Cramér–Rao sensitivity limits for astronomical instruments: implications for interferometer design

Jonas Zmuidzinas

Division of Physics, Mathematics, and Astronomy, California Institute Institute of Technology, 320-47 Pasadena, California 91125

Multiple-telescope interferometry for high-angular-resolution astronomical imaging in the optical–IR–far-IR bands is currently a topic of great scientific interest. The fundamentals that govern the sensitivity of direct-detection instruments and interferometers are reviewed, and the rigorous sensitivity limits imposed by the Cramér–Rao theorem are discussed. Numerical calculations of the Cramér–Rao limit are carried out for a simple example, and the results are used to support the argument that interferometers that have more compact instantaneous beam patterns are more sensitive, since they extract more spatial information from each detected photon. This argument favors arrays with a larger number of telescopes, and it favors all-on-one beam-combining methods as compared with pairwise combination. © 2003 Optical Society of America

OCIS codes: 030.4280, 030.5260, 030.5290, 110.4280, 120.6200, 270.5290, 350.1270.

1. INTRODUCTION

Astronomical spatial interferometry, which is the technique of combining the radiation gathered by several separated telescopes, is of great current interest because of the scientific potential of very-high-angular-resolution observations, combined with numerous technological advances that now make interferometry feasible at optical–IR wavelengths. As a result, this field is very active, and serious investments are being made: Numerous ground-based facilities are in development, large arrays are being discussed, and ambitious space missions are being considered. Detailed reviews of this field, which describe the scientific motivation and results, technical challenges and approaches, and existing and planned facilities and which contain extensive bibliographies, have been made recently.[1,2] Additional information is readily available.[3]

Of course, interferometry is very well developed at radio wavelengths, at which large arrays of telescopes, such as the National Radio Astronomy Observatory Very Large Array (NRAO VLA),[4] routinely provide high-angular-resolution synthetic aperture imaging. Ground-based interferometry at optical wavelengths is inherently more difficult owing to the limitations imposed by atmospheric phase fluctuations; dealing with these fluctuations is perhaps the key issue for ground-based systems and provides strong motivation to consider interferometry in space. As a result, NASA is pursuing the Space Interferometry Mission,[5] an optical astrometric interferometer with a 10-m baseline, scheduled for launch at the end of the decade. The Space Interferometry Mission will perform a wide range of science, including searches for extrasolar planets as well as synthetic aperture imaging of the centers of galaxies. Another important advantage of space interferometry is the unobstructed transmission and low background over the entire IR–far-IR–submillimeter spectrum. Looking further into the future, IR interfer-

ometers are being considered for missions such as the Terrestrial Planet Finder[6] and Darwin,[7] which have the ambitious goal of detecting and characterizing Earth-like planets around nearby stars. Direct-detection space interferometers at very long (far-IR–submillimeter) wavelengths that use cold telescopes have also been proposed, such as the NASA Submillimeter Probe of the Evolution of Cosmic Structure and Space Infrared Interferometric Telescope (SPECS/SPIRIT) concepts.[8–10] Such interferometers would give the angular resolution needed to break through the spatial confusion limit,[11] which becomes severe at these long wavelengths, and would allow a detailed study of the properties of the newly discovered class of submillimeter-luminous, dusty galaxies at high redshifts.[12–14]

In spite of this high level of activity, it appears that important design considerations for optical interferometers are not yet resolved. One major issue is the number of telescopes that should be used; another issue is the method of beam combination. Of course, various practical constraints may limit the range of design options; for instance, it is likely to be important to minimize the number of telescopes for a space interferometer. Although a number of papers have addressed these various issues,[15–24] there appears to be no general consensus on which design approaches give the best sensitivity or even if there is much difference between them.[1] It is essential to have a full understanding of the fundamental issues that determine the sensitivity of optical interferometers; advancing that understanding is the goal of this paper. We will therefore ignore important technical issues, such as the methods used to deal with atmospheric fluctuations, mechanical and thermal perturbations, detector noise, etc., and discuss only interferometers with nearly ideal characteristics.

Our approach will focus on the instantaneous (not synthesized) angular response function or beam pattern as-

sociated with a given detector that receives the light gathered by the interferometer. We will argue on general grounds, supported by the Cramér–Rao lower bound on the uncertainty of statistical inference, that it is important to make the instantaneous angular response function as compact as possible in order to extract the maximum amount of spatial information from each detected photon. This argument favors certain interferometer designs over others. For instance, pairwise beam combination in which the light from T telescopes is split and interfered on $T(T-1)/2$ detectors (one detector per baseline), which is essentially the approach adopted for radio interferometry, is, in general, less sensitive for direct-detection interferometric imaging than schemes in which the light from all T telescopes is coherently combined onto the detectors. The reason for this is that the angular response functions are less compact for the case of baseline pair combination, and each photon detected provides less information about the spatial structure of the source. This distinction vanishes for the radio case, in which one is dealing with amplified signals that have high occupation numbers for the photon modes, and so photon bunching plays an important role. The connections and distinctions between the optical and the radio cases will be discussed in more detail in a future paper.[25]

The structure of the paper is as follows. Sections 2–5 are devoted to giving a rigorous definition of the photon-detection probability matrix of an arbitrary astronomical instrument, such as an interferometer, in terms of electromagnetic scattering matrices. The formalism described in these sections is quite powerful and allows us to handle any interferometer configuration or beam-combining method. Many of the key results described in these sections are well known and will likely be familiar. However, the use of scattering matrices and photon-detection probability matrices to describe astronomical instruments and interferometers is not common and is therefore covered in some depth.

The key mathematical results are described in Section 6. There we discuss the Cramér–Rao sensitivity bound, as well as two techniques (maximum likelihood and least squares) that can be used to obtain the source intensity distribution from the observed photon counts; the overall goal is to argue that the Cramér–Rao lower bound actually gives a good estimate of the instrument performance. We also show that the only instruments that can achieve ideal sensitivity are those that do not mix up photons from different spatial or spectral channels. Unfortunately, such instruments are difficult to build; interferometers do, in fact, mix up photons spatially (and spectrally, in some cases).

To demonstrate the consequences of a nonideal response, in Section 7 we present the results of a numerical calculation of the Cramér–Rao bound for the simple but illustrative case of one-dimensional interferometer arrays. To do this, we draw on the material presented in Sections 2–6. Here we demonstrate that arrays with a larger number of telescopes are superior and that N-way beam combining is superior to two-way combining. The paper closes with comments on these results and indicates areas for future research.

2. SCATTERING MATRIX DESCRIPTION OF OPTICAL SYSTEMS

The response of any electromagnetic system, such as a collection of optical elements (mirrors, lenses, etc., but not detectors), may be described by a classical scattering matrix S. Scattering matrices are common in electrical engineering,[26,27] in which they are used to describe linear N-port circuits:

$$b_i = \sum_j S_{ij} a_j. \qquad (1)$$

The indices $1 \leq i,\ j \leq N$ label the ports. A set of N transmission lines attached to the ports carries incoming waves with amplitudes a_i and outgoing waves with amplitudes b_i. The standard practice is to normalize these amplitudes to give simple expressions for the power carried by the waves; for instance, the total incident power is $Q_{\text{inc}} = \Sigma_i |a_i|^2$. With this choice of normalization, it is straightforward to show that a lossless circuit has a unitary scattering matrix, $SS^\dagger = I$ (here I is the identity matrix). Reciprocity, which has its roots in time-reversal symmetry and applies to most passive circuits, implies that $S^T = S$. When the wave amplitudes are expressed in terms of voltages and currents at the ports, the scattering matrix S may be related to more familiar quantities such as the impedance matrix Z. Finally, we note that (classical) noise can be treated very naturally within the framework of scattering matrix theory.[26,28–30] The extension to include quantum effects, such as photon-counting statistics, is straightforward.[25]

We can apply the scattering matrix concept to characterize an optical system, such as the collection of telescopes and beam-combining optics that make up an interferometer. This approach was developed for antenna problems during the World War II radar effort[31] and is particularly convenient at radio wavelengths, since it allows circuit and antenna concepts to be treated in a unified fashion. Although scattering matrices are frequently applied to antenna problems,[32,33] they are not often used to describe optical systems (though they do find occasional use[34]). Scattering matrices are very useful for describing guided-mode optics, which are, in fact, of substantial interest for astronomical interferometry.[35,36] Scattering matrices are also very well suited for problems involving the quantum-mechanical nature of the radiation field, such as photon-counting statistics, because one directly deals with the modes of the radiation field. For these reasons, we adopt the scattering matrix approach. To make the paper reasonably self-contained, and to establish our notation, we will review this approach in detail.

We begin by describing the radiation field in terms of incoming and outgoing plane waves. For simplicity, we will continue to use a classical description for the electromagnetic field; it is straightforward to adapt our formalism to the case of a quantum electromagnetic field.[25] Assuming a time-harmonic $\exp(+j\omega t)$ time dependence, the electric field of the incoming wave arriving at the telescope system (or antenna) can be expressed as

$$\mathbf{E}_{\mathrm{inc}}(\mathbf{r}) = \frac{\sqrt{2\,\eta_0}}{\lambda} \int d\Omega\, \mathbf{a}(\Omega) \exp[+jk\,\hat{\mathbf{n}}(\Omega) \cdot \mathbf{r}]. \quad (2)$$

Here $\mathbf{a}(\Omega)$ represents the amplitude and polarization distribution of the incoming plane waves, λ is the free-space wavelength, and $\eta_0 = 377\Omega$ is the free-space impedance. As usual, Ω represents the polar angles (θ, ϕ) with respect to the chosen coordinate system; the unit vector $\hat{\mathbf{n}}(\Omega) = \hat{\mathbf{z}}\cos\theta + \sin\theta(\hat{\mathbf{x}}\cos\phi + \hat{\mathbf{y}}\sin\phi)$ describes the direction from which a plane-wave component is arriving. Similarly, the outgoing wave can be expressed as

$$\mathbf{E}_{\mathrm{out}}(\mathbf{r}) = \frac{\sqrt{2\,\eta_0}}{\lambda} \int d\Omega\, \mathbf{b}(\Omega) \exp[-jk\,\hat{\mathbf{n}}(\Omega) \cdot \mathbf{r}]. \quad (3)$$

The normalization is again chosen to give simple expressions for power, for instance,

$$Q_{\mathrm{inc}} = \int d\Omega\, |\mathbf{a}(\Omega)|^2. \quad (4)$$

In the usual case of incoherent light emitted by astronomical sources, the wave amplitudes can be considered to be random quantities with certain statistical properties.

