convert the designed continuous-tone mask into a half-toned version. A small region of the continuous-tone pattern is subdivided into an array of even smaller subregions. Each of these subregions is assigned a transmission of either 0% or 100%, such that the net transmission in the region matches the gray-scale value of the continuous-tone pattern. Provided that the conversion happens on a spatial scale that is smaller than the pixelization of the detector plane, no significant difference should be detectable.

There are a variety of half-toning algorithms available for optimizing the conversion. We have recently acquired a half-toned version of our harmonic mask pattern described above and have begun testing.

C. Optical Distortions and Corrections

The internal optics of the spectrometer can have a significant effect on the performance of the system. The optical properties of a static MMS deviate from a traditional instrument in a critical manner. Because the MMS encodes spectral information across the detector plane in a highly nonlocal way, optical errors anywhere have a nonlocal effect on the reconstruction, introducing noise and errors at regions throughout the spectral range.

What are the primary optical errors we must worry about? We assume above that the incoherent imaging kernel is given by $H(x, x', y, y'; \lambda) = \delta(y - y')\delta\{x - [x' + \alpha(\lambda - \lambda_c)]\}$. Significant deviation from this assumption leads to degraded (or erroneous) spectral reconstructions. Thus we have three primary optical requirements:

1. The spectral resolution of the instrument should be limited by the width of a feature on the input mask Δx. This requires that the size of the incoherent impulse response be small compared with Δx. Further, the size of the impulse response should not vary significantly across the input and output fields.
2. The impulse responses in the x and y directions should be uncorrelated. This requires that the optical system have low distortion across the input and output fields.
3. The input intensity profile should be unaffected by propagation through the system (aside from a wavelength-dependent shift in the x direction). This requires that there be no field-dependent intensity modulations (vignetting) in the system.

There is an additional way that our ideal imaging kernel can break down. Unfortunately, this issue exists even for an ideal optical system and must be dealt with either through special modifications to the hardware or through software corrections of the detector image prior to spectral reconstruction.

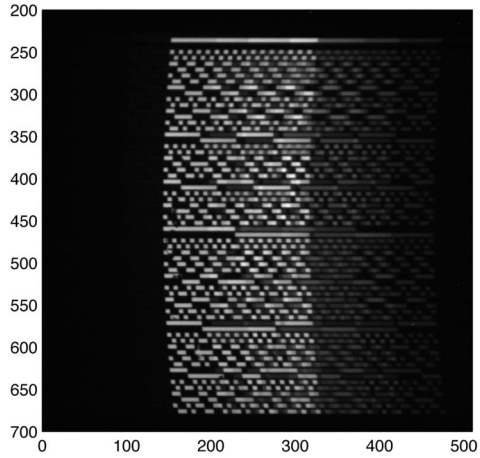

Fig. 7. Raw intensity image captured at the focal plane. The smile distortion is clearly visible. The spectral source has only sharp spectral lines, so the image contains only a few, crisp images of the mask pattern.

It is well known that imaging an aperture through a diffraction grating results in an image that is curved in the direction of the dispersion.[21] In terms of our imaging kernel, this manifests as a λ_c that is y dependent. This curvature, which is sometimes referred to as smile distortion, is the result of the particular geometry of the wave-normal sphere. For high-$F/\#$ systems, the curvature is minimal and can be ignored. However, since we are concerned with maximizing étendue, a static MMS is almost always constructed at a low $F/\#$. As a result, the curvature is significant, as can be seen in Fig. 7. This curvature can be corrected in two possible ways: either the input aperture can be predistorted to compensate for the applied curvature, or the collected image can be computationally processed to straighten the patterns prior to reconstruction. We choose the latter approach. The smile distortion is a simple quadratic. During the calibration phase, we fit the leading edge of an aperture image to a parabola and determine the required circular shift to apply to each row of the image to straighten the patterns. The result of the procedure is shown in Fig. 8.

5. Experimental Results

Over the past eighteen months, we have designed and constructed more than fifteen different spectrometers based on the ideas and codes described above (specifically, row-doubled implementations of the Hadamard masks described in Subsection 3.C). The different instruments have been used for Raman, fluorescence, and absorption spectroscopy; have spanned the spectral range from UV to the near-IR; have demonstrated both reflective and transmissive geometries; and have achieved spectral resolutions in the range $\Delta\lambda \approx 0.1-3$ nm. The performance of the instruments has invariably been excellent, signifi-

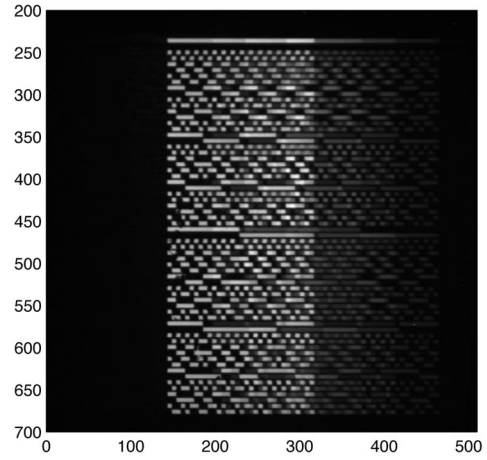

Fig. 8. Corrected intensity image after smile distortion was removed via software processing. The leftmost edge of the sharp mask image in Fig. 7 was fit to a parabola to determine the amount of shift to be applied to each row of the image.

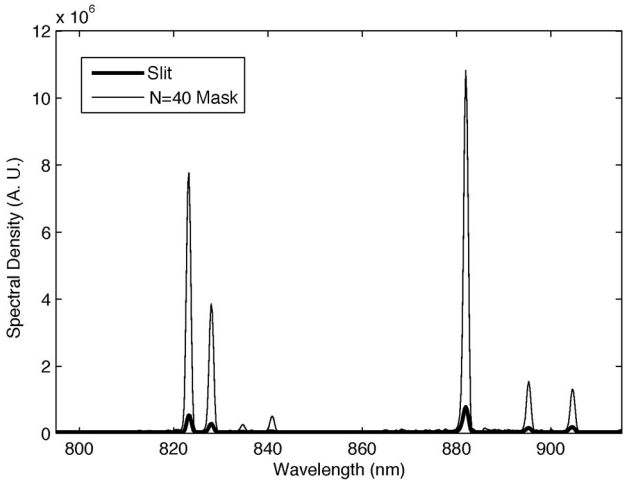

Fig. 9. Comparison between the reconstructed spectrum from a mask based on $\hat{H}_{40}$ and one based on a slit aperture. The mask aperture clearly captures significantly more light, while maintaining an equivalent spectral resolution.

cantly outperforming traditional spectrometers on diffuse sources.

This section presents a series of experimental results collected on one of the static MMS systems. The primary goal is to clearly demonstrate the existence of the Jacquinot and Fellgett advantages and to show that the performance scales as expected.

In all the experiments below, the spectral source for the experiments was a xenon discharge lamp operated in conjunction with a diffuser. The light from the diffuser was allowed to fall directly on the mask aperture—no relay optics of any kind were used. Unless otherwise noted, the CCD integration time was 160 ms. The particular spectrometer has a spectral range of $\Delta\Lambda \approx 775$–$900$ nm. The spectral resolution depends on the mask, and for the majority of the masks used, it was $\delta\lambda \approx 0.65$ nm. The masks consisted of chrome deposited on a quartz substrate. The smallest mask feature was 36 μm, corresponding to 4 pixels on the CCD.

Figure 9 compares the spectrum reconstructed by using the mask generated from $\hat{H}_{40}$ and from a slit with a width (36 μm) equal to the feature size of the mask. Clearly, the coded aperture collects significantly more light without sacrificing spectral resolution.

We tested row-doubled Hadamard masks, $\hat{H}_n$, of a variety of orders ($n = 40, 32, 24, 16, 12$). In Fig. 10 we plot the results from the different masks. The signal strength increases as the mask order increases, as we would expect. However, determining the throughput advantage is complicated by the fact that as the mask order increases, there is an increase not only in the number of openings on a given row of the mask but also in the number of mask rows. To check the throughput scaling, we normalize the total counts collected for a given mask by dividing by the total counts collected with a slit that occupies an equal number of rows on the CCD. In a row-doubled Hadamard mask, there are $N/2$ openings on any row. As such, we would expect the normalized counts to also scale by this amount. The results are plotted in Fig. 11.

We see that the observed scaling is approximately $N/4$, rather than the expected $N/2$. We believe the discrepancy can be attributed to the optical system in the spectrometer. Because the reduction in light collection is a constant factor of ≈ 2 regardless of mask size, we can rule out vignetting as the cause. Rather, we believe the effect arises from the modulation transfer function (MTF) of the optics. In the horizontal (x) direction, the Hadamard masks and the slit have the same range of spatial frequencies. In the vertical (y) direction, however, the slit contains only a dc component, while the masks contain high spatial frequencies from the row doubling. Experimentally,

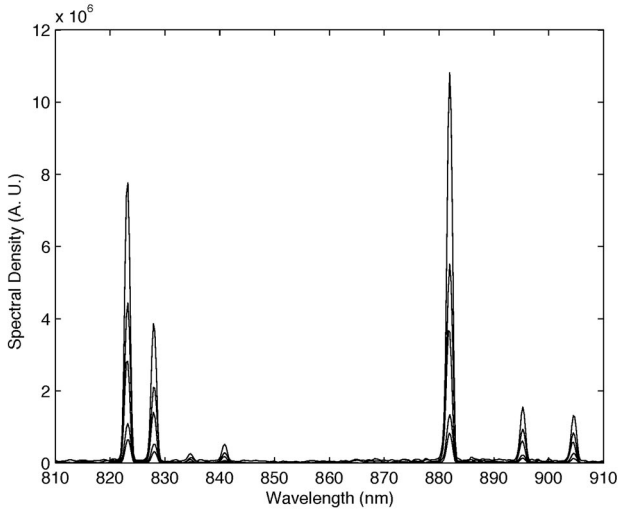

Fig. 10. Comparison of reconstructed spectra from row-doubled Hadamard masks of various orders ($n = 40, 32, 24, 16, 12$). The system throughput increases as the mask order increases, as expected. The codes increase the system throughput without affecting spectral resolution.

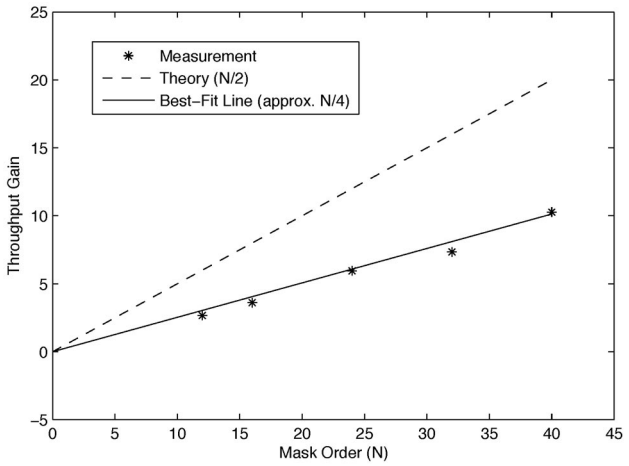

Fig. 11. Throughput gain achieved by order-N masks compared to slits of equivalent height. Theoretically, the gain should scale as $N/2$. We observe approximately $N/4$. For reasons discussed in the text, we attribute this discrepancy to the MTF of the optical system. Note that even with this discrepancy, a mask of moderate complexity is capable of increasing throughput by an order of magnitude without sacrificing spectral resolution.

when we compare the counts on a single row of the CCD between the mask and the slit, we observe a ratio of approximately $N/4$ as measured for the entire pattern. If we instead compare the counts on a row between the mask and a square pinhole, we observe a ratio of approximately $N/2$ as theory would predict. Thus we conclude that the discrepancy is related to the MTF of the optical system.

Finally, we attempt to quantify the improvement in the SNR that accompanies the increase in throughput. Figure 12 shows a region of the xenon spectrum containing a very small peak (so weak that it is not visible at the scales of the previous figures). The top graph of Fig. 12 shows the peak as reconstructed by the row-doubled, order-40 Hadamard mask. The bottom plot is the peak as measured by the slit aperture. If we define the SNR of the peak to be its height divided by the RMS value of the region near the peak, we find that the SNR for the mask aperture is $\approx$23.7, while the SNR for the slit is $\approx$7.0. This is a SNR gain of $23.7/7.0 \approx 3.4$. From Fig. 11, we see that the mask provided a throughput advantage of $\approx$10.3. For a shot-noise process, we would expect this throughput gain to result in an SNR gain of $\sqrt{10.3} \approx 3.2$, which is indeed close to the observed value.

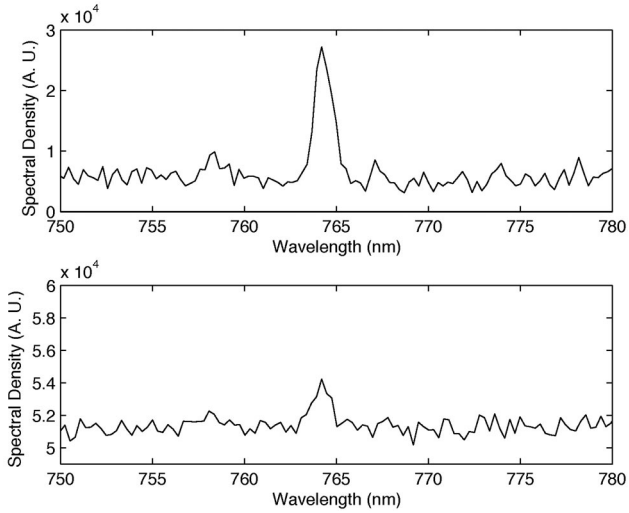

Fig. 12. Comparison of a small spectral peak as reconstructed by a mask based on $\hat{H}_{40}$ and a slit. The SNR gain with the mask is consistent with the measured throughput gain.

The capability to dramatically increase the throughput of the system clearly demonstrates the Jacquinot advantage of the system. The presence of the Fellgett advantage is also easily deduced. By having multiple openings on a row of the input mask, any detector pixel sees a combination of spectral channels. As such, the signal level on the pixel is increased over what it would be in the absence of multiplexing. Since the additive detector noise remains constant, the SNR on the pixel must increase. Thus our system also exhibits the Fellgett advantage.

6. Summary

In this paper we have described a new class of coded-aperture spectrometer, which we refer to as a static MMS. A static MMS is optimized for working with diffuse sources and achieves both the Jacquinot and the Fellgett advantages in a single shot. Experimental results support these claims. The result is an inexpensive, mechanically robust, high-performance spectrometer.

The required aperture patterns for a static MMS are based on orthogonal function theory and can be generated for a variety of schemes. We have presented several specific families of particular interest.

In future publications we will discuss the special imaging properties of spectrometers based on these apertures, as well as the application of static MMS to chemometric studies of biological systems.

This research was supported by a grant from the National Institutes of Health and the National Institute on Alcohol Abuse and Alcoholism.

References

1. D. J. Brady, "Multiplex sensors and the constant radiance theorem," Opt. Lett. **27,** 16–18 (2002).
2. P. Jacquinot, "New developments in interference spectroscopy," Rep. Prog. Phys. **23,** 267–312 (1960).
3. P. B. Fellgett, "The multiplex advantage," Ph.D. dissertation (University of Cambridge, Cambridge, UK, 1951).
4. M. J. E. Golay, "Multislit spectrometry," J. Opt. Soc. Am. **39,** 437–444 (1949).
5. M. J. E. Golay, "Static multislit spectrometry and its application to the panoramic display of infrared spectra," J. Opt. Soc. Am. **41,** 468–472 (1951).
6. A. Girard, "Spectrometre a Grilles," Appl. Opt. **2,** 79–87 (1963).
7. A. S. Hedayat, N. J. A. Sloane, and J. Stufken, *Orthogonal Arrays: Theory and Applications* (Springer-Verlag, 1999).
8. R. N. Ibbett, D. Aspinall, and J. F. Grainger, "Real-time multiplexing of dispersed spectra in any wavelength region," Appl. Opt. **7,** 1089– 1093 (1968).

9. J. A. Decker and M. O. Harwitt, "Sequential encoding with multislit spectrometers," Appl. Opt. **7,** 2205–2209 (1968).
10. J. A. Decker, "Experimental realization of the multiplex advantage with a Hadamard-transform spectrometer," Appl. Opt. **10,** 510–514 (1971).
11. P. Hansen and J. Strong, "High resolution Hadamard transform spectrometer," Appl. Opt. **11,** 502–506 (1972).
12. P. G. Phillips and D. A. Briotta, "Hadamard-transform spectrometry of the atmopheres of Earth and Jupiter," Appl. Opt. **13,** 2233–2235 (1974).
13. R. D. Swift, R. B. Wattson, J. A. Decker, R. Paganetti, and M. O. Harwitt, "Hadamard transform imager and imaging spectrometer," Appl. Opt. **15,** 1595–1609 (1976).
14. M. O. Harwitt and N. J. A. Sloane, *Hadamard Transform Optics* (Academic, 1979).
15. S. B. Mende, E. S. Claflin, R. L. Rairden, and G. R. Swenson, "Hadamard spectroscopy with a two-dimensional detecting array," Appl. Opt. **32,** 7095–7105 (1993).
16. R. Riesenberg and U. Dillner, "HADAMARD imaging spectrometer with micro slit matrix," in *Imaging Spectrometry V*, M.R. Descour and S.S. Shen, eds., Proc. SPIE **3753,** 203–213 (1999).
17. R. A. de Verse, R. M. Hammaker, and W. G. Fateley, "Realization of the Hadamard multiplex advantage using a programmable optical mask in a dispersive flat-field near-infrared spectrometer," Appl. Spectrosc. **54,** 1751–1758 (2000).
18. R. Riesenberg, G. Nitzsche, and W. Voigt, "HADAMARD encoding and other optical multiplexing," VDI Ber. **1694,** 345–350 (2002).
19. P. Jacquinot, "How the search for a throughput advantage led to Fourier transform spectroscopy," Infrared Phys. **2–3,** 99–101 (1984).
20. G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists,* 6th ed. (Academic, 2005).
21. D. J. Schroeder, *Astronomical Optics* (Academic, 1987).