

Spectrum Estimation from Quantum-Limited Interferograms

Daniel R. Fuhrmann, Chrysanthe Preza, Joseph A. O'Sullivan
Donald L. Snyder, William H. Smith

Abstract—A quantitative model for interferogram data collected in a quantum-limited hyperspectral imaging system is derived. This model accounts for the geometry of the interferometer, the Poisson noise, and the parameterization of the mean of the noise in terms of the autocorrelation function of the incident optical signal. The Cramer-Rao bound on the variance of unbiased spectrum estimates is derived and provides an explanation for what is often called the “multiplex disadvantage” in interferometer-based methods. Three spectrum estimation algorithms are studied: maximum likelihood via the expectation-maximization (EM) algorithm, least squares (LS), and the fast Fourier transform (FFT) with data pre-correction. Extensive simulation results reveal advantages and disadvantages with all three methods in different signal-to-noise ratio regimes.

Index Terms—Spectrum Estimation, Interferometry, Interferograms, EM Algorithm, Cramer-Rao Bound, Multiplex Disadvantage.

I. Introduction

Hyperspectral imaging (HSI), or imaging spectrometry, is an emerging technology which is showing significant promise for applications in remote sensing, surveillance, and astronomical imaging. In HSI, a sensor acquires an image of a scene, normally in the visible or infrared regions of the optical spectrum, in which each pixel is resolved into many, perhaps hundreds, of spectral bands. This fine spectral resolution, it is anticipated, will create new opportunities for scene analysis, target recognition, and anomaly detection which would not be possible in conventional monochromatic or trichromatic imaging systems. The growth of this technology is rekindling an interest in the classical problems of spectrum estimation as well.

In many systems, the resolution of each pixel in the spectral dimension is accomplished by dispersing light over the focal plane either by a diffraction grating or

by a prism. The signal processing required for spectrum estimation is minimal, and consequently these systems do not concern us here. There is a second class of systems which use interferometric techniques [1], [2], [3] to acquire a signal called an interferogram, which is nominally a sampled version of the autocorrelation function of the incident optical signal. Typically one obtains the desired spectrum by computing the discrete Fourier transform of the interferogram. We feel that this method, which computationally efficient, is suboptimal from a statistical signal processing perspective and that other methods should be considered. The subject of this paper is the statistical estimation of the power spectral density from the interferogram, with particular emphasis on low light-level conditions.

The subject of spectrum estimation is a familiar one to many in the signal processing community. In early work on the subject, the problem was often posed as the mathematical one of identifying a positive spectrum which is consistent with a finite known set of autocorrelation lags. In his original work, Burg [4] considered the problem of finding an extension of the truncated autocorrelation which maximized the spectral entropy subject to constraints given by the known lags. While this application of the maximum entropy formalism was a breakthrough philosophically, the technique was at first widely misapplied when it was used on sample autocorrelation lags formed from time series data. Subsequent research led to a wide variety of techniques for spectrum estimation from time series data, many of which are summarized in [5]. In the present work we return to the problem of spectrum estimation from autocorrelation data - but in this case a noisy, truncated, and distorted version of the autocorrelation called the interferogram.

This problem was considered by Neira and Constantinides [6], [7] who addressed it using the tools of frame theory and maximum entropy. They observed that if the autocorrelation is oversampled, then a smooth and more accurate version of the autocorrelation can be recovered from the noisy samples. This approach does not treat the problem directly as a statistical estimation problem.

Bialkowski [8] approached the problem by observing that the desired spectrum is a linear transformation (the Fourier transform) of the true autocorrelation, and that the acquired autocorrelation data are subject to Poisson statistics. Under this model, he set up the appropriate maximum-likelihood estimation problem and solved it

This work has been submitted to the IEEE Transactions for Signal Processing for possible publication. Copyright may be transferred without notice, after which this version will be superseded. D. R. Fuhrmann is with the Department of Electrical Engineering, Washington University, St. Louis, MO 63130. E-mail: danf@ee.wustl.edu. C. Preza is with the Department of Electrical Engineering, Washington University, St. Louis, MO 63130. E-mail: preza@ee.wustl.edu. J. A. O'Sullivan is with the Department of Electrical Engineering, Washington University, St. Louis, MO 63130. E-mail: jao@ee.wustl.edu. D. L. Snyder is with the Department of Electrical Engineering, Washington University, St. Louis, MO 63130. E-mail: dls@ee.wustl.edu. W. H. Smith is with the Department of Earth and Planetary Sciences, Washington University, St. Louis, MO 63130. E-mail: whsmith@levee.wustl.edu.

numerically using the Expectation-Maximization (EM) algorithm [9], [10], [11]. The motivation in this work is interesting, and relevant to hyperspectral imaging. Bialkowski was attempting to overcome a phenomenon known as the “multiplex disadvantage” [12], in which the effect of sensor noise is seen across all spectral bins, rather than just those in which there is signal energy, as would occur with direct spectroscopic measurements. The concern is that multiplicative noise associated with a single strong source could mask a weak signal at a separate frequency, a concern which we share as well.

In this paper, we extend the statistical estimation approach of Bialkowski in several ways. First, we derive a more realistic physics-based model for the relationship between the desired spectrum and the acquired data. This model includes the effects of time-warping and apodization of the interferogram, as well as possible misalignment effects in the sensor. The parameters of the resulting sensor response model can be determined empirically. Using this model, we determine a performance bound on any unbiased estimate of the spectrum. Analysis of this performance bound confirms, and provides insight into, the difficulties of performing wideband spectrum estimation in the low-SNR regime using interferometric techniques. Finally, three spectrum estimation methods are derived: 1) maximum-likelihood via the EM algorithm; 2) least squares; and 3) fast Fourier transform (FFT) with data pre-correction. Results obtained with these methods as well as trade-offs in performance and computational complexity are presented.

II. Sensor Model

Jacquinet [13] reviewed in detail the interrelation between light handling capability, spectral resolving power, and the physical nature of specific spectroscopic devices. While these considerations are important for the design and analysis of modern hyperspectral sensors, there is an additional impact due to the multiplex and/or multi-channel signal encoding properties. We study this impact for the specific case of low flux quantum noise-limited illumination in a two-beam interferometer.

A. Physical Model

The sensor model we adopt in this paper is that of a simple two-beam interferometer. While simple and intuitive, this model captures many of the important features of more complex real-life sensors such as the Digital Array Scanned Interferometer (DASI) [2], [14] shown in Figure 1. Our purpose is to show how the effects of time-domain warping, apodization, and Poisson noise can all be accommodated in the proposed spectrum estimation method. Although this model has been presented also in [15] it is included here for completeness.

The interferometer model we consider is shown in Figure 2 corresponding to the classical Young’s experiment. All of the arguments below are based on geometric ray tracing (no diffraction). The optical wavefront is

incident on a shutter in which there are two narrow slits, separated by distance d . Light passes through the slits and illuminates the focal plane at a distance f . The light from the two paths recombines at the focal plane to form an interference pattern.

Let the incident field at the two apertures be $E(t)$. Ignoring polarization, $E(t)$ will be treated as a scalar field. It is modeled as a wide-sense-stationary stochastic process with power spectral density $S(f)$ and corresponding autocorrelation function $R(\tau)$. The incident field at position x along the focal plane axis is proportional to $E(t - \tau_1(x)) + E(t - \tau_2(x))$, where $\tau_i(x) = \frac{r_i(x)}{c}$, $r_i(x)$ are the optical path lengths in Figure 2, and c is the speed of light. The intensity of the light at position x , is the average squared value of the field, given by

$$I(x) = \langle |E(t - \tau_1(x)) + E(t - \tau_2(x))|^2 \rangle. \quad (1)$$

The intensity in (1) can be written in terms of the autocorrelation function as

$$I(x) = 2R(0) + 2R(\tau(x)), \quad (2)$$

where

$$\tau(x) = \tau_2(x) - \tau_1(x). \quad (3)$$

The two distances $r_1(x)$ and $r_2(x)$ are given by

$$r_1(x) = \left[f^2 + \left(x + \frac{d}{2} \right)^2 \right]^{1/2} \quad (4)$$

$$r_2(x) = \left[f^2 + \left(x - \frac{d}{2} \right)^2 \right]^{1/2}. \quad (5)$$

When the resulting expression for $\tau(x)$ is expressed as a first-order Taylor series in x in the neighborhood of $x = 0$ we get

$$\tau(x) \approx \frac{d}{2f} \frac{x}{c}. \quad (6)$$

However, this linear approximation is not good over the entire range of x . The nonlinear relationship between x and $\tau(x)$ must be accounted for in the spectrum estimation.

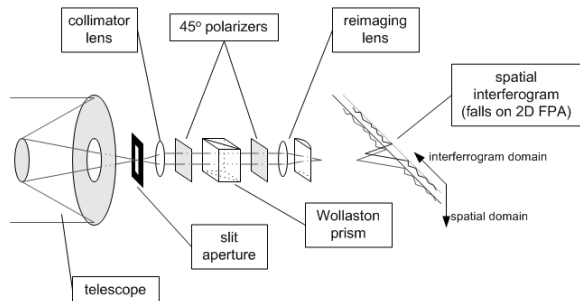


Fig. 1. Schematic of a birefringent DASI sensor configuration. Light passing through a slit is sheared into two components by the Wollaston prism which are forced to interfere by the second polarizer. The interference pattern or interferogram can be detected by a two-dimensional focal plane array (FPA).

If a charge-coupled device (CCD) array is placed along the focal-plane axis (x -axis), then each element has a response which is proportional to the intensity of the incident field. In this way, the sensor acquires an image of the autocorrelation function, added to an $R(0)$ pedestal as indicated by (2). The data acquired by the CCD elements and their associated instrumentation form what is called the interferogram.

A model which takes into account imperfect alignment of various parts of the sensor is also possible, as shown in Figure 3. In this model, the direction-of-arrival of the incident field makes an angle ϕ with the vertical, and the focal plane makes an angle θ with the horizontal. Furthermore, the CCD array is not centered beneath the midpoint of the two slits. While the first two effects are undesirable, the third may be deliberate in that it allows for acquisition of autocorrelations at larger $\tau(x)$ than is possible in the symmetric sensor. The equations for time delay at position x are:

$$\tau(x) = \frac{r_2(x) - r_1(x)}{c} - \tau_0 \quad (7)$$

$$\tau_0 = \frac{d}{c} \sin \phi \quad (8)$$

$$r_1(x) = \left[(f - x \sin \theta)^2 + \left(x \cos \theta + \frac{d}{2} \right)^2 \right]^{1/2} \quad (9)$$

$$r_2(x) = \left[(f - x \sin \theta)^2 + \left(x \cos \theta - \frac{d}{2} \right)^2 \right]^{1/2} \quad (10)$$

Another important effect which must be included in our model is the attenuation in intensity which occurs naturally as a result of the angle which the ray path makes with the normal to the CCD element (the obliquity). This attenuation of $R(\tau)$ with increasing τ is called apodization. From a signal processing perspective, this is equivalent to windowing except that it occurs naturally rather than being introduced within the signal processing.

To derive the apodization function for our simplified model, we assume that the illumination from the two slits

is distributed uniformly across the angle in the plane of the diagram in Figure 2, and then we compute the angle subtended by each pixel at the midpoint M between the two slits. Define Ψ as the angle formed by the ray from M to the pixel at location x with the normal ray from M to the CCD line; see Figure 4. The angle subtended by pixel x is $\Delta\Psi$, which is well-approximated by $\frac{d\Psi}{dx} \Delta x$ where Δx is the width of one pixel, assumed constant. In our model for the sensor with imperfect component alignment, we have

$$\Psi = \tan^{-1} \left(\frac{x - f \sin \theta}{f \cos \theta} \right) \quad (11)$$

and hence

$$\frac{d\Psi}{dx} = \frac{1}{1 + \left(\frac{x - f \sin \theta}{f \cos \theta} \right)^2} \frac{1}{f \cos \theta} \quad (12)$$

If we consider this a space-dependent gain $g(x)$ and normalize it so that it has maximum value 1, then we arrive at the apodization function

$$g(x) = \frac{1}{1 + \left(\frac{x - f \sin \theta}{f \cos \theta} \right)^2} \quad (13)$$

Note that in this model, the peak of the apodization function ($x = f \sin \theta$) does not necessarily coincide with the time center of the measured autocorrelation function ($x = 0$).

Summarizing the sensor model to this point, we have that the light intensity at position x is given by

$$I(x) = Cg(x)[R(0) + R(\tau(x))] \quad (14)$$

$$= Cg(x) \int S(f)(1 + \cos(2\pi f\tau(x)))df \quad (15)$$

where $\tau(x)$ is determined from the time-delay equations, $g(x)$ is the apodization term, and C is some constant. Note the linear relationship between the desired spectrum $S(f)$ and the intensity $I(x)$.

As stated in the beginning of this section, this model captures most of the key features of an interferometric sensor which we need for illustrating the spectrum estimation method. However, there is one important feature

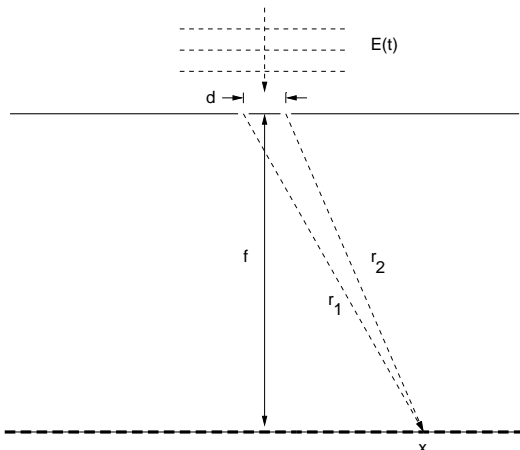


Fig. 2. Two-slit interferometer

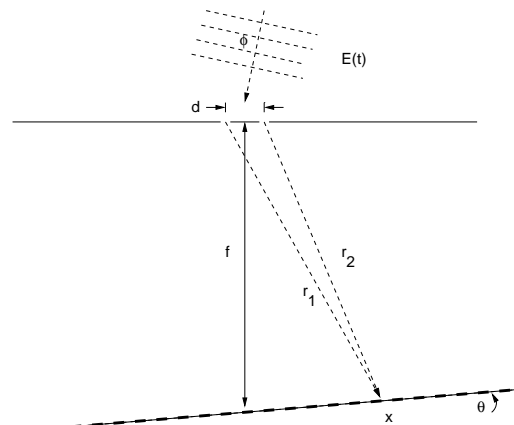


Fig. 3. Interferometer with imperfect alignment.

of the birefringent form DASI [2] worth mentioning at this point. In the DASI (Figure 1), the two light paths are created by a single slit aperture, a Wollaston prism, and an appropriate arrangement of polarizing filters. With this configuration, it is simple to create an interference pattern of the form $R(0) - R(\tau)$ instead of $R(0) + R(\tau)$ by a 90-degree rotation of one of the polarizers [14]. When both data sets are collected (separate measurements) one can obtain estimates of $g(x)R(0)$ and $g(x)R(\tau(x))$ by simple addition and subtraction. In our simulation example of Section II-D, it will be assumed that both measurements are available, although in two of the three spectrum estimation methods the two data sets remain distinct.

B. Sensor Calibration

The performance bounds and spectrum estimation methods to be proposed in later sections depend on accurate knowledge of the sensor parameters and knowledge of the time-delay and apodization curves. Detailed methods for calibration of these functions is beyond the scope of this paper, but are discussed in [16], [15]. In brief, the apodization curve can be autocalibrated using simple polynomial curve-fitting techniques. Estimation of the time-delay curve requires calibration sources, one wideband and one narrowband. The wideband source is used to determine the center of the interferogram, whereas the narrowband source is used to produce an interferogram which takes the form of a phase-modulated sinusoid. The phase modulation can be deduced by discrete-time quadrature demodulation, and this in turn determines the time delay. In [15] it is assumed that the sources have adequate SNR so that statistical estimation methods are not required for calibration.

C. Statistical Measurement Model

According to the physical model above, there is a linear relationship between the intensity $I(x)$ and the desired spectrum $S(f)$. Assume now that there are M CCD elements at locations $x_1 \dots x_M$. This arrangement discretizes the intensity $I(x)$ into M distinct values $I_m = I(x_m)$. If we further discretize the spectrum $S(f)$ into N spectral components at frequencies $f_1 \dots f_N$, we have the linear

equation

$$I_m = \sum_{n=1}^N a_{mn} s_n \quad (16)$$

where the a_{mn} are elements of an $M \times N$ matrix \mathbf{A} incorporating all the effects described in the model above and $\mathbf{s} = [s_1 \dots s_N]^T$ is the desired spectral vector. It is assumed that \mathbf{A} is known from modeling and calibration. When in-phase and out-of phase measurements are available, then there are two $M \times N$ matrices describing the sensor effects which can be concatenated to form a single $2M \times N$ matrix, which we still call \mathbf{A} .

Under low light-level conditions, we model the outputs of the individual CCD elements as Poisson random variables $K_1 \dots K_M$ with means proportional to intensity, i.e., $\lambda_m = \alpha I_m$ where α is a constant which depends on physical parameters of the CCD element and the integration time. The probability mass function for the data is

$$P(K; \mathbf{s}) = \prod_{i=1}^M e^{-\lambda_i} \frac{\lambda_i^{k_i}}{k_i!} . \quad (17)$$

Without loss of generality, we can assume that $\alpha = 1$ and incorporate its effect into the matrix \mathbf{A} . Thus, \mathbf{s} and λ are related through the $M \times N$ system of equations

$$\lambda = \mathbf{A} \mathbf{s} ,$$

where $\lambda = [\lambda_1 \dots \lambda_M]^T$.

The statistical estimation problem which now arises naturally for our model is the following: given the data $K_1 \dots K_M$ and the linear intensity model (16), estimate the spectral values $s_1 \dots s_N$. In brief, we seek to estimate a nonnegative function $S(f)$ from Poisson data where the rates are linearly related to $S(f)$ via a nonnegative kernel. Put this way, the problem bears a close resemblance to similar problems in astronomical imaging [17] and positron emission tomography (PET) [10].

D. Simulation Example

We now present an example of the sensor modeled in the previous sections, and two examples of the data which might be acquired by such a sensor.

Using the model for the two-slit interferometer with imperfect alignment, the following parameters are chosen: the distance between the two slits is $d = 0.1$ mm, the focal length is $f = 2.0$ cm, and the tilt angles are $\theta = 5^\circ$ and $\phi = 10^\circ$. The CCD array consists of 1280 elements spaced uniformly along a line in the focal plane from $x = -15$ mm to $x = +25$ mm. In Figure 5, there are two plots. The first shows the nonlinear relationship between x and the differential time delay $\tau(x)$; the second shows the apodization function $g(x)$.

The first example is the simulated optical power spectral density for a true extraterrestrial solar power spectrum covering the range 400-800 THz shown in Figure 6. The extraterrestrial solar spectrum over a wide range of wavelengths, in W/cm^{-1} , was obtained from a publicly available source [18]. This spectrum was converted to a

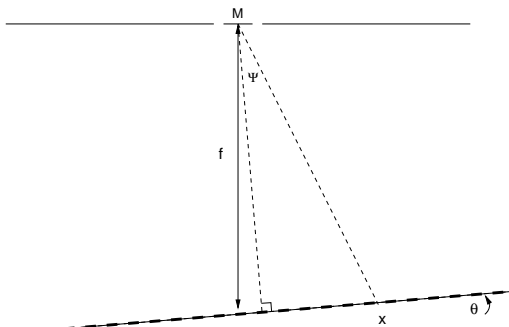


Fig. 4. Geometry for modeling apodization

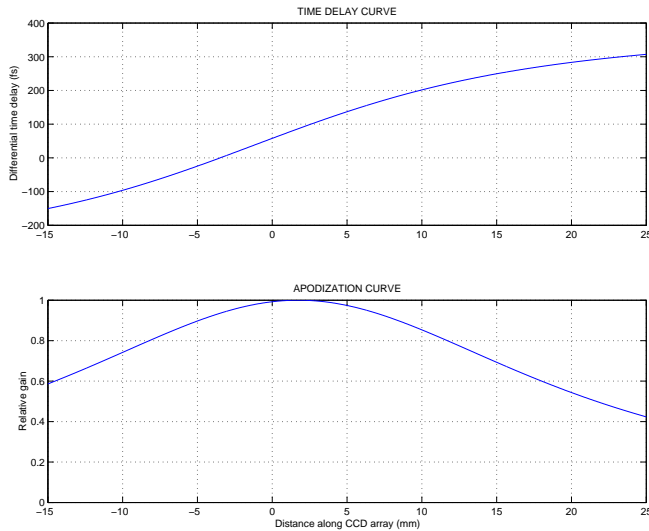


Fig. 5. Simulated time-delay and apodization curves.

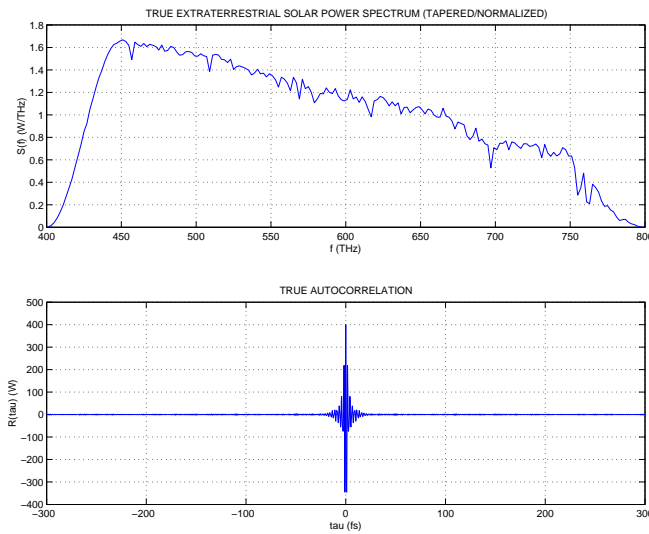


Fig. 6. Simulated optical solar spectrum and autocorrelation function.

power spectral density with units of W/THz, then sampled at 2 THz spacing from 400 THz to 800 THz, yielding 200 spectral values covering roughly the visible optical spectrum. The spectrum was windowed using a cosine taper over the ranges 400-450 THz and 750-800 THz to make it go smoothly to 0 at the edges. The chosen resolution value is consistent with the maximum time delay $\tau_{max} = 300$ fsec which the instrument can measure, and is also as good or better than the spectral resolution of several published specifications for hyperspectral sensors in the visible range [19]. This value was chosen for our simulation to avoid any issues of ill-posedness in the inverse problem: too fine a discretization in frequency will cause the \mathbf{A} matrix in the previous section to be rank-deficient, which raises a host of problems regarding the uniqueness of spectrum estimates that we do not wish to address here. The true autocorrelation function of the spectrum, found by computing the inverse Fourier

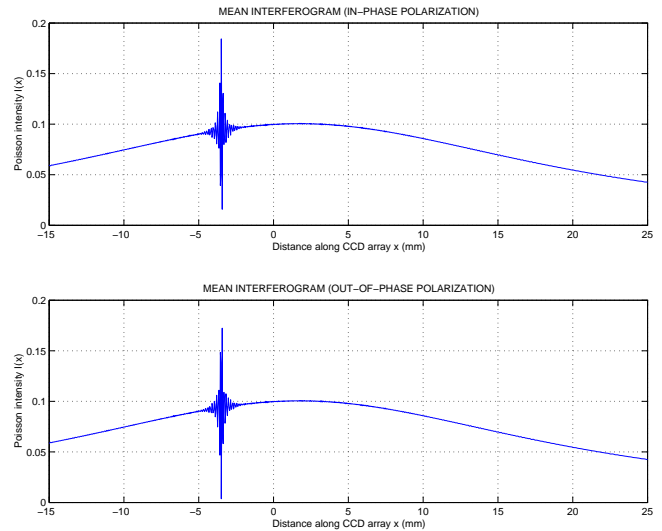


Fig. 7. Simulated optical intensity for the solar spectrum corresponding to the in-phase (top panel) and out-phase (bottom panel) configuration of the sensor.

transform of the solar spectrum on a very fine grid in time (τ), is shown over the range $\tau = -300$ to 300 fsec in Figure 6.

The simulated interferogram predicted by our model was computed for both in-phase and out-of-phase measurements. The simulated interferometer has 1280 elements uniformly spaced from -15 mm to +25 mm. As a result, the \mathbf{A} matrix, relating observed measurements to the spectral values, has dimension 2560×200 . Measurements were simulated by first computing $\mathbf{A}\mathbf{s}$ and then using the `poissrnd` function in MATLAB to generate Poisson data with mean $c\mathbf{A}\mathbf{s}$ where c is a constant that controls the overall signal-to-noise (SNR) defined as the average number of counts per spectral bin. In other words, $\text{SNR} = \sum(c\mathbf{A}\mathbf{s})/200$, where 200 is the number of spectral bins.

The intensity $I(x)$ of the simulated interferogram predicted by our model is shown in Figure 7 for both the in-phase measurement, and the out-of-phase polarization measurement. Note the asymmetry in this intensity and the warping of the delay axis for large x . Shown in Figure 8 are typical Poisson-distributed data generated according to the intensity in Figure 7.

Using the same simulation parameters, a second example of a predicted interferogram for a simulated optical power spectral density of a line spectrum is presented. The line spectrum contains two equal power Gaussian functions centered at 495 and 505 THz with equal standard deviation of 2 THz and a constant (white noise) background. The powers were chosen such that 3/4 of the total power is in the peaks, and 1/4 in the background. The line spectrum and its true autocorrelation function, shown over the range $\tau = -300$ to 300 fsec, are given in Figure 9. The interferogram intensity, $I(x)$, predicted by our model for the in-phase and out-of-phase configuration is shown in Figure 10.

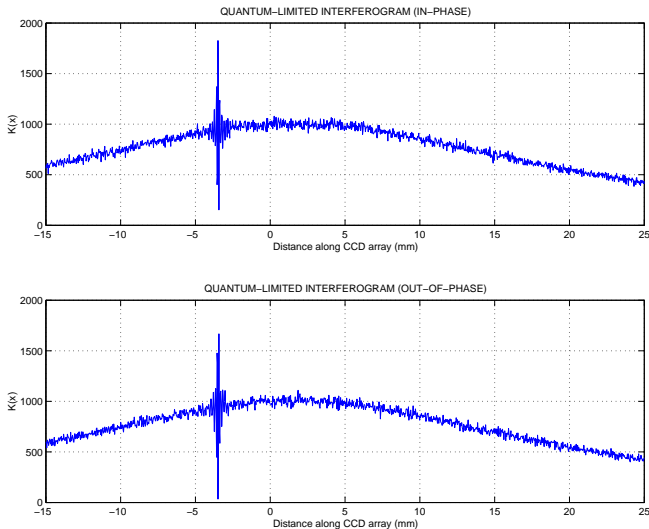


Fig. 8. Poisson-distributed data based on the simulated optical intensity for the solar spectrum corresponding to the in-phase (top panel) and out-phase (bottom panel) configuration of the sensor.

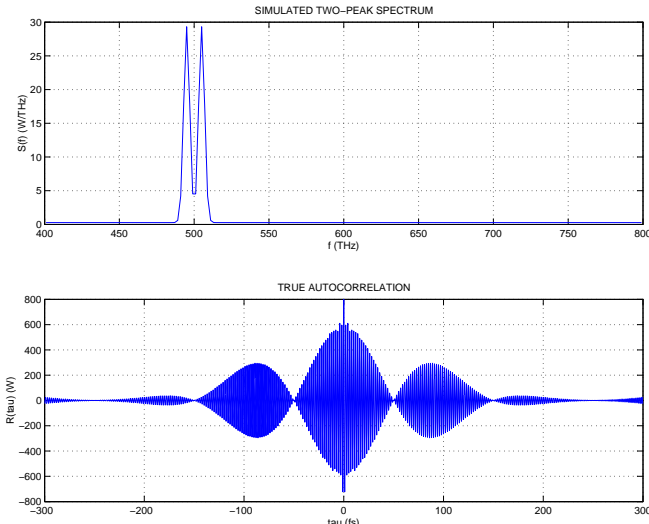


Fig. 9. Simulated optical line spectrum and autocorrelation function.

III. Performance Bounds

For the statistical measurement model of Section II-C above, it is possible to use the Cramer-Rao (CR) bound to determine a matrix lower bound on the covariance matrix of errors for any unbiased estimate $\hat{\mathbf{s}}$ of the spectral vector \mathbf{s} . The maximum-likelihood estimator of Section IV-A appears biased in simulation; the least-squares estimator of Section IV-B is unbiased, provided that nonnegativity is not enforced. In any case, the log-likelihood satisfies the regularity conditions of Cramer [20] for the estimator to be asymptotically unbiased, in which case the CR bound results are useful, and furthermore the maximum-likelihood estimates are asymptotically efficient.

The Fisher information matrix is given by

$$\mathbf{J}_s = -E \left[\frac{\partial^2}{\partial \mathbf{s} \partial \mathbf{s}^T} \ell(\mathbf{s}; \mathbf{k}) \right] \quad (18)$$

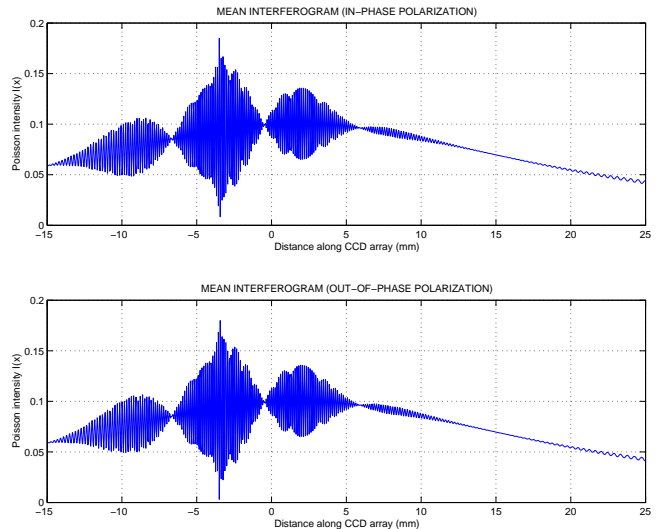


Fig. 10. Simulated optical intensity for the line spectrum corresponding to the in-phase (top panel) and out-phase (bottom panel) configuration of the sensor.

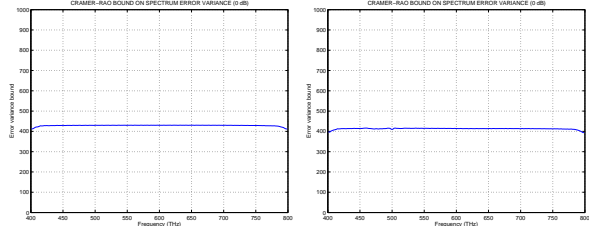


Fig. 11. Cramer-Rao bound on spectrum error variance at 0 dB for: the solar spectrum (left panel) and the line spectrum (right panel).

where ℓ is the log-likelihood for the data \mathbf{k} and the second derivative is evaluated at the true value of \mathbf{s} . The Fisher information matrix for the PET problem, mathematically identical to the one considered here, was derived by Hero and Fessler [21]. Defining the diagonal matrix Λ according to

$$\Lambda = \text{diag}(\lambda) \quad ,$$

then the Fisher information matrix is

$$\mathbf{J}_s = \mathbf{A}^T \Lambda^{-1} \mathbf{A} \quad . \quad (19)$$

The inverse of \mathbf{J}_s is a matrix lower bound on the covariance matrix of any unbiased estimator for \mathbf{s} . That is, if $\mathbf{R}_{\hat{\mathbf{s}}}$ is the covariance of some unbiased estimator $\hat{\mathbf{s}}$, then $\mathbf{R}_{\hat{\mathbf{s}}} - \mathbf{J}_s^{-1}$ is a nonnegative definite matrix. In particular, the diagonal elements of the \mathbf{J}_s^{-1} are lower bounds on the variances of the corresponding individual elements of $\hat{\mathbf{s}}$.

The representation of a Poisson log-likelihood function for a linear measurement model that is used to derive the EM algorithm may also be used to quantify the multiplex disadvantage. We show here that the measurement matrix \mathbf{A} introduces a loss in information in the following sense: the Fisher information matrix for the case where the data are Poisson with mean \mathbf{s} , is greater than the Fisher information matrix in the case where the data are Poisson with mean $\mathbf{A}\mathbf{s}$, assuming that the expected total number

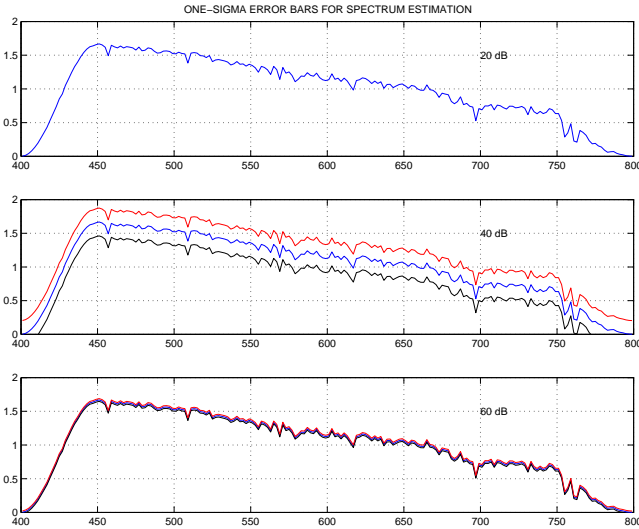


Fig. 12. One sigma error bars for the solar spectrum estimation for 20, 40, and 60 dB. Blue lines are the true spectrum, red and black lines correspond to the mean \pm one standard deviation.

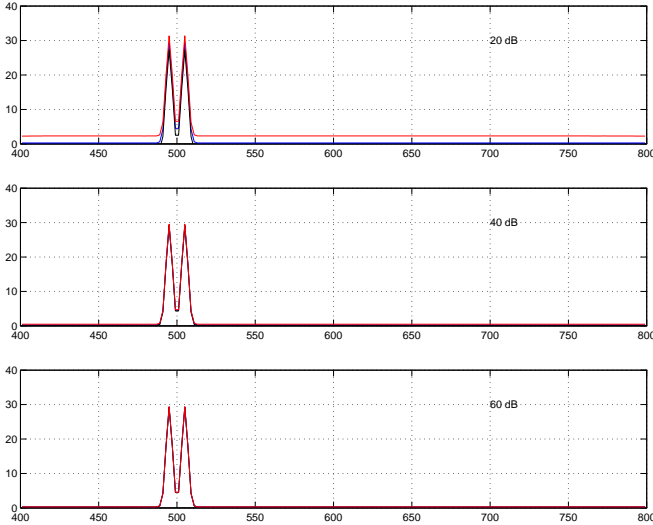


Fig. 13. One sigma error bars for the line spectrum estimation for 20, 40, and 60 dB. Blue lines are the true spectrum, red and black lines correspond to the mean \pm one standard deviation.

of counts is the same. Our proof technique is based on an alternating maximization methodology for Poisson maximum-likelihood problems [22], [23].

The Poisson log-likelihood function, the logarithm of (17), is given by

$$\ell(\mathbf{k} : \mathbf{s}) = \sum_{m=1}^M k_m \ln \left[\sum_{n=1}^N a_{mn} s_n \right] - \sum_{n=1}^N a_{mn} s_n \quad (20)$$

$$= \max_{\pi \in \mathcal{P}} - \sum_{m=1}^M \sum_{n=1}^N \left[\pi_{n|m} k_m \ln \frac{\pi_{n|m}}{a_{mn} s_n} + a_{mn} s_n \right] \quad (21)$$

$$= \max_{\pi \in \mathcal{P}} \mathbf{L}(\mathbf{k} : \mathbf{s}, \pi), \quad (22)$$

where \mathcal{P} is the set of conditional probabilities $\mathcal{P} = \{\pi_{n|m} \geq 0 : \sum_{n=1}^N \pi_{n|m} = 1\}$. The EM algorithm presented in the next section is obtained by alternately

maximizing \mathbf{L} over \mathbf{s} and over $\pi \in \mathcal{P}$. The maximization over π , viewed as a function of the spectrum, is achieved by

$$\pi_{n|m}(\mathbf{s}) = \frac{a_{mn} s_n}{\sum_{\tilde{n}} a_{m\tilde{n}} s_{\tilde{n}}}. \quad (23)$$

The function \mathbf{L} can be written as the sum of two terms, the first being

$$Q(\mathbf{k} : \mathbf{s}, \pi) = \sum_{n=1}^N \sum_{m=1}^M [\pi_{n|m} k_m \ln s_n - a_{mn} s_n], \quad (24)$$

the expected value of the complete data log-likelihood function given \mathbf{k} , where $\pi_{n|m}$ is interpreted as the probability that a measurement in bin m came from spectral bin n . The second term is a measure of the entropy of $\pi_{n|m}$.

The Fisher information matrix in (19) is found by taking the second derivatives of (20) yielding

$$\mathbf{J}_s(j, k) = \sum_{i=1}^M \lambda_i^{-1} a_{ij} a_{ik}. \quad (25)$$

The Fisher information matrix can be found equivalently using the second expression (21). The result of course is the same as using the first expression (20). The advantage is the following. Consider the matrix of second partial derivatives of \mathbf{L} with respect to elements of $\{\mathbf{s}, \pi\}$. The submatrices have entries

$$\frac{\partial^2 \mathbf{L}}{\partial s_n \partial s_{n'}} = - \sum_{m=1}^M \left[\frac{\pi_{n|m} k_m \delta_{nn'}}{s_n^2} \right] \quad (26)$$

$$\frac{\partial^2 \mathbf{L}}{\partial s_n \partial \pi_{n'|m}} = \frac{k_m \delta_{nn'}}{s_n} \quad (27)$$

$$\frac{\partial^2 \mathbf{L}}{\partial \pi_{n|m} \partial \pi_{n'|m'}} = - \frac{k_m \delta_{mm'} \delta_{nn'}}{\pi_{n|m}}, \quad (28)$$

where δ_{nm} is the Kronecker delta. The second derivatives with respect to s_n yield a diagonal matrix whose mean is the negative of the Fisher information matrix that would be obtained if the data $\sum_{m=1}^M \pi_{n|m} k_m$ were Poisson with mean s_n and these data were measured directly. These diagonal entries are

$$\frac{1}{s_n} \sum_{m=1}^M \frac{a_{mn} k_m}{\sum_{\tilde{n}} a_{m\tilde{n}} s_{\tilde{n}}} \quad (29)$$

and have mean $\sum_{m=1}^M a_{mn} / s_n$.

The Fisher information matrix for our problem is obtained from this larger matrix using the first derivatives of $\pi_{n|m}(\mathbf{s})$ with respect to the s_n ,

$$\frac{\partial \pi_{n|m}}{\partial s_{n'}} = \frac{a_{mn} \delta_{nn'}}{\sum_{\tilde{n}=1}^N a_{m\tilde{n}} s_{\tilde{n}}} - \frac{a_{mn} s_n a_{mn'}}{\left[\sum_{\tilde{n}=1}^N a_{m\tilde{n}} s_{\tilde{n}} \right]^2}. \quad (30)$$

The matrix of second derivatives with entries (28) is singular. One set of N singular vectors, indexed by n , has the first N entries equal to $\delta_{nn'}$ and last MN entries equal to $\frac{a_{mn} \delta_{nn'}}{\sum_{\tilde{n}=1}^N a_{m\tilde{n}} s_{\tilde{n}}}$, the first term in (30). Thus the Fisher information matrix is determined by the second

term in (30) and the second partial derivatives of \mathbf{L} with respect to elements of π .

An alternative expression for the Fisher information matrix clearly shows that the information decreases due to not measuring the entries of s_n directly. Plugging (30) into the partial derivatives, keeping the first term (26) separate, and simplifying the remaining terms (the three remaining terms all equal plus or minus the second term here) yields

$$\begin{aligned} [\tilde{\mathbf{J}}_s]_{nn'} &= \frac{\delta_{nn'}}{s_n s_{n'}} \sum_{m=1}^M \pi_{n|m} k_m \\ &- \frac{1}{s_n s_{n'}} \sum_{m=1}^M k_m \sum_{\tilde{n}=1}^N [\delta_{n\tilde{n}} - \pi_{n|m}] \pi_{\tilde{n}|m} [\delta_{n'\tilde{n}} - \pi_{n'|m}]. \end{aligned} \quad (31)$$

The Fisher information matrix is found by taking the expected value of $\tilde{\mathbf{J}}_s$. The second matrix in this expression is clearly nonnegative definite, yielding a decrease in the Fisher information matrix and hence an increase in Cramér-Rao lower bounds.

For the solar and line spectra considered in Section II-D, lower bounds on the variances of the individual spectral values obtained from the diagonal of \mathbf{J}_s^{-1} , are plotted in Figure 11. These plots indicate that the loss of information in comparison with the direct-measurement case can be substantial. For a broad spectrum with an average of one count per spectral bin, the CR bound of the direct-measurement case would be the spectrum \mathbf{s} itself, with an average value of 1. With the interferometric measurement, we see an average variance of over 400 - a drop of over 26 dB in SNR. This loss can be attributed to the Poisson noise in the $R(0)$ pedestal in the tails of the interferogram, which are uninformative in the broad spectrum case. Better results are obtained in the line spectrum case, where the interferogram contains more useful signal information in the tails and the spectral peaks rise above the noise floor. Here the SNR in the peak region of the spectrum is higher than it would be in the direct measurement case. Standard deviations superimposed on the true spectrum are shown in Figure 12 for the solar spectrum and in Figure 13 for the line spectrum. Although the Fisher Information Matrix and its inverse are full-rank in this example, there is considerable correlation exhibited in the off-diagonal elements of \mathbf{J}_s^{-1} . Although this is difficult to display graphically, it means that the errors are correlated across the spectral values in \mathbf{s} . This may also be interpreted as another manifestation of the ‘‘multiplex disadvantage’’.

IV. Spectrum Estimation Methods

In this section we present three different methods for estimating the spectrum S given the Poisson-distributed observations $K_1 \cdots K_M$. The first is based on maximum-likelihood methodology, where we use the EM algorithm to maximize numerically the likelihood function. The second is based on least squares, and the third is a more conventional FFT approach, after preprocessing steps to compensate for the time-delay curve and apodization.

A. Maximum Likelihood via the EM Algorithm

We seek to maximize the log-likelihood (20) with respect to $s_1 \cdots s_N$, subject to the positivity constraint $s_n \geq 0$, $n = 1 \cdots N$.

The EM algorithm derived for other problems can be applied to this problem; the result is an alternating maximization algorithm which guarantees the positivity of the iterates $s_n^{(j)}$ at each iteration j and monotonicity of the log-likelihood function. The initial iterate could be any reasonable guess, such as a flat spectrum with level determined by the sum total of all data or one of the computationally efficient spectrum estimates described below. Derivation of the EM algorithm can be found in [8], [10], [22]. As mentioned in Section III, the algorithm alternately maximizes $\mathbf{L}(\mathbf{k} : \mathbf{s}, \pi)$ over \mathbf{s} and over π . With π evaluated using (23) at the previous estimate $\mathbf{s}^{(j)}$,

$$\begin{aligned} s_n^{(j+1)} &= \sum_{m=1}^M \frac{\pi_{n|m}(\mathbf{s}^{(j)}) k_m}{\sum_{m'=1}^M a_{m'n}} \\ &= \left(\frac{s_n^{(j)}}{\sum_{m'=1}^M a_{m'n}} \right) \sum_{m=1}^M a_{mn} \frac{k_m}{(\sum_{n'=1}^N a_{mn'} s_{n'}^{(j)})} \end{aligned} \quad (32)$$

With proper data organization, each step of the algorithm is easily implemented using 2 matrix-vector multiplications, 1 pointwise vector multiplication and 2 pointwise vector divisions. (The required number of iterations is difficult to determine a priori, but in our experience, with the simulation parameters of Section II-D, 100 to several hundred iterations are required.)

B. Least Squares

As an alternative to the maximum-likelihood method, we consider a least-squares approach. The least squares solution is that one which minimizes the squared error between the data and the synthetic model based on the chosen parameter.

The intensity vector λ is related to the underlying spectrum through the linear relation $\lambda = \mathbf{A}\mathbf{s}$ as seen previously. The observation \mathbf{k} has expectation λ . The least-squares estimation of \mathbf{s} is then that value of \mathbf{s} which minimizes

$$|\mathbf{k} - \mathbf{A}\mathbf{s}|^2. \quad (34)$$

The solution, found by straightforward matrix calculus, is

$$\hat{\mathbf{s}}_{LS} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{k}. \quad (35)$$

The pseudo-inverse

$$\mathbf{A}^\# = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \quad (36)$$

could be computed off-line if the calibration parameters are known and the estimation is to be performed multiple times.

It is easily seen that

$$E[\hat{\mathbf{s}}_{LS}] = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{A} \mathbf{s} = \mathbf{s}, \quad (37)$$

and thus, the least-squares estimate is unbiased and the Cramer-Rao bound derived in Section III is applicable.

However, an exact variance expression is available in this case, so the CR bound is not especially useful. Note that

$$\text{cov}(\mathbf{s}) = \mathbf{A}^\# \text{cov}(\mathbf{k}) \mathbf{A}^{\#T} . \quad (38)$$

Since the components of \mathbf{k} are independent and Poisson, we have that $\text{cov}(\mathbf{k}) = \text{diag}(\lambda) = \Lambda$, and hence

$$\text{cov}(\mathbf{s}) = \mathbf{A}^\# \Lambda \mathbf{A}^{\#T} . \quad (39)$$

The diagonal elements of this matrix comprise the variances of the spectral values.

Note that a nonnegativity constraint was not enforced on the spectrum estimates in any of the above derivations. This could be done by either 1) clipping the spectrum estimates, so that any negative spectral values obtained by (35) are set to 0, or 2) finding $\hat{\mathbf{s}}$ which satisfies the nonnegative least squares problem, i.e., minimizes (34) subject to nonnegativity constraints. Either method leads to positively-biased spectrum estimates; the latter approach is computationally prohibitive for the dimensions involved. Instead, least-squares estimates were computed using the Matlab pseudo-inverse routine and then negative values were set to zero.

C. FFT with Data Precorrection

A computationally efficient and reasonably effective method for estimating the spectrum can be crafted based on the discrete Fourier transform (DFT). In fact, the fundamental idea of measuring the autocorrelation and taking the inverse Fourier transform to get the spectrum is the basis of all interferometers. As shown above, such an approach ignores the Poisson statistics of the data, and indeed is based on a noiseless data model. However, with appropriate attention to other sensor effects, a reasonable spectrum estimation algorithm is obtainable.

Aside from the noise, the two sensor effects are the apodization curve and the nonuniform sampling of the time-delay axis by the uniformly-spaced CCD elements. We propose simple methods to mitigate these effects by preprocessing the data. First, all measurements $K(x)$ can be divided pointwise by the apodization curve $g(x)$ to arrive at a measurement which has uniform gain along the entire interferogram. Second, if data corresponding to both $R(0) + R(\tau)$ and $R(0) - R(\tau)$ are available, these can be subtracted to arrive at a single composite measurement of $R(\tau)$. Finally, the autocorrelation can be resampled from the nonuniform grid afforded by the CCD elements, to a uniform grid in τ space. This interpolation can be found using some bandlimited interpolation function (such as a sinc), or by using cubic spline interpolation. The maximum value of τ should be the same as that corresponding to the most distance CCD element from the interferogram center, and the grid spacing should be determined by the largest frequency present in the spectrum, using sampling theorem concepts. Note that a grid spacing finer than what is given by the measurements themselves will lead to meaningless results at high frequencies.

The autocorrelation function is an even function of τ , and this symmetry can also be used to advantage. Once

the interferogram has been resampled onto a uniform grid, values of $R(\tau)$ for which there are two data points on either side of the center can be replaced by their average. Values of $R(\tau)$ for which only one data point exists can be set equal to this value on either side of $\tau = 0$. Furthermore, this symmetrized autocorrelation can be zero-padded out to some convenient number of time delays, presumably a power of 2, for purposes of efficiently computing the DFT via a fast Fourier transform (FFT) algorithm. Details on this method can be found in [15].

In our simulation results, resampling of the autocorrelation is accomplished using the Matlab routine `spline` and the spectrum is computed using a 4096-point FFT on the symmetrized interferograms.

V. Simulation Summary

Results obtained with the three methods (EM, LS, and FFT) and slight variations thereof are summarized in this section. Throughout we adopt the conventions that 1) the spectrum vector \mathbf{s} is normalized so that the average component is equal to 1.0, and 2) the columns of the matrix \mathbf{A} are normalized to unit sum, then multiplied by some constant c . The data are Poisson with mean $\Lambda \mathbf{s}$, as described previously. For Poisson random variables, the mean equals the variance and therefore the SNR (mean squared over variance) equals the intensity. Hence we can report that the input SNR, by which we mean the average intensity per spectral bin, is equal to the constant c , or $10 \log c$ if we use a dB scale. By keeping the spectrum constant at one count per spectral bin, and changing the number of Poisson counts through variation of c , it is possible to compare the performance of the various algorithms directly as a function of SNR.

Poisson data with SNR equal to 20-60 dB, in 10 dB increments, were processed with the three methods. Figures 14-16 show the results for the solar spectrum, and results for the line spectrum are shown in Figures 17-19.

Figure 14 illustrates the result of processing one realization of the Poisson data for the solar spectrum at SNR = 40 dB. It is evident that the level of noise in all three spectrum estimates obscures the fine detail in the spectrum at this SNR. Also shown in this figure are the computation times for the three methods, for the Sun Sparcstation used to carry out the simulations. Note that there is about an order of magnitude difference between the FFT and LS methods, and another order of magnitude between the LS and EM methods.

In Figure 15 we compare the three methods at a higher SNR of 60 dB. In this case the fine detail in the spectrum is revealed in the LS and EM estimates. The FFT method also reveals this detail, although there is an apparent systematic bias at higher frequencies.

The result of a large simulation which summarizes the performance of the various methods as a function of SNR is shown in Figure 16. Each data point shown is the result of 10000 simulated spectrum estimates from Poisson data, and the plotted results are the average squared error between true and estimated spectral, per spectral

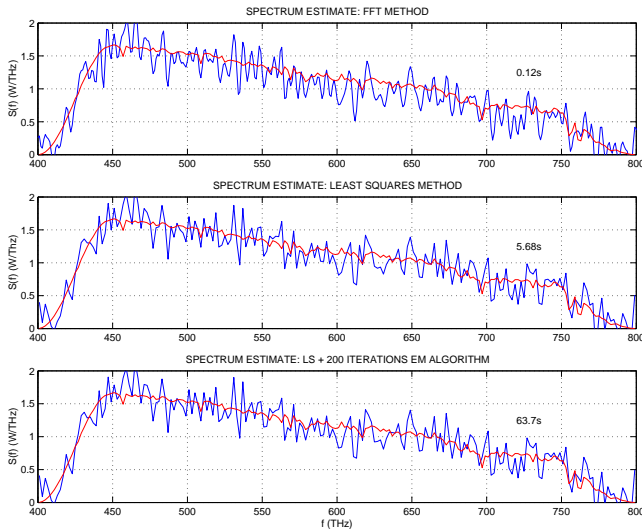


Fig. 14. Solar spectrum estimated from one realization of the data with SNR = 40 dB with three different methods: FFT method (top panel), LS method (middle panel), and EM algorithm (bottom panel). The red line plots the true solar spectrum. The time it took to compute each estimate is indicated in seconds in each panel. The LS result was used as the initial estimate for the EM algorithm.

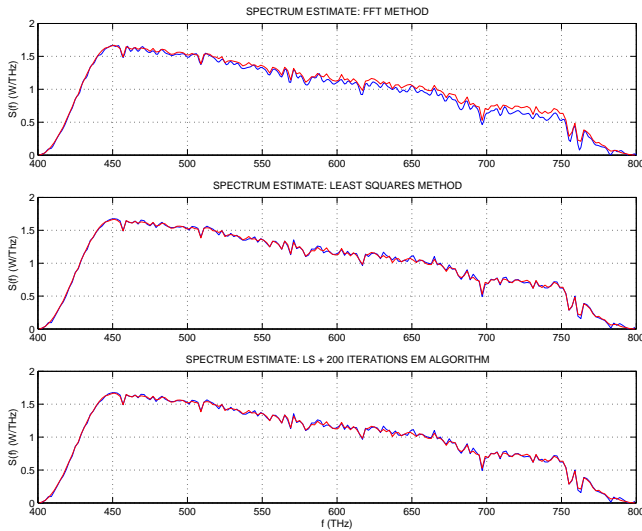


Fig. 15. Solar spectrum estimated from one realization of the data with SNR = 60 dB with three different methods: FFT method (top panel), LS method (middle panel), and EM algorithm (bottom panel). The red line plots the true solar spectrum. The LS result was used as the initial estimate for the EM algorithm.

bin, expressed on a dB scale. Two versions each of the LS and FFT algorithms were considered, one in which the spectral values are thresholded at 0 (indicated by the term CLIP), and one in which they were not. Since in the true spectra there is an average signal level of one count per spectral bin, the negative of the result plotted can also be interpreted as the average output SNR. The average Cramer-Rao bound, found by averaging the diagonal elements of the inverse of the Fisher Information Matrix, is also plotted in these figures.

All the methods yield an average squared error that is lower than the CR bound, except least-squares (LS) where

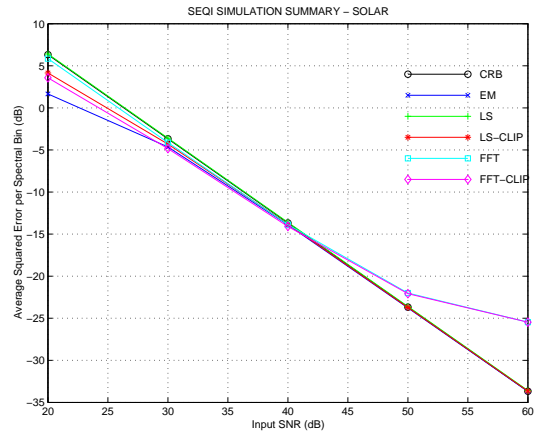


Fig. 16. Average squared error per spectral bin, for the solar spectrum computed for 5 different methods: 1) EM; 2) LS; 3) LS-CLIP which is the result of thresholding the LS result at 0; 4) FFT; and 5) FFT-CLIP which is the result of thresholding the FFT result at 0. CRB is the Cramer-Rao bound.

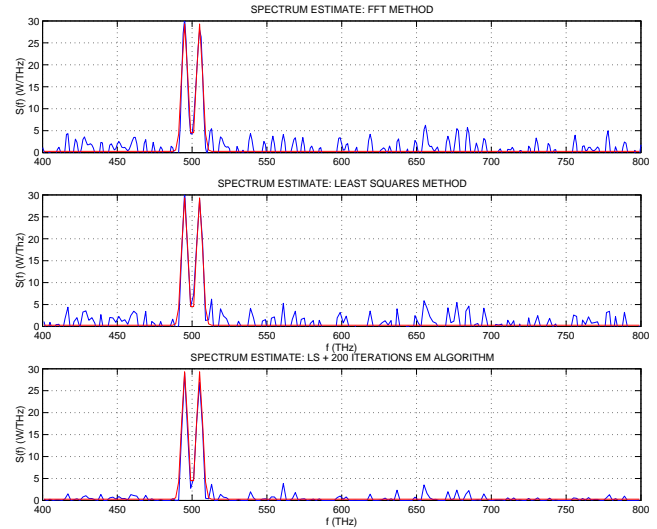


Fig. 17. Line spectrum estimated from one realization of the data with SNR = 20 dB with three different methods: FFT method (top panel), LS method (middle panel), and EM algorithm (bottom panel). The red line plots the true line spectrum.

we can in fact compute the squared error analytically. This is not too surprising since all the other methods are biased, so the CR bound is not really applicable. At low SNR, EM yields the smallest squared error. Around 40 dB (10000 counts per spectral bin), all methods are about the same and lead to an output SNR of 14-15 dB. At high SNR, the relative performance of FFT methods falls off. We believe this is due to systematic errors in the various interpolations used, which begin to dominate the error for that method in this regime. Also at high SNR, the CR bound and the performances of the EM and LS algorithms essentially coincide.

Figure 17 depicts the results for processing one realization of the line spectrum at an SNR of 20 dB. Here we notice a significant improvement in performance of the EM algorithm relative to the LS approach, unlike what

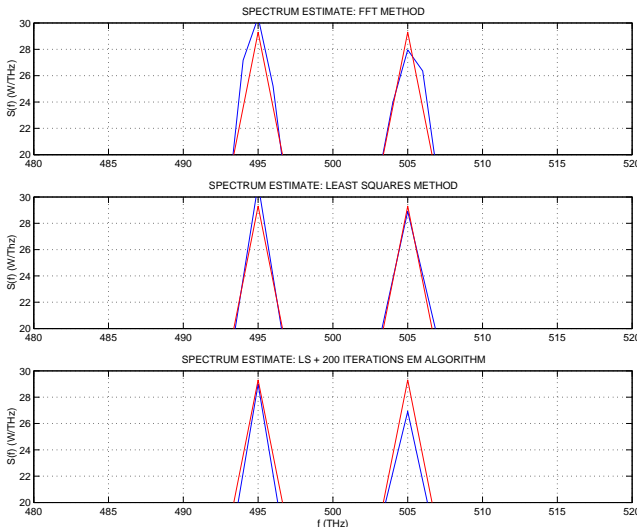


Fig. 18. Line spectrum estimated from one realization of the data with SNR = 20 dB with three different methods: FFT method (top panel), LS method (middle panel), and EM algorithm (bottom panel). The red line plots the true line spectrum.

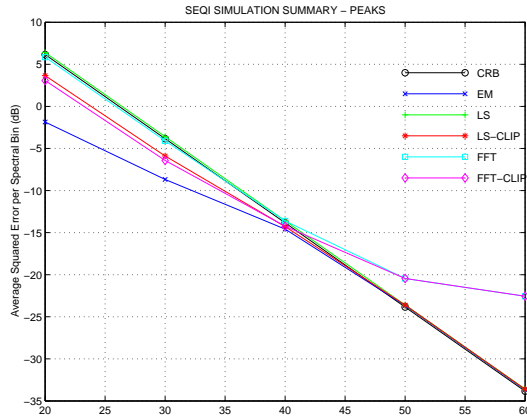


Fig. 19. Average squared error per spectral bin for the line (or peaks) spectrum computed for 5 different methods: 1) EM; 2) LS; 3) LS-CLIP which is the result of thresholding the LS result at 0; 4) FFT; and 5) FFT-CLIP which is the result of thresholding the FFT result at 0. CRB is the Cramer-Rao bound.

was observed in the solar case. An expanded view of the estimated spectra in the vicinity of the peaks is shown in Figure 18. Figure 19 summarizes the performance of the various methods for this spectrum, analogous to what is shown in Figure 16 and with similar results.

VI. Discussion

The results of our analysis and simulations reveal a number of interesting issues in the design of interferometric spectrometers for low light-level applications. Because our results are all based on rather generic models and ideal simulations, we do not attempt to draw definitive conclusions or recommendations about specific instruments or the algorithms used to process data. However, some considerations which undoubtedly will arise in such designs are worth pointing out.

The first issue has to do with the multiplex disadvantage which is often associated with using interferometric sensors. In our analysis, we compare the error in the spectrum estimates based on Poisson data with mean cs (the so-called direct measurement) with the error based on Poisson data with mean As , where the matrix A is chosen so that the total number of Poisson counts is preserved. The theoretical drop in performance is dramatic - as much as 26 dB based on the parameters we have chosen. However, this observation must be tempered by the fact that we are comparing a realistic interferometer model to a highly idealized direct-measurement instrument. The resolving power of a real dispersive instrument would certainly play a role here, as would many other optical and electronic effects which have not been included in our model, such as diffraction, stray light, dark current, and flat-fielding effects.

The source of the multiplex disadvantage seems clear: it is the Poisson noise associated with the $R(0)$ pedestal that extends across the entire length of the interferogram, when most of the interesting signal information is confined to a very short interval surrounding the point $\tau = 0$ (see Figures 7 and 8). This suggests a design trade-off between the length of the interferogram, which will affect spectrum resolution, and the output SNR.

For certain kinds of signals, namely monochromatic light or signals with line spectra, our analysis indicates that the multiplex disadvantage may actually become a multiplex advantage in those regions of the spectrum with signal energy. Because the effects of the noise are spread across the entire spectrum, the output SNR at the spectrum peaks could be higher than what it would have been in the direct measurement case. The price to pay for this improved performance near the peaks is the increased noise level in the background spectrum.

The fact that all methods except LS exhibit squared-error performance superior to the CR bound indicates that these methods are biased. A complete study of the nature of this bias has not been carried out. The two example spectra described in this paper were the only ones studied, and were not chosen to show the statistical spectrum estimation algorithms in a favorable light. We leave open the possibility that, because of the bias, there may exist other true spectra for which the algorithms have performance worse than that indicated by the CR bound. Such a study is outside the scope of this paper.

The choice of spectrum estimation algorithm will depend on the design goals and applications of a particular instrument. Our results indicate that the EM algorithm is generally superior but that it is computationally expensive and therefore may not always be the best choice. At low SNR, the advantage of the EM algorithm is most obvious, but one must make the decision as to whether the spectrum so obtained has any value at all at such a low output SNR. The answer may be yes in the case of line spectra, and furthermore there exist in the literature a host of modifications, constraints, and regularizations (such as complexity regularization [24], [25]) which could

be introduced in this regime to ameliorate some of the noise artifacts. At medium SNR levels all methods were comparable and thus the FFT approach seems attractive because of its computational efficiency. At high SNR, the systematic shortcomings of the FFT approach begin to appear, and the LS method emerges as a good choice due to the combination of statistical and computational performance. Again, the final choice would depend on a thorough analysis of the instrument, the light level, and the application.

VII. Conclusions

A quantitative model for interferometric data collection in a quantum-limited hyperspectral imaging system was described. This model accounts for the nonlinear relationship between CCD position x and differential time delay τ , the effects of apodization, and the Poisson statistics of the acquired data. The multiplex disadvantage, by which we mean the drop in performance of an interferometric instrument relative to an idealized direct-measurement instrument, was evident both in the Cramer-Rao bound analysis and in the simulations. The problem of estimating the spectrum was considered using a variety of techniques, namely the EM algorithm (often applied in other Poisson inverse problems), least squares, and the FFT with data precorrection. Simulation results indicate advantages, disadvantages, and design trade-offs with all the methods, and that the correct choice would most likely depend on a more thorough analysis and experimentation with a particular instrument. Future research in this area could include extending the model to include nonuniformities in the focal plane array and for dark current, as in [17], and the consideration of design trade-offs in the choice of the interferometer parameters.

VIII. Acknowledgments

This work was supported in part by a grant from the Boeing-McDonnell Foundation and by ONR MURI grant N00014-98-1-06-06.

References

- [1] Robert J. Bell. *Introductory Fourier Transform Spectroscopy*. Academic Press, 1972.
- [2] William H. Smith and W. V. Schempp. Digital array scanned interferometers. *Experimental Astronomy*, 1:389–405, 1991.
- [3] Leonard J. Otten III et al. On board spectral imager data processor. In M. Descour and S. Shen, editors, *Imaging Spectrometry V*, volume 3753 of SPIE Proc., pages 86–94, 1999.
- [4] John P. Burg. *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford University, 1975.
- [5] Petre Stoica and Randolph Moses. *Introduction to Spectral Analysis*. Prentice-Hall, 1997.
- [6] L. Rebollo Neira and A. G. Constantinides. Power spectrum estimation from values of noisy autocorrelations. *Signal Processing*, 50:223–231, 1996.
- [7] L. Rebollo Neira and A. G. Constantinides. Power spectrum estimation of noisy and limited autocorrelations: a maximum entropy approach. *Signal Processing*, 56:135–147, 1997.
- [8] Stephen E. Bialkowski. Overcoming the multiplex disadvantage by using maximum-likelihood inversion. *Applied Spectroscopy*, 52:591–598, 1998.

- [9] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Ann. Royal Statistical Soc.*, B-36:1–38, 1977.
- [10] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Medical Imaging*, 1:113–122, 1982.
- [11] Michael I. Miller and Donald L. Snyder. The role of likelihood and entropy in the incomplete-data problems: applications to estimating point-process intensities and Toeplitz-constrained covariances. *Proc. IEEE*, 75, 1987.
- [12] Edward Voigtman and James D. Winefordner. The multiplex disadvantage and excess low-frequency noise. *Applied Spectroscopy*, 41:1182–1184, 1987.
- [13] P. Jacquiot. New developments in interference spectroscopy. *Rep. Prog. Phys.*, 23:267–312, 1960.
- [14] William H. Smith and Phillip D. Hammer. Digital array scanned interferometer, sensors and results. *Applied Optics*, 35:2902–2909, 1996.
- [15] Daniel R. Fuhrmann and William H. Smith. Empirical modeling and calibration of Fourier transform spectrometers I: linearization and normalization of interferograms. *Optical Engineering*, 2002. in review.
- [16] Daniel N. Harres, Daniel R. Fuhrmann, and William H. Smith. Compensation for optical distortion in Fourier Transform spectrometers. In M. Descour and S. Shen, editors, *Imaging Spectrometry V*, volume 3753 of SPIE Proc., pages 142–151, 1999.
- [17] Donald L. Snyder, Abed Hammoud, and R. White. Image recovery from data acquired with a charge-coupled device camera. *J. Optical Soc. Amer. A*, 10:1014–1023, 1993.
- [18] The 1985 Wehrli Standard Extraterrestrial Solar Irradiance Spectrum. <http://rredc.nrel.gov/solar/standards/am0/#spectrum>.
- [19] M. Descour and S. Shen, editors. *Imaging Spectroscopy V*, volume 3753. SPIE, 1999.
- [20] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, 1946.
- [21] Alfred O. Hero and Jeffrey A. Fessler. A recursive algorithm for computing CR-type bounds on estimator covariance. *IEEE Trans. Information Theory*, 40:1205–1210, 1994.
- [22] Joseph A. O'Sullivan, Donald L. Snyder, and Richard E. Blahut. Information-theoretic image formation. *IEEE Trans. Information Theory*, 44:2094–2123, 1998.
- [23] Joseph A. O'Sullivan. *Alternating Minimization Algorithms: from Blahut-Arimoto to Expectation-Maximization*. In A. Vardy, editor, *Codes, Curves, and Signals: Common Threads in Communications*, pages 173–192. 1998.
- [24] P. Moulin and J. Liu. Statistical imaging and complexity regularization. *IEEE Trans. Information Theory*, 46:1762–1777, 2000.
- [25] J. Liu and P. Moulin. Complexity-regularized image denoising. *IEEE Trans. Image Processing*, 10:841–851, 2001.