An antenna is usually thought of as having one or more well-defined ports, or terminals, that one can attach to other circuits, such as an amplifier. In the case of a radio telescope, the terminal may well be the output waveguide of a feedhorn. The situation is a little more subtle for the case of an optical telescope system, in which the light is focused directly on the bare pixels of a detector array, such as a CCD. [In reality, the distinction between optical and radio techniques is blurring. For instance, the introduction of single-mode optical fibers for collecting and transporting light from the focus of a telescope is completely analogous to the use of radio feedhorns and waveguides, and bare pixel detector arrays find use even at very long (submillimeter) wavelengths.] We can imagine describing the radiation incident on a given detector pixel, which we label α, in terms of a modal expansion. At the detector surface, these spatial modes may be constrained to have a nonzero amplitude only over the region occupied by the pixel (for pixel sizes exceeding $\sim\lambda$, this constraint makes the modes for different pixels orthogonal, which simplifies the form of the scattering matrix). The ith such mode for pixel α has incoming wave amplitudes $a_{i\alpha}$ and outgoing wave amplitudes $b_{i\alpha}$, which we assume have the usual power normalization. We will define the terms "incoming" and "outgoing" in reference to the telescope system rather than the detector pixels, so that the light gathered by the telescope system that arrives at the detectors is characterized by the amplitude $b_{i\alpha}$. Assuming perfect absorption by the detector, the power absorbed can be expressed as

$$Q_\alpha = \sum_i |b_{i\alpha}|^2. \quad (5)$$

Imperfect absorption by the detector (quantum efficiency below unity), as well as varying sensitivities to the different spatial modes, can be incorporated into the definition of the scattering matrix of the optical system, which we discuss below. In the case of a system designed for diffraction-limited imaging, most of the light absorbed by the detector will be contained in a single mode (or two modes, for polarization-insensitive detectors).

The telescopes and associated optical system (e.g., beam-combining optics) can be characterized by a generalized scattering operator S. This operator acts on a Hilbert space consisting of vectors of the form

$$a = \begin{bmatrix} \mathbf{a}(\Omega) \\ a_{i\alpha} \end{bmatrix}, \quad (6)$$

where $\mathbf{a}(\Omega)$ is square integrable. The operator S can be partitioned into four blocks:

$$S = \begin{bmatrix} \mathcal{S}^{(\mathrm{scat})} & \mathbf{S}^{(\mathrm{rec})} \\ \mathbf{S}^{(\mathrm{trans})} & S^{(\mathrm{refl})} \end{bmatrix}. \quad (7)$$

The first block, $\mathcal{S}^{(\mathrm{scat})}(\Omega, \Omega')$, describes the scattering of an incoming plane wave arriving from Ω' to an outgoing plane wave traveling toward Ω. The script font reminds us that this is a 3×3 matrix to account for polarization. The second and third blocks are off diagonal: $\mathbf{S}^{(\mathrm{rec})}_{i\alpha}(\Omega')$ describes the (vector) receiving antenna pattern for the detector pixel mode $i\alpha$, and $\mathbf{S}^{(\mathrm{trans})}_{j\beta}(\Omega)$ describes the transmitting antenna pattern. The fourth block is an ordinary matrix, $S^{(\mathrm{refl})}_{i\alpha j\beta}$, that represents the scattering (reflection) of the optical system between the various detector pixel modes. The meaning of these quantities becomes clear when we write expressions for the outgoing waves in terms of the incoming waves:

$$\mathbf{b}(\Omega) = \int d\Omega'\, \mathcal{S}^{(\mathrm{scat})}(\Omega, \Omega')\mathbf{a}(\Omega') + \sum_{j\beta} \mathbf{S}^{(\mathrm{trans})}_{j\beta}(\Omega)a_{j\beta}, \quad (8)$$

$$b_{i\alpha} = \int d\Omega'\, \mathbf{S}^{(\mathrm{rec})}_{i\alpha}(\Omega') \cdot \mathbf{a}(\Omega') + \sum_{j\beta} S^{(\mathrm{refl})}_{i\alpha j\beta}a_{j\beta}. \quad (9)$$

Assuming that the optical system contains only reciprocal elements (e.g., no Faraday rotation isolators), we know that the scattering operator must equal its transpose. This implies that the transmitting and receiving patterns are the same:

$$\mathbf{S}^{(\mathrm{trans})}_{i\alpha}(\Omega) = \mathbf{S}^{(\mathrm{rec})}_{i\alpha}(\Omega), \quad (10)$$

which is the well-known reciprocity theorem for antennas. In addition, the radiation scattering operator obeys

$$\mathcal{S}^{(\mathrm{scat})}(\Omega, \Omega') = [\mathcal{S}^{(\mathrm{scat})}]^T(\Omega', \Omega), \quad (11)$$

and the pixel-to-pixel scattering matrix is reciprocal, $S^{(\mathrm{refl})} = [S^{(\mathrm{refl})}]^T$.

The output power emanating from a passive optical system cannot exceed the input power, which imposes an important constraint on the scattering operator: $I - S^\dagger S$ must be nonnegative definite. Combined with the reciprocity theorem, this can be used to demonstrate that the following matrix must also be nonnegative definite:

$$M_{i\alpha,j\beta} = \delta_{i\alpha,j\beta} - \int d\Omega [\mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega)]^* \cdot \mathbf{S}_{j\beta}^{(\mathrm{rec})}(\Omega)$$

$$- \sum_{k\gamma} [S_{k\gamma,i\delta}^{(\mathrm{refl})}]^* S_{k\gamma,j\beta}^{(\mathrm{refl})}. \tag{12}$$

In particular, the diagonal elements must be nonnegative, which implies

$$\int d\Omega |\mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega)|^2 \leqslant 1 - \sum_{j\beta} |S_{j\beta i\alpha}^{(\mathrm{refl})}|^2 \leqslant 1. \tag{13}$$

Thus the overall normalization of the receiving patterns is not arbitrary. If we demand that the optical system be lossless and if the detector modes are perfectly coupled [$S^{(\mathrm{refl})} = 0$], the receiving patterns must be orthonormal:

$$\int d\Omega [\mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega)]^* \cdot \mathbf{S}_{j\beta}^{(\mathrm{rec})}(\Omega) = \delta_{i\alpha,j\beta}. \tag{14}$$

We now calculate the power received by any detector pixel. We assume that any imperfect absorption (including reflection) associated with the detector pixel has been incorporated into the definition of S. Furthermore, we assume that the detectors do not radiate into the optical system, so that $a_{i\alpha} = 0$. (The detectors are usually operated at a temperature low enough that the thermal radiation they emit is negligible.) The power absorbed by pixel α is

$$Q_\alpha = \sum_i |b_{i\alpha}|^2 = \sum_i \left| \int d\Omega \mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega) \cdot \mathbf{a}(\Omega) \right|^2. \tag{15}$$

3. ASTRONOMICAL SOURCES

Astronomical sources emit radiation that is spatially and temporally incoherent. This means that we should regard the amplitude $\mathbf{a}(\Omega, \nu)$ at frequency ν as a complex random variable with mean zero and with a correlation function of the form

$$\langle a_q(\Omega, \nu) a_{q'}^*(\Omega', \nu') \rangle = A_{qq'}(\Omega, \nu)\delta(\Omega - \Omega')$$

$$\times \delta(\nu - \nu'). \tag{16}$$

Here $a_q(\Omega, \nu) = \hat{\mathbf{e}}_q^*(\Omega) \cdot \mathbf{a}(\Omega, \nu)$ and $q, q' \in \{1, 2\}$ are vector (polarization) indices, corresponding to two orthonormal polarization vectors $\hat{\mathbf{e}}_q(\Omega)$, each orthogonal to the propagation direction $\hat{\mathbf{n}}(\Omega)$. We note that the plane-wave expansion of the incoming field is unique in the sense that the correlation function is diagonal in the spatial variable. Had we chosen some other modal representation, e.g., a vector spherical harmonic expansion, the amplitude correlation matrix, in general, would not be diagonal in the mode indices.

The physical interpretation of $A_{qq'}(\Omega, \nu)$ follows from a calculation of the flux $F(\hat{\mathbf{p}}, \hat{\mathbf{s}}, \nu)$, the power per unit bandwidth per unit area, in a given polarization $\hat{\mathbf{p}}$ that is incident on a surface with normal $\hat{\mathbf{s}}$:

$$F(\hat{\mathbf{p}}, \hat{\mathbf{s}}, \nu) = \frac{1}{\lambda^2} \sum_{qq'} \int d\Omega \hat{\mathbf{n}}(\Omega) \cdot \hat{\mathbf{s}}$$

$$\times \left[\frac{\hat{\mathbf{p}}^* \cdot \hat{\mathbf{e}}_q(\Omega) A_{qq'}(\Omega, \nu) \hat{\mathbf{e}}_{q'}^*(\Omega) \cdot \hat{\mathbf{p}}}{|1 - \hat{\mathbf{n}}(\Omega) \cdot \hat{\mathbf{p}}|^2} \right]. \tag{17}$$

For sources occupying a small solid angle near zenith ($\hat{\mathbf{s}} = \hat{\mathbf{z}}$), the total flux for both polarizations simplifies to

$$F_{\mathrm{total}}(\nu) = \frac{1}{\lambda^2} \sum_q \int d\Omega A_{qq}(\Omega, \nu). \tag{18}$$

Thus $A_{qq}(\Omega, \nu)\lambda^{-2}$ is the specific intensity (flux per unit solid angle) arriving from the direction Ω in polarization $\hat{\mathbf{e}}_q(\Omega)$. For unpolarized emission, we can write

$$A_{qq'}(\Omega, \nu) = h\nu\, n(\Omega, \nu)\delta_{qq'}, \tag{19}$$

where $n(\Omega, \nu)$ is the mean photon occupation number. This simplifies to $A_{qq} = k_B T$ for a blackbody in the Rayleigh–Jeans limit.

4. RESPONSE OF AN OPTICAL SYSTEM TO AN ASTRONOMICAL SOURCE

The average power per unit bandwidth received from an astronomical source by a detector pixel can be calculated with Eqs. (15) and (16):

$$\langle Q_\alpha(\nu) \rangle = \sum_{qq'} \int d\Omega A_{qq'}(\Omega, \nu) \Bigg\{ \sum_i [\hat{\mathbf{e}}_q \cdot \mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega, \nu)]$$

$$\times [\mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega, \nu) \cdot \hat{\mathbf{e}}_{q'}]^* \Bigg\}. \tag{20}$$

For an unpolarized source, $A_{qq'}(\Omega, \nu) = A(\Omega, \nu)\,\delta_{qq'}$, this reduces to

$$\langle Q_\alpha(\nu) \rangle = \int d\Omega A(\Omega, \nu) R_\alpha(\Omega, \nu), \tag{21}$$

where we have defined the angular response function corresponding to this detector:

$$R_\alpha(\Omega, \nu) = \sum_i |\mathbf{S}_{i\alpha}^{(\mathrm{rec})}(\Omega, \nu)|^2. \tag{22}$$

For a source that has uniform brightness over the area sampled by the response function $R_\alpha(\Omega, \nu)$, the power received is $\langle Q_\alpha(\nu) \rangle = m_\alpha(\nu) A(\Omega, \nu)$, where the effective number of modes m_α (spatial and polarization) coupled to the detector is defined as

$$m_\alpha(\nu) = \int d\Omega R_\alpha(\Omega, \nu). \tag{23}$$

According to inequality (13), $m_\alpha(\nu) \leqslant \Sigma_i 1$, and so $m_\alpha(\nu)$ cannot exceed the number of modes received by the detector that are illuminated by the telescope.

For the opposite extreme, we take the case of a point source located at Ω^p. The power received by one detector is

$$\langle Q_\alpha(\nu) \rangle = \tfrac{1}{2} F(\nu)\lambda^2 R_\alpha(\Omega^p, \nu), \tag{24}$$

where $F(\nu)$ is the flux (both polarizations) of the point source. On the other hand, for a telescope system with total collecting area A_{tel}, the power collected by all the detectors cannot exceed $F(\nu) A_{\mathrm{tel}}$. Thus

$$\sum_\alpha R_\alpha(\Omega^p, \nu) \leq \frac{2A_{\mathrm{tel}}}{\lambda^2}. \qquad (25)$$

Since the response functions are all positive, each individual response function must also obey this inequality. Therefore if the response function $R_\alpha(\Omega, \nu)$ has a flat-top shape extending over a solid angle $\Delta\Omega_\alpha$, the effective number of modes obeys

$$m_\alpha \leq \frac{2A_{\mathrm{tel}}\Delta\Omega_\alpha}{\lambda^2}. \qquad (26)$$

This statement is often called the antenna theorem; one cannot increase the number of modes coupled to a given detector without simultaneously broadening the angular response function.

Inequality (25) suggests that we renormalize our response functions:

$$\rho_\alpha(\Omega, \nu) = \frac{\lambda^2}{2A_{\mathrm{tel}}} R_\alpha(\Omega, \nu), \qquad (27)$$

so that they obey

$$\sum_\alpha \rho_\alpha(\Omega, \nu) \leq 1. \qquad (28)$$

The power received by a detector [Eq. (21)] can now be expressed as

$$\langle Q_\alpha(\nu)\rangle = \frac{2A_{\mathrm{tel}}}{\lambda^2} \int d\Omega A(\Omega, \nu)\rho_\alpha(\Omega, \nu). \qquad (29)$$

We can discretize this integral by splitting the source into small patches or "pixels" centered at positions Ω_s, allowing each patch to have a different size $\Delta\Omega_s$ and assuming that the source has uniform intensity across each patch:

$$A(\Omega, \nu) = \sum_s \bar{A}(\Omega_s, \nu)U_s(\Omega), \qquad (30)$$

where the indicator function $U_s(\Omega)$ has a value of unity over the patch $\Delta\Omega_s$ and is zero otherwise [some sort of restriction on the form of $A(\Omega, \nu)$ is necessary, since a discrete set of data cannot uniquely determine a function of a continuous variable]. This gives

$$\langle Q_\alpha(\nu)\rangle = \sum_s \frac{2\Delta\Omega_s A_{\mathrm{tel}}}{\lambda^2} \bar{A}(\Omega_s, \nu)\bar{\rho}_\alpha(\Omega_s, \nu), \quad (31)$$

where the average of the response function over patch s is

$$\bar{\rho}_\alpha(\Omega_s, \nu) = \frac{1}{\Delta\Omega_s} \int_{\Delta\Omega_s} d\Omega \rho_\alpha(\Omega, \nu). \qquad (32)$$

Finally, using this result along with $A(\Omega, \nu) = h\nu\, n(\Omega, \nu)$ [Eq. (19)], we arrive at a very simple and illuminating expression for the average number of photons detected in a unit bandwidth during an integration time τ.

$$\langle N_\alpha\rangle = \int d\nu \sum_s \langle N(\Omega_s, \nu)\rangle\bar{\rho}_\alpha(\Omega_s, \nu), \qquad (33)$$

where

$$\langle N(\Omega_s, \nu)\rangle = \tau \frac{2\Delta\Omega_s A_{\mathrm{tel}}}{\lambda^2}\bar{n}(\Omega_s, \nu) \qquad (34)$$

is just the maximum average number of photons (per unit bandwidth) that a single-aperture telescope with area A_{tel} could detect in a time τ from the solid-angle patch $\Delta\Omega_s$. The interpretation of the normalized response function is simple: $\rho_\alpha(\Omega, \nu)$ is the probability that a photon of frequency ν that was emitted from position Ω and was collected by the instrument is actually detected by detector α. The total probability for detection is at most unity, according to inequality (28).

If we wish, we can take the additional step of discretizing the frequency integral into spectral channels $\Delta\nu_f$, which leads to

$$\langle N_\alpha\rangle = \sum_{s,f} \langle N(\Omega_s, \nu_f)\rangle\bar{\rho}_\alpha(\Omega_s, \nu_f), \qquad (35)$$

where the maximum average number of detectable photons in frequency channel $\Delta\nu_f$ is

$$\langle N(\Omega_s, \nu_f)\rangle = \tau\Delta\nu_f \frac{2\Delta\Omega_s A_{\mathrm{tel}}}{\lambda^2}\bar{n}(\Omega_s, \nu_f) \qquad (36)$$

and the response function is now also averaged over frequency:

$$\bar{\rho}_\alpha(\Omega_s, \nu_f) = \frac{1}{\Delta\Omega_s\Delta\nu_f} \int_{\Delta\nu_f} d\nu \int_{\Delta\Omega_s} d\Omega \rho_\alpha(\Omega, \nu). \quad (37)$$

Equations (35) and (36) tell us that the response of any photon direct-detection instrument, regardless of spectral or spatial resolution, whether single telescope or interferometer, etc., can be reduced to a probability matrix $p_{\alpha c}$. Here, for simplicity, the combined spatial–spectral channel index c replaces both indices s and f:

$$p_{\alpha c} = \bar{\rho}_\alpha(\Omega_s, \nu_f). \qquad (38)$$

This matrix describes the probability of a photon, which was emitted by the source in some spatial–spectral channel c and is collected by the instrument, to be absorbed in a given detector α. Let

$$\lambda_c = \langle N(\Omega_s, \nu_f)\rangle \qquad (39)$$

represent the mean number of incident photons from channel c arriving at the instrument. The mean number μ_α of photons detected by detector α is given by

$$\mu_\alpha = \langle N_\alpha\rangle = \sum_c p_{\alpha c}\lambda_c. \qquad (40)$$

A few comments can be made about the properties of the probability matrix P.

1. From inequality (28) we have

$$\sum_\alpha p_{\alpha c} \leq 1, \qquad (41)$$

so that the mean number of photons detected cannot exceed the mean number of incident photons, $\Sigma_\alpha \mu_\alpha \leq \Sigma_c \lambda_c$.

2. Since all elements are nonnegative, we also must have

$$0 \leq p_{\alpha c} \leq 1. \qquad (42)$$

3. The dimensions of the matrix are $N_{\text{detec}} \times N_{\text{chan}}$. Thus the rank of the probability matrix, $r(P)$, obeys $r(P) \leq N_{\text{chan}}$. Here N_{detec} is the number of detectors and N_{chan} is the number of spatial–spectral channels.

4. If we wish to determine uniquely the source intensity distribution λ_c in N_{chan} spatial–spectral channels, we should require P to have full rank: $r(P) = N_{\text{chan}}$. In particular, this condition implies $N_{\text{detec}} \geq N_{\text{chan}}$.

These properties and their implications will be discussed in more detail in Section 6.

5. RESPONSE OF SINGLE-MODE INTERFEROMETERS

Section 4 gives us the tools to rigorously calculate the photon-detection probability matrix for an interferometer. The primary reason for constructing interferometers is to obtain high spatial resolution; the overall field of view is often a secondary concern. Thus it is interesting to examine the case of an interferometer in which each telescope collects light from a single diffraction-limited beam, which sets the field of view. For simplicity, we will assume that the interferometer consists of T identical telescopes, each with area A_1. The total collecting area is $A_T = TA_1$. The receiving pattern corresponding to the single diffraction-limited beam of telescope t (where $t = 1 \ldots T$) will be denoted by $\mathbf{S}_t^{(\text{rec})}(\Omega, \nu)$, following our established notation. These receiving patterns are assumed to be identical, apart from the fact that the telescopes are located at different positions. We denote the telescope positions with respect to an arbitrary (but common) origin using the displacement vectors $\mathbf{r}_t$. The receiving patterns can then be written as

$$\mathbf{S}_t^{(\text{rec})}(\Omega, \nu) = \exp[+ik\hat{\mathbf{n}}(\Omega) \cdot \mathbf{r}_t] \mathbf{S}^{(\text{rec})}(\Omega, \nu). \quad (43)$$

Here the pattern $\mathbf{S}^{(\text{rec})}(\Omega, \nu)$ denotes the receiving pattern of a telescope located at the origin. Each of these telescopes produces a single-mode output beam, described by an outgoing wave amplitude b_t:

$$b_t(\nu) = \int d\Omega \mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \cdot \mathbf{a}(\Omega, \nu). \qquad (44)$$

The receiving patterns from different telescopes $t \neq t'$ are orthogonal to a high degree,

$$\int d\Omega [\mathbf{S}_t^{(\text{rec})}(\Omega, \nu)]^* \cdot \mathbf{S}_{t'}^{(\text{rec})}(\Omega, \nu)$$

$$= \int d\Omega |\mathbf{S}^{(\text{rec})}(\Omega, \nu)|^2 \exp[-ik\hat{\mathbf{n}}(\Omega) \cdot (\mathbf{r}_t - \mathbf{r}_{t'})]$$

$$\approx 0, \qquad (45)$$

because of the oscillations of the exponential factor. This means that the telescopes are not coupled significantly (as can happen for closely packed antenna arrays), and it

is therefore possible to achieve nearly perfect coupling to the single-mode outputs. For this case, the telescope patterns are orthonormal [see Eq. (14)]. If necessary, coupling losses can be accounted for in the beam combiner, which we introduce next.

The purpose of the beam combiner is to interfere the light from different telescopes before it is detected. There are various ways to do this: The telescopes can be combined in pairs, analogous to the way that radio-correlation interferometers operate, or all the light from the telescopes can be interfered simultaneously, as is done in Fizeau interferometry. Beam combination is a key issue for interferometer design. We can describe any type of beam-combining scheme using a scattering matrix $S^{(\text{comb})}$. This matrix tells us how the wave amplitudes b_t from the single-mode telescope feeds are coupled to the wave amplitudes $b_{i\alpha}$ arriving at the detectors:

$$b_{i\alpha}(\nu) = \sum_t S_{i\alpha,t}^{(\text{comb})} b_t(\nu). \qquad (46)$$

By combining Eqs. (44) and (46), we have

$$b_{i\alpha}(\nu) = \sum_t S_{i\alpha,t}^{(\text{comb})}(\nu) \mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \cdot \mathbf{a}(\Omega, \nu). \quad (47)$$

For astronomical sources, we are interested in the response function $R_\alpha(\Omega, \nu)$ [see Eq. (22)], which in this case is

$$R_\alpha(\Omega, \nu) = \sum_i \left| \sum_t S_{i\alpha,t}^{(\text{comb})}(\nu) \mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \right|^2. \quad (48)$$

These response functions must obey the inequality expressed in inequality (25), with the factor of 2 removed, since we are collecting only a single mode (single polarization). For single-mode interferometers, we would define

$$\rho_\alpha(\Omega, \nu) = \frac{\lambda^2}{A_T} R_\alpha(\Omega, \nu), \qquad (49)$$

and we would also omit the corresponding factor of 2 in Eq. (36). Using Eqs. (38), (48), and (49), we can calculate the photon-detection probability matrix $p_{\alpha c}$ for an interferometer. Although we have assumed identical telescopes, it is straightforward to generalize our expressions to include heterogeneous arrays.

Aperture synthesis imaging refers to the technique of repeatedly observing a given astronomical source with different configurations of the telescopes that form an interferometer. This is a standard technique in radio astronomy; interferometers are constructed with multiple possible positions or stations for each telescope, and the telescopes are physically moved to these different stations. In addition, aperture synthesis imaging usually relies on the rotation of the Earth, which, in essence, rotates the interferometer as a function of time with respect to the astronomical source. The source is assumed not to vary over the time taken to gather the aperture synthesis observations.

It is quite straightforward to take into account multiple telescope configurations within our formalism. The photon-detection probability matrix $p_{\alpha c;a}$ depends on the telescope configuration, which we label by the discrete index a (the different orientations produced by Earth rotation may be binned). We can replace the indices α and a

with a combined index β. If there are N_{config} telescope configurations and the interferometer has N_{detec} physical detectors, then β will take on $N_{\text{config}}N_{\text{detec}}$ different values, and we can consider β to be an index for a set of virtual detectors. This allows us to define a new photon-detection probability matrix that describes the entire set of aperture synthesis observations:

$$p_{\beta c} = p_{\alpha c;a} w_a, \qquad (50)$$

where the weighting factor w_a is simply the fraction of the total observing time that was spent in configuration a,

$$w_a = \frac{T_a}{\displaystyle\sum_{a'} T_{a'}}, \qquad (51)$$

where T_a is the time spent observing in configuration a. The properties of the photon-detection probability matrix discussed at the end of Section 4 also hold for the case of aperture synthesis (with N_{detec} reinterpreted as the number of virtual detectors).

6. POISSON DECONVOLUTION PROBLEM AND THE CRAMER–RAO BOUND

A. Introduction and Overview
We have seen that any optical instrument, including an interferometer used for aperture synthesis imaging, may be described by a photon-detection probability matrix P, with elements $p_{\alpha c}$. According to Eq. (40), this matrix relates the mean photon counts μ_α registered by the detectors to the source distribution λ_c. Some sort of inversion procedure must be applied to obtain an estimate $\hat{\lambda}_c$ of the source distribution from the observed counts N_α. How well can this deconvolution be done? What is the noise in the deconvolved image (or spectrum)? The answer to this question depends on

- the statistics of the photon counts,
- the deconvolution method that is used,
- the nature of the probability matrix P.

We will assume that the photon counts have independent Poisson distributions (as we discuss below). Clearly, we should expect in all cases that the variance of $\Delta\hat{\lambda}_c = \hat{\lambda}_c - \lambda_c$ obey the inequality

$$\sigma_c^2 = \langle(\Delta\hat{\lambda}_c)^2\rangle \geq \lambda_c, \qquad (52)$$

since the number of photons emitted from the source in channel c has a Poisson distribution with mean λ_c. This inequality will be derived, and the conditions on P that are necessary and sufficient to achieve this limit will be given.

A wide variety of deconvolution methods have been proposed[2]; in this paper we will discuss two: a simple linear least-squares (LS) method and the maximum-likelihood (ML) method. In general, the ML method is superior, but it is nonlinear and is therefore more difficult to compute and analyze. Since the LS method is not optimal, its performance provides only an upper bound to

the sensitivity. However, in the important special case in which all detectors receive the same photon flux (μ_α all equal), we will see that the LS upper bound actually coincides with the Cramér–Rao lower bound (described below), and therefore both give the actual sensitivity.

The Cramér–Rao theorem gives a rigorous lower bound for the sensitivity that holds for any deconvolution method. We will use this result to derive inequality (52) and to show that the ideal sensitivity limit can be achieved only if the instrument does not allow photons from different spatial–spectral channels to arrive at the same detector. In other words, an ideal instrument does not mix up photons from different channels, and a photon-detection event can be unambiguously assigned to the appropriate spatial–spectral channel. Consequently, such an instrument would achieve the lower bound on sensitivity given by inequality (52). The definition of an ideal instrument is given more precisely below. Unfortunately, interferometers, in general, do not obey our definition and do mix up photons spatially (and spectrally) and therefore cannot achieve the ideal sensitivity limit.

The Poisson deconvolution problem arises in many other contexts, such as positron-emission tomography for medical imaging[37]; as a result, there is extensive literature on this subject. In fact, iterative ML algorithms for Poisson deconvolution do exist, e.g., the Richardson–Lucy or the expectation maximization algorithm (see the review by Molina et al.[38] and the references therein). The existence of ML algorithms is important, since the noise performance of the ML method is guaranteed to asymptotically approach the Cramér–Rao bound in the limit of high signal-to-noise ratio (SNR). However, in cases that the likelihood function has multiple local maxima, the Cramér–Rao bound often fails to give a realistic estimate of the achievable sensitivity and does not give even a qualitative description of how the sensitivity varies with the SNR. This occurs because the estimation problem is ambiguous—the various local maxima represent alternate possible solutions. The Cramér–Rao bound becomes useful only when the SNR is large enough so that a unique solution can be chosen among these various possibilities. Examples include time delay and bearing estimation problems[39–41]; in such cases, other techniques such as the Ziv–Zakai bound must be used. Fortunately, this situation does not occur for the problem we are studying. We shall see that if the rank of the probability matrix is equal to the number of spatial–spectral channels [$r(P) = N_{\text{chan}}$], then a finite Cramér–Rao bound exists, the LS deconvolution also exists, and the likelihood function has a single maximum that defines the unique ML deconvolution. Conversely, if $r(P) < N_{\text{chan}}$, a finite Cramér–Rao bound does not exist, the LS deconvolution may fail, and the ML method may yield a degenerate subspace of possible solutions instead of a single unique solution.

Thus a finite Cramér–Rao bound should correspond to the actual sensitivity of an astronomical instrument or interferometer for bright objects or for long integrations; one simply needs to ensure that the photon counts are large, $N_\alpha \gg 1$. The Cramér–Rao bound can therefore provide a useful tool for quantitatively comparing the sensitivity of various interferometer design options.

B. Poisson Statistics

It is well established that the photon counts registered by the detectors in an optical instrument follow statistically independent Poisson distributions, so that the fluctuations of the counts in different detectors are uncorrelated. To be more precise, this situation holds for the case of thermal emission (from the source, the atmosphere, the telescope, etc.) in which the mean photon occupation numbers of the modes incident on the detectors are low, $n \ll 1$. In the high occupancy limit, $n \gg 1$, photon bunching becomes important in that it changes the counting statistics and can introduce correlations among the detectors. We will discuss only the first case, $n \ll 1$, which applies to most astronomical observations at optical and infrared wavelengths.

C. Definition of an Ideal Instrument

We define an ideal instrument in terms of its probability matrix $p_{\alpha c}$, which is required to obey the condition

$$p_{\alpha c} p_{\alpha c'} = 0 \tag{53}$$

for all detectors α and for all $c' \neq c$. This condition simply states that if $p_{\alpha c} \neq 0$, then $p_{\alpha c'} = 0$. In words, if a detector receives photons from the spatial–spectral channel c, it cannot receive photons from some other channel c'. This allows us to group the detectors into disjoint sets that may be indexed by the channel that the detectors respond to. Furthermore, we require that the total detection probability for any channel c be unity: $\Sigma_\alpha p_{\alpha c} = 1$. It is evident that such an instrument can achieve the ideal sensitivity limit set by inequality (52); to estimate the source intensity in channel c, we simply sum the photon counts from the set of detectors responding to channel c. By considering this set of detectors as an equivalent single detector, one can readily see that any ideal instrument is equivalent to an instrument that has a probability matrix equal to the identity matrix, $p_{\alpha c} = \delta_{\alpha c}$.

The spatial and spectral resolution of an ideal instrument would therefore be set by our definition of the spatial–spectral channels c, which is, in principle, arbitrary. Do ideal instruments really exist, at least in principle? Achieving the required spectral resolution without mixing up photons spectrally is not a fundamental difficulty; all one needs is a large enough grating. Achieving the required spatial resolution without mixing up photons spatially is a more subtle issue, since the parameters λ_c refer to a fixed telescope collecting area, according to the definition in Eqs. (36) and (39). However, we can use a single-aperture telescope, with a diameter sufficiently large to achieve the spatial resolution required, as long as we reduce the transmission (i.e., use a neutral-density filter) to keep the effective collecting area (and the parameters λ_c) constant. Of course, doing so would be foolish; this argument serves only as an existence proof.

D. Cramér–Rao Sensitivity Limit

The Cramér–Rao theorem provides a strict lower limit for the variance of a quantity that is estimated from a set of noisy measurements. This theorem can be applied to determine the minimum noise in the determination of the intensity in some spatial–spectral channel by use of an instrument with a nonideal response, one whose probability matrix differs from the identity matrix. The Hubble Space Telescope, whose initial point-spread function suffered from spherical aberration that has since been corrected, provides a particularly well-known example of the substantial sensitivity degradation that occurs as a result of a nonideal response, which cannot be undone by image restoration techniques such as the maximum-entropy algorithm.[42] In fact, the Cramér–Rao limit was applied to exactly this situation by Jakobsen *et al.*[43] The Cramér–Rao theorem has also been used to evaluate similar information loss effects in other imaging problems, for instance, gamma-ray imaging in nuclear medicine.[44]

We start by quickly reviewing the Cramér–Rao theorem.[45–47] Let us consider an experiment that delivers a set of measurements, denoted by the vector x, but that has scatter that is due to measurement noise. The measurement process can be described by a probability distribution $f(x|\theta)$ to obtain a result x. Here the vector θ represents the unknown parameters or quantities that the experiment is sensitive to, such as the source intensity distribution in our case. The usual goal is to determine one or more of these parameters from the measured data, say, θ_i. To do this, we must construct some estimator $\hat{\theta}_i(x)$ that uses the measured data vector x to estimate θ_i. For simplicity, we assume that this estimator is unbiased; the results could be generalized to include bias. First, we define the matrix

$$M_{ij}(\theta) = \left\langle \frac{\partial \ln f}{\partial \theta_i} \frac{\partial \ln f}{\partial \theta_j} \right\rangle$$
$$= \int \mathrm{d}x f(x|\theta) \frac{\partial \ln f(x|\theta)}{\partial \theta_i} \frac{\partial \ln f(x|\theta)}{\partial \theta_j}. \tag{54}$$

This is known as the Fisher information matrix and is symmetric and nonnegative definite. In fact, it is positive definite, unless there is some linear combination of parameters θ_i that the function $f(x|\theta)$ is completely independent of, in which case we should reparameterize to eliminate that linear combination. Thus we assume M has an inverse, M^{-1}, which is also positive definite. For a detailed analysis of the case in which the Fisher information matrix is singular, we refer the reader to a recent paper by Stoica and Marzetta.[48]

The Cramér–Rao theorem states that $C \geq M^{-1}$, where C is the covariance matrix of the estimators, $C_{ij} = \langle (\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) \rangle$, and the matrix inequality is understood to mean that $C - M^{-1}$ is nonnegative definite. In particular, the diagonal elements give

$$\sigma_i^2 = \langle (\hat{\theta}_i - \theta_i)^2 \rangle \geq (M^{-1})_{ii}. \tag{55}$$

A weaker limit can also be given:

$$\sigma_i^2 \geq (M^{-1})_{ii} \geq (M_{ii})^{-1}. \tag{56}$$

The second (weaker) limit can be obtained by considering the case in which all other parameters are known except for θ_i.

We now apply this to the photon-detection problem. The detector counts have independent Poisson distributions:

$$f(N|\lambda) = \prod_\alpha \frac{\mu_\alpha^{N_\alpha}}{N_\alpha!} \exp(-\mu_\alpha), \qquad (57)$$

so

$$\frac{\partial \ln f}{\partial \lambda_c} = \sum_\alpha \left(-p_{\alpha c} + \frac{N_\alpha p_{\alpha c}}{\mu_\alpha} \right), \qquad (58)$$

$$M_{cc'}(\lambda) = \left\langle \frac{\partial \ln f}{\partial \lambda_c} \frac{\partial \ln f}{\partial \lambda_{c'}} \right\rangle = \sum_\alpha \frac{p_{\alpha c} p_{\alpha c'}}{\mu_\alpha}$$

$$= \sum_\alpha \frac{p_{\alpha c} p_{\alpha c'}}{\sum_{c''} p_{\alpha c''} \lambda_{c''}}, \qquad (59)$$

by virtue of $\langle N_\alpha N_{\alpha'} \rangle = \mu_\alpha \mu_{\alpha'} + \mu_\alpha \delta_{\alpha \alpha'}$. The Cramér–Rao sensitivity limit for channel c is

$$\sigma_c^2 \geq (M^{-1})_{cc}. \qquad (60)$$

Equation (59) and inequality (60) are key results, since they give us a quantitative way to set lower limits to the sensitivity of optical instruments. Alternatively, for some purposes we may wish to use the weaker bound, $\sigma_c^2 \geq (M_{cc})^{-1}$, which is the result quoted by Jakobsen et al.[43]:

$$\sigma_c^2 \geq \left[\sum_\alpha \frac{p_{\alpha c}^2}{\sum_{c'} p_{\alpha c'} \lambda_{c'}} \right]^{-1}. \qquad (61)$$

It is easy to verify that either form gives $\sigma_c^2 = \lambda_c$ in the case of an ideal instrument, for which $p_{\alpha c} = \delta_{\alpha c}$. However, we note that although the Cramér–Rao bound given by inequality (60) is asymptotically achievable by the ML method, this is not true, in general, for the weaker bound.

These sensitivity limits depend on the values of λ_c, i.e., the structure of the source. For the case of a point source located in channel c, inequality (61) gives $\sigma_c^2 = \lambda_c (\Sigma_\alpha p_{\alpha c})^{-1} \geq \lambda_c$; the equality holds if all the photons are detected. The sensitivity limit given by the stronger Cramér–Rao bound [inequality (60)] is typically only somewhat worse than this. The reason that the point-source sensitivity does not vary much with the instrument response is that we are simply summing all the photons counted by all the detectors; it does not matter much how the photons are distributed among the detectors. Thus calculations that only compare the sensitivities of interferometers to point sources do not tell the whole story.

It is illuminating to write Eq. (59) with matrix notation:

$$M = P^T \Gamma^{-1} P, \qquad (62)$$

where $\Gamma = \mathrm{diag}(\mu_1, \mu_2, ...)$ is a diagonal matrix. From this expression, it is evident that the Fisher information matrix will be singular if the probability matrix P does not have full rank [$r(P) < N_{\mathrm{chan}}$].

It may appear that Eq. (62) presents a mathematical difficulty for the case of an ideal instrument observing a point source, for which $\mu_\alpha = 0$ for the off-source channels. A similar situation may also occur for nonideal instruments. However, one can imagine adding a low-level background to the point source, so that now $\mu_\alpha = \epsilon$ for the off-source channels. The Fisher information matrix M is now well defined and can be inverted to obtain the Cramér–Rao bound. Finally, the limit $\epsilon \to 0$ can be taken.

E. Sensitivity Bound and Ideal Instruments
We now prove the sensitivity bound given in inequality (52) and show that only ideal instruments can actually achieve this limit. For some channel c, the weaker bound states that

$$\frac{1}{\sigma_c^2} \leq M_{cc} = \sum_{\alpha \in D_c} \frac{p_{\alpha c}^2}{\mu_\alpha}, \qquad (63)$$

where $D_c = \{\alpha | p_{\alpha c} \neq 0\}$. Now

$$\mu_\alpha = \sum_{c'} p_{\alpha c'} \lambda_{c'} \geq p_{\alpha c} \lambda_c \geq 0, \qquad (64)$$

since all the terms in the sum are nonnegative. Thus

$$\frac{1}{\sigma_c^2} \leq \sum_{\alpha \in D_c} \frac{p_{\alpha c}^2}{p_{\alpha c} \lambda_c} = \sum_{\alpha \in D_c} \frac{p_{\alpha c}}{\lambda_c} \leq \frac{1}{\lambda_c}, \qquad (65)$$

since $\Sigma_\alpha p_{\alpha c} \leq 1$. This proves inequality (52). The conditions necessary to actually achieve this limit are evident from the derivation: $\mu_\alpha = p_{\alpha c} \lambda_c$ and $\Sigma_\alpha p_{\alpha c} = 1$. Since the source distribution λ_c is, in general, arbitrary, the first condition can be met only if $p_{\alpha c'} = 0$ for all $c' \neq c$. Thus the instrument must be an ideal instrument. We have therefore shown that only ideal instruments can (and do) achieve the sensitivity bound given by inequality (52).

F. Maximum-Likelihood Deconvolution
Deconvolution that uses the ML method is desirable, since the performance of the ML method is guaranteed to asymptotically approach the Cramér–Rao bound in the high SNR limit. Apart from a constant, the logarithm of the likelihood function for some estimate $\hat{\mu}$ of the mean count rate is [see Eq. (57)]:

$$\ln f = \sum_\alpha (N_\alpha \ln \hat{\mu}_\alpha - \hat{\mu}_\alpha). \qquad (66)$$

The goal of the ML method is to find a source distribution $\hat{\lambda}$ that maximizes this function, where $\hat{\mu} = P\hat{\lambda}$. In general, we would seek the solution by setting the gradient to zero, $\partial \ln f / \partial \hat{\lambda}_c = 0$. However, the case in which $N_\alpha = 0$ for one or more detectors must be treated with care. The difficulty may be illustrated by first considering $\ln f$ to be a function of $\hat{\mu}$ directly. The maximum of this function is unique and occurs at $\hat{\mu}_\alpha = N_\alpha$, and, in fact, $\partial \ln f / \partial \hat{\mu}_\alpha = -1$ for the detectors that have $N_\alpha = 0$. The problem is that the maximum occurs at the boundary of the physically acceptable region $\hat{\mu}_\alpha \geq 0$. We can remedy this problem by making all detector counts nonzero, replacing

$N_\alpha = 0$ with $N_\alpha = \epsilon$, where $\epsilon \ll 1$. We then proceed to find the ML solution using calculus; at the end, we take the limit $\epsilon \to 0$.

We therefore assume that $N_\alpha > 0$ for all detectors. The local maxima of the likelihood function must then obey [see Eq. (58)]

$$\frac{\partial \ln f}{\partial \hat{\lambda}_c} = \sum_\alpha p_{\alpha c}\left(-1 + \frac{N_\alpha}{\hat{\mu}_\alpha}\right) = 0 \qquad (67)$$

or, in matrix notation, $P^T v = 0$, where the components of the vector v are $v_\alpha = -1 + N_\alpha/\hat{\mu}_\alpha$. Thus v must be in the null space of P^T, which is the annihilator of the range of P. If there are N_{detec} detectors and P has rank $r(P)$, v is constrained to lie in a subspace of dimension $N_{\text{detec}} - r(P)$. In addition, $\hat{\mu}$ must lie in the range of P, as $\hat{\mu} = P\hat{\lambda}$, which gives $r(P)$ additional constraints. Overall, we would get N_{detec} constraint equations for v, which could be solved to yield the local maxima of the likelihood function. However, since some of the constraints apply linearly to v and others apply linearly to $\hat{\mu}$, the combined set of equations are, in fact, nonlinear. As mentioned earlier, iterative algorithms for finding the solution(s) do exist.

We now demonstrate that, in general, there is only a single local maximum of the likelihood function. Suppose instead that there were two distinct local maxima, corresponding to two different source distributions $\hat{\lambda}^{(a)}$ and $\hat{\lambda}^{(b)}$. We can write $\hat{\mu}^{(a)} = P\hat{\lambda}^{(a)}$ and $\hat{\mu}^{(b)} = P\hat{\lambda}^{(b)}$ and introduce the vectors $v^{(a)}$ and $v^{(b)}$, defined by their components

$$v_\alpha^{(a)} = -1 + \frac{N_\alpha}{\hat{\mu}_\alpha^{(a)}}, \qquad (68)$$

and similarly for $v_\alpha^{(b)}$. Since these correspond to local maxima, we know that we must have $P^T v^{(a)} = P^T v^{(b)} = 0$, which tells us that both $v^{(a)}$ and $v^{(b)}$ must lie in the annihilator of the range of P. Since $\hat{\mu}^{(a)}$ and $\hat{\mu}^{(b)}$ lie in the range of P, $v^{(a)}$ and $v^{(b)}$ must each be orthogonal to both $\hat{\mu}^{(a)}$ and $\hat{\mu}^{(b)}$:

$$\sum_\alpha \hat{\mu}_\alpha^{(a)} v_\alpha^{(a)} = \sum_\alpha \hat{\mu}_\alpha^{(b)} v_\alpha^{(b)} = 0, \qquad (69)$$

$$\sum_\alpha \hat{\mu}_\alpha^{(a)} v_\alpha^{(b)} = \sum_\alpha \hat{\mu}_\alpha^{(b)} v_\alpha^{(a)} = 0. \qquad (70)$$

These conditions tell us that

$$\sum_\alpha \hat{\mu}_\alpha^{(a)} = \sum_\alpha \hat{\mu}_\alpha^{(b)} = \sum_\alpha N_\alpha \qquad (71)$$

and also that

$$\sum_\alpha N_\alpha\left[\frac{\hat{\mu}_\alpha^{(a)}}{\hat{\mu}_\alpha^{(b)}} - 1\right] = \sum_\alpha N_\alpha\left[\frac{\hat{\mu}_\alpha^{(b)}}{\hat{\mu}_\alpha^{(a)}} - 1\right] = 0. \qquad (72)$$

Define $x_\alpha = \hat{\mu}_\alpha^{(a)}/\hat{\mu}_\alpha^{(b)} \geqslant 0$. Adding the equations above yields

$$\sum_\alpha N_\alpha g(x_\alpha) = 0, \qquad (73)$$

where $g(x) = x - 2 + 1/x$. Now, the function $g(x)$ has a single minimum in the domain $x \geqslant 0$, located at $x = 1$, and $g(1) = 0$. Since $N_\alpha > 0$ and $g(x) \geqslant 0$, we must actually have $g(x_\alpha) = 0$ and so all $x_\alpha = 1$, and therefore $\hat{\mu}_\alpha^{(a)} = \hat{\mu}_\alpha^{(b)}$. Thus $P\hat{\lambda}^{(a)} = P\hat{\lambda}^{(b)}$. If P has full rank $[r(P) = N_{\text{chan}}]$, we can conclude that $\hat{\lambda}^{(a)} = \hat{\lambda}^{(b)}$, and so, in fact, there is only a single local maximum, and the ML deconvolution is unique. If P does not have full rank, the solution is ambiguous, and $\hat{\lambda}^{(a)} = \hat{\lambda}^{(b)} + l$, where l can be any vector in the null space of P. As discussed above, the Fisher information matrix will be singular in this situation, so the calculation of the Cramér–Rao bound will fail. Also, LS deconvolution will fail, since $P^T P$ will not have an inverse (see below).

G. Least-Squares Deconvolution
To obtain a simpler (but suboptimal) linear method, we could attempt to set $v = 0$, which would give

$$\hat{\mu}_\alpha = \sum_c p_{\alpha c}\hat{\lambda}_c = N_\alpha. \qquad (74)$$

Since, in general, $N_{\text{detec}} > r$, this system of linear equations overconstrains $\hat{\lambda}$ but may be solved in a LS sense. Another way of explaining this is that the vector $\hat{\mu}$ must lie in the range of P, but, in general, the vector N does not. The LS algorithm is obtained by first projecting N into the range of P and then solving the resulting linear system for $\hat{\lambda}$:

$$\hat{\lambda}^{(\text{LS})} = (P^T P)^{-1} P^T N. \qquad (75)$$

In general, $P^T P$ is symmetric and nonnegative definite. In fact, it is positive definite and has an inverse if and only if P has full rank $[r(P) = N_{\text{chan}}]$. Thus the existence of a finite Cramér–Rao bound guarantees both that the LS deconvolution exists and that the likelihood function has a single maximum.

Because the LS deconvolution is linear, it is straightforward to calculate its bias and its noise performance. In fact, the LS method is unbiased, since

$$\langle \hat{\lambda}^{(\text{LS})}\rangle = (P^T P)^{-1} P^T \mu = (P^T P)^{-1} P^T P\lambda = \lambda. \qquad (76)$$

The covariance matrix of the LS deconvolution is given by

$$C^{(\text{LS})} = \langle \Delta\hat{\lambda}^{\text{LS}}(\Delta\hat{\lambda}^{\text{LS}})^T\rangle = (P^T P)^{-1} P^T \Gamma P(P^T P)^{-1}, \qquad (77)$$

where $\Gamma = \text{diag}(\mu_1, \mu_2, ...)$, as defined earlier. Since the LS method is not optimal, in general, this covariance matrix can be used to establish an upper bound to the sensitivity, to complement the Cramér–Rao lower bound.

In the case that $\mu_\alpha = \mu$ are all equal, $\Gamma = \mu I$, we see that the covariance matrix simplifies to

$$C^{(\text{LS})} = \mu(P^T P)^{-1}. \qquad (78)$$

Meanwhile, the Fisher information matrix [Eq. (62)] for this case is $M = \mu^{-1} P^T P$, so, in fact, the performance of LS deconvolution achieves the Cramér–Rao bound. This is admittedly a special case. However, it may often occur in practice, at least approximately, e.g., for a uniform source or when the photon counts are dominated by the

contribution of a uniform background such as thermal emission from the atmosphere or the telescopes.

Another interesting special case occurs when the probability matrix P is square and has full rank, $r(P) = N_{\text{chan}}$. Under these conditions, P can be inverted, and one can show that the LS and ML estimators are, in fact, the same:

$$\hat{\lambda}^{(\text{LS})} = \hat{\lambda}^{(\text{ML})} = P^{-1}N. \qquad (79)$$

Also, both estimators achieve the Cramér–Rao bound.

H. Interferometry, Aperture Synthesis, and the Cramér–Rao Bound

Astronomical aperture synthesis observations often do not obtain enough information to uniquely determine the source intensity distribution (map) at the desired resolution. In other words, the condition $r(P) < N_{\text{chan}}$ applies, and so a finite Cramér–Rao bound does not exist. In such cases, a linear or nonlinear regularization method (such as the maximum-entropy method) may be applied to select a single representative map that has certain desirable characteristics (e.g., smoothness, positivity) out of the infinite set of maps that are consistent with the measured data.

What good is the Cramér–Rao bound in such cases? All the regularization techniques have the general feature of varying the effective local spatial (or spectral) resolution, depending on the interferometer response and the SNR. One might therefore redefine the spatial–spectral channels *a posteriori*, reducing N_{chan} by degrading the resolution in the appropriate portions of the map, so as to achieve the condition $r(P) = N_{\text{chan}}$ required for the calculation of the Cramér–Rao bound. Alternatively, a singular value decomposition of the probability matrix could be used to determine which components of the source distribution can be recovered from the data.[48]

However, the utility of the Cramér–Rao bound transcends its application to any one particular set of aperture synthesis observations. The real power of the Cramér–Rao bound is that it provides a rigorous way to compare the sensitivity of different interferometer designs and observing strategies. An *a priori* choice of the spatial–spectral channels sets the problem: What area must be mapped and at what resolution? This choice is ultimately dictated by the science we wish to do. Once the channels are defined, we can calculate the Cramér–Rao bound for any interferometer design, assuming some set of telescope configurations. If the bound does not exist, the set of configurations must be expanded or the interferometer design must be modified and the calculation repeated. The optimization of the observing strategy for a particular interferometer design amounts to determining the set of configurations, and their time allocations, that gives the best Cramér–Rao bound.

7. APPLICATION TO ONE-DIMENSIONAL ARRAYS

A. Introduction

In Section 6 we have seen that the only way to achieve the best possible sensitivity for a measurement of the spectral and spatial intensity distribution of a source is to build an instrument that separates the photons into separate spectral–spatial channels prior to detection. Unfortunately, interferometers with separated telescopes generally cannot achieve this goal, since the instantaneous (not synthesized) angular response function will not be highly localized, as is the case for a single-aperture telescope, but will instead have multiple sidelobes or fringes. In the case of a two-element interferometer, this response function is just the single-telescope pattern, modulated by the interference fringes that correspond to the baseline between the two telescopes. Thus, although these fringes can provide much higher angular resolution than the individual telescopes, the spatial information provided by each detected photon is less than would have been obtained with a more localized response function.

It is important to determine quantitatively the magnitude of this effect, to be able to compare various options in the design of an interferometer, such as the number of telescopes, their configurations, and the method of beam combining. This comparison can be done with the Cramér–Rao lower bound. In this section we present the results of a numerical calculation of the Cramér–Rao sensitivity limits for a one-dimensional aperture synthesis array. A homogeneous one-dimensional array provides a nice case study because the parameter space is limited and the computations are fast, and yet the principal implications of the Cramér–Rao bound can be readily demonstrated. In addition, we calculate an upper bound to the sensitivity derived from the LS method. A comparison of these two bounds demonstrates that, in fact, the Cramér–Rao approach is highly useful for this problem.

B. Interferometer Response

The interferometer consists of T identical equally-spaced telescopes spread along the x axis, each with length L_1, and the overall collecting length is $L_T = TL_1$. We choose uniform illumination, so that the single-element amplitude pattern has the form

$$S^{(\text{rec})}(\theta) = \left(\frac{L_1}{\lambda}\right)^{1/2} \text{sinc}\left(\frac{\pi L_1}{\lambda}\theta\right). \qquad (80)$$

Here θ is the angle from zenith, and we have assumed that $L_1 \gg \lambda$, which allows a small-angle approximation to be made. The full width at half-power of this single-element pattern is $0.886\lambda/L_1$. For our comparison, we use $L_1 = 1000\lambda$, although the normalized sensitivities do not depend on the telescope size. For telescope t located at position x_t with respect to the origin,

$$S_t^{(\text{rec})}(\theta) = \exp\left(i\frac{2\pi x_t}{\lambda}\theta\right)S^{(\text{rec})}(\theta). \qquad (81)$$

The response of the interferometer for a given choice of beam combination is still given by Eq. (48) (note that the modal sum over i can be omitted, since we assume single-mode detectors); however, the proper normalization for the detection probability function is

$$\rho_\alpha(\theta, \nu) = \frac{\lambda}{L_T}R_\alpha(\theta, \nu). \qquad (82)$$

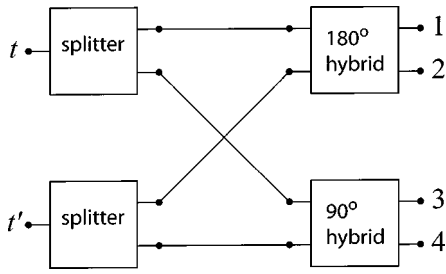For the numerical examples, we assume a single narrow frequency channel and discretize the spatial variable θ

Fig. 1. Schematic diagram of the pairwise beam combination scheme for a single baseline between telescopes t and t'. The inputs t and t' on the left represent the single-mode beams from the two telescopes, after division $T - 1$ ways. The four outputs on the right are sent to photon-counting detectors.

into N_{chan} uniform pixels, labeled by the index $c = 1, \ldots, N_{\text{chan}}$, extending across the field of view of a single telescope, $-0.75\lambda/L_1 \leqslant \theta \leqslant 0.75\lambda/L_1$. The photon-detection probability matrix $p_{\alpha c}$, defined by Eq. (38), is calculated with a one-dimensional analog of Eq. (32).

C. Pairwise Combination

We compare two types of beam combination. The first case is the usual pairwise combination, in which we focus on measuring the fringe visibilities and fringe phases that can be obtained from the $T(T - 1)/2$ telescope baseline pairs. To do this, the light from each telescope must first be split into $T - 1$ beams. We assume that the fringe measurement is done using four detectors per baseline, as shown in Fig. 1. The scattering matrix of the beam combiner produces the following four linear combinations:

$$b_1(t, t') = \frac{1}{2\sqrt{T - 1}}(b_t + b_{t'}),$$

$$b_2(t, t') = \frac{1}{2\sqrt{T - 1}}(b_t - b_{t'}),$$

$$b_3(t, t') = \frac{1}{2\sqrt{T - 1}}(b_t + ib_{t'}),$$

$$b_4(t, t') = \frac{1}{2\sqrt{T - 1}}(ib_t + b_{t'}).$$

Here b_t and $b_{t'}$ represent the single-mode wave amplitudes corresponding to the light collected by telescopes t and t', and the four amplitudes $b_k(t, t')$ (here $k = 1, \ldots, 4$) represent the four combinations of light from the two telescopes that are being detected to produce the corresponding photon counts $N_k(t, t')$. To avoid repeating telescope pairs, we require $T \geqslant t > t' \geqslant 1$. The total number of detectors in this scheme is $2T(T - 1)$. The detector index $\alpha = 1, \ldots, 2T(T - 1)$ uniquely specifies a baseline pair (t, t') as well as one of the four beam-combiner outputs k; we can set $\alpha(t, t', k) = [4(T - 1) - 2t')](t' - 1) + 4(t - 2) + k$. For example, Eq. (48) reads

$$R_\alpha(\Omega, \nu) = \frac{1}{4(T - 1)}\left|S_t^{(\text{rec})}(\theta) + iS_{t'}^{(\text{rec})}(\theta)\right|^2 \quad (83)$$

for the case that $\alpha = \alpha(t, t', k = 3)$.

It is straightforward to verify that the first two beam combinations produce symmetric angular response functions, whereas the latter two produce antisymmetric response functions (apart from a constant offset term). Thus both types of beam combination are needed to uniquely determine the image of a source. The latter two beam combinations are readily produced with a 50% beam splitter; at microwave frequencies, the equivalent device is known as a 90° 3-dB hybrid. The first two combinations can be obtained by use of the optical equivalent of a microwave 180° 3-dB hybrid; such devices are currently being investigated for nulling interferometry.[49,50] All four combinations can be gotten simultaneously by use of a two-way power splitter (or a 50% beam splitter), as shown in Fig. 1. It is straightforward to verify that for this beam combination scheme all the power (photons) collected by the telescopes is absorbed by the detectors. Equivalently, the scattering matrix of the beam combiner is unitary and only couples input ports to output ports.

D. Butler Combination

As shown in Fig. 2, the second type of beam combination method we investigate is the standard Butler matrix beam-forming network,[51] which is used with microwave phased-array antennas to produce a set of localized beams, each pointing in a different direction. This approach is analogous to the image-plane beam recombination used in Fizeau interferometry. Although similar free-space optical techniques could be used for microwave beam forming, Butler beam formers use guided-wave components (coaxial or waveguide) and are therefore much smaller physically.

The key idea behind the Butler matrix is to produce a linear-stepped phase gradient across the antenna or telescope array in order to steer the beam of the array. Mathematically, the T outputs b_α, which are sent to the photon-counting detectors, are given in terms of the T single-mode inputs b_t from the telescopes by

$$b_\alpha = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} b_t \exp\left(i\frac{2\pi\alpha t}{T}\right), \quad (84)$$

which, in essence, is just a discrete Fourier transform. The Butler matrix is actually a hardware implementation of this concept that is analogous to the fast-Fourier-transform algorithm. Power conservation, or unitarity of the beam-combiner scattering matrix, follows from Parseval's theorem. The typical angular response functions for pairwise and Butler combining are compared in Fig. 3.
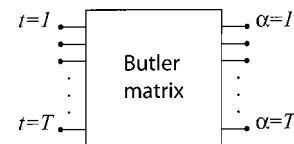

Fig. 2. Schematic diagram of the Butler matrix all-on-one beam combination scheme. The inputs on the left represent the single-mode beams from all T telescopes; each of the T outputs on the right, which are sent to photon-counting detectors, contain some contribution from all T telescope inputs.
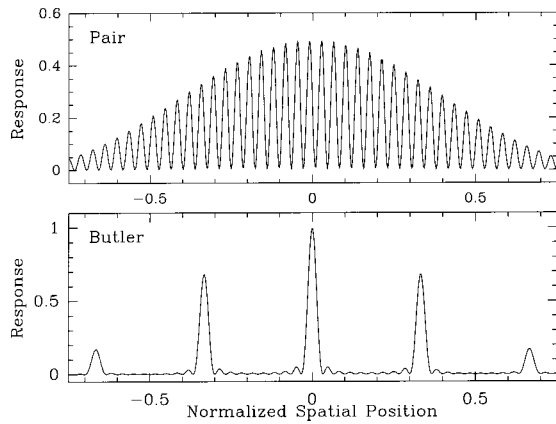
Fig. 3.   Comparison of the angular response functions for pair-wise and Butler beam combining. Top, typical response corresponding to two telescopes in a pair-combined array; the separation between the telescopes is $27L_1$ for this example. Bottom, typical response of a Butler-combined array; in this case, there are 10 uniformly spaced telescopes, with a distance $3L_1$ between telescopes, so that the array size is $27L_1$.

Table 1.   Array Configurations[a]

T^b	$N_{\text{spac}}{}^c$	$S_{\min}{}^d$	$S_{\max}{}^d$	ΔS^d	$B_{\max}{}^e$
3	44	1.0	22.5	0.5	45.0
6	17	1.0	9.0	0.5	45.0
10	9	1.0	5.0	0.5	45.0

[a] All dimensions are scaled to the telescope size L_1.
[b] The number of telescopes.
[c] The number of element spacings.
[d] Minimum, maximum, and step size for the element spacings.
[e] The maximum baseline.

E.   Numerical Calculations

We consider array configurations of three, six, and ten uniformly spaced telescopes, as shown in Table 1. Since the maximum baseline of all the array configurations shown in Table 1 is $B_{\max} = 45L_1$ and the maximum baseline controls the spatial resolution, we use $N_{\text{chan}} = 41$ spatial pixels in the calculations.

As shown in Table 1, the calculation also includes N_{spac} different telescope spacings, ranging from close-packed to dilute arrays. The source is observed repeatedly with different telescope spacings; this is an example of aperture synthesis in one dimension. For example, we assume that the three-telescope array observes the source in a total of $N_{\text{spac}} = 44$ different configurations. The first configuration corresponds to a close-packed array, in which the telescope spacing is $S = L_1$. The various configurations are obtained by incrementing the telescope spacing S by ΔS; thus the second configuration has $S = 1.5L$, the third has $S = 2.0L_1$, and so on. The most dilute configuration has $S = 22.5L_1$. The corresponding positions of the three telescopes are $x_1 = -S$, $x_2 = 0$, and $x_3 = S$.

To account for these various configurations, we imagine that there are actually $N_{\text{spac}}N_{\text{detec}}$ different virtual detectors, now indexed by β, where N_{detec} is the physical number of detectors (labeled by α) in any one configuration. Note that we must have $N_{\text{chan}} < N_{\text{spac}}N_{\text{detec}}$; otherwise,

the probability matrix P cannot have full rank, and the Fisher information matrix [Eq. (59)] will be singular. As described in the discussion of aperture synthesis at the end of Section 5, the detection probabilities for any one configuration are reduced by the factor $(N_{\text{spac}})^{-1}$. We are splitting up the total observing time into N_{spac} different sessions of equal duration, one per configuration; the total probability over the course of the entire integration for a photon to be detected by the array in some particular configuration is $(N_{\text{spac}})^{-1}$. As a check, we verified that in all cases the total combined probability for detecting photons from the central spatial pixels c in any of the virtual detectors β was near unity: $\Sigma_{\beta=1}^{(N_{\text{spac}}N_{\text{detec}})} p_{\beta c} \approx 1$.

We also consider two types of source: uniform sources and point sources. For uniform sources, we set $\lambda_c = 1$ for all N_{chan} spatial pixels; for point sources, we set $\lambda_c = 1$ only for the central pixel and set all others to zero.

F.   Results

Figure 4 shows the results for a uniform source with Butler beam combination. The vertical axis is the normalized Cramér–Rao sensitivity bound, calculated with inequality (60). This sensitivity would be unity for an ideal instrument, that is, a single telescope with an aperture large enough to resolve the spatial pixels but with the same total effective collecting area as the interferometer (i.e., with a neutral-density filter or attenuator to reduce the total number of detected photons to match the interferometer). A sensitivity above unity implies that the use of an interferometer incurs a penalty owing to its inability to fully determine from which spatial pixel each detected photon came. As Fig. 4 shows, adding more telescopes improves the sensitivity; the reason for this is that the quality of the instantaneous beam pattern improves. For this particular example, the sensitivity pen-
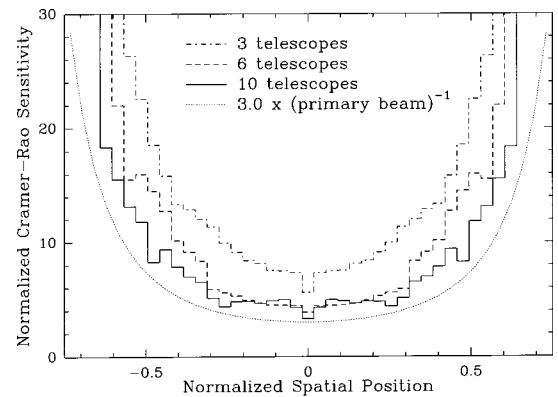

Fig. 4.   Variation of the normalized sensitivity with telescope array size for the case of Butler beam combination. The source is assumed to have a uniform spatial distribution. The horizontal axis gives the spatial position θ in units of λ/L_1; note that the full width at half-power of the single-element pattern is $0.886\lambda/L_1$. The vertical axis gives the Cramér–Rao normalized sensitivity bound (see text for details). For Butler beam combination, increasing the array size improves the normalized sensitivity. The dotted curve shows that the sensitivity degradation toward the edges of the field of view scales as the reciprocal of the single-element beam pattern. The upper sensitivity limit calculated with the LS deconvolution method is indistinguishable from Cramér–Rao bound; thus this plot gives the actual achievable sensitivity.
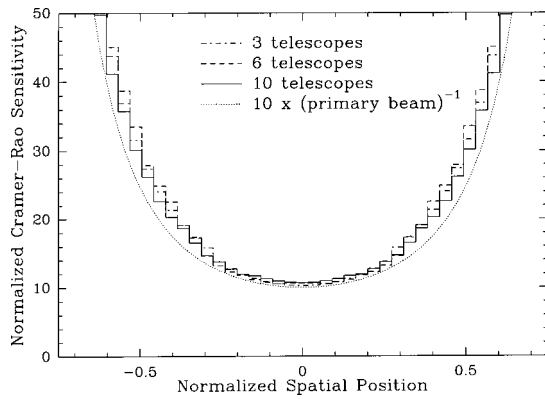
Fig. 5. Similar to Fig. 4 but calculated for pairwise beam combination. Normalized sensitivity is essentially independent of array size. Again, the upper limit from the LS method is calculated to be the same as the Cramér–Rao lower bound.
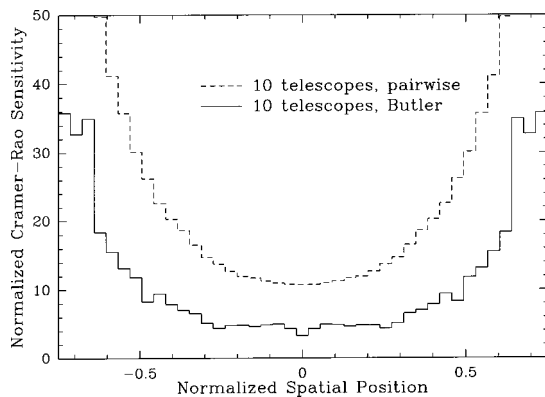

Fig. 6. Comparison of normalized sensitivities for Butler versus pairwise beam combining for a ten-element array when uniform sources are observed; the sensitivity advantage for Butler combining is more than a factor of 3.

alty for the Butler-combined interferometer is a factor of ~3 for a ten-telescope array. The variation of the sensitivity across the field of view is seen to scale with the inverse of the single-element power pattern, as shown by the dotted curve in Fig. 4. The reason for this is that the Cramér–Rao sensitivity bound includes the effects of noise cross talk between the spatial pixels that arise from the nonideal interferometer beam patterns. For this example, the upper limit to sensitivity derived from the LS deconvolution method [by use of Eq. (77)] coincides with the lower limit set by the Cramér–Rao bound; thus either gives the true sensitivity.

Figure 5 shows the comparable uniform-source results with pairwise beam combination. In this case, the normalized sensitivity (which takes out the effect of the total collecting area) shows no improvement as the array size is increased. This is because the quality of the instantaneous beam patterns remains unchanged: The beam patterns are always those of two-element interferometers. Again, the sensitivity scales inversely with the single-element pattern; however, the sensitivity penalty relative to an ideal instrument is now a factor of ~10. Again, the LS and Cramér–Rao bounds coincide, so either yields the true sensitivity. The comparison between the Butler-combined and the pairwise-combined ten-element arrays for uniform sources is shown more directly in Fig. 6; the

Butler-combined array enjoys a sensitivity advantage in excess of a factor of 3 for this example.

A similar comparison for the case of point sources is shown in Figs. 7–9. There now is a difference between the LS upper bound and the Cramér–Rao lower bound for the case of Butler combining (Fig. 7) but not for pairwise combining (Fig. 8). However, we see that the LS upper bound still allows us to demonstrate the superiority of arrays with larger numbers of elements and of Butler combining over pairwise combining.

The normalized Cramér–Rao sensitivities for the point source itself are quite comparable in all cases, which we expect, because we can, in essence, sum all the photons detected to estimate the brightness of the source. However, the sensitivity for the off-source pixels tells a much different story. Here Butler combination enjoys a large sensitivity advantage, approximately an order of magnitude for the ten-element array. Note that for the Butler-combined arrays, the sensitivity of the off-source pixels is actually substantially better than for the on-source pixels. This is highly desirable: It gives the array more sensitivity to see faint sources in the presence of a brighter
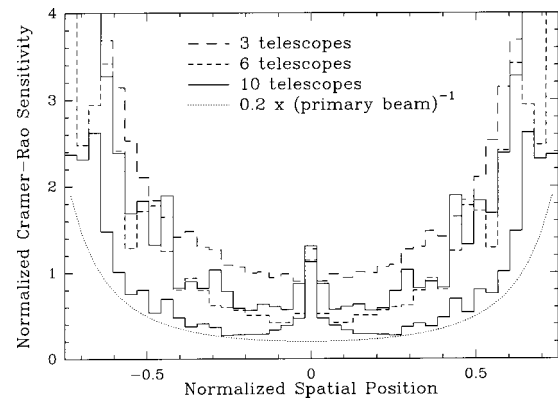

Fig. 7. Similar to Fig. 4 but calculated for a point source in the center of the field instead of for a uniform source. Note that although the normalized sensitivity to the point source is nearly constant, the sensitivity for off-source pixels improves substantially with array size. The upper limit from the LS method is somewhat worse than the Cramér–Rao lower bound and is shown as the upper solid curve for the ten-element array. This curve lies near the Cramér–Rao lower bound for the six-element array and below the bound for the three-element array.
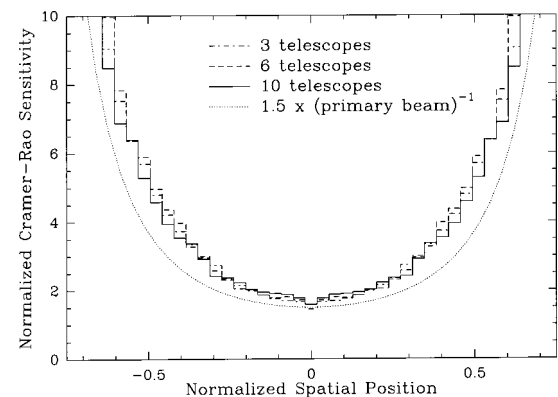

Fig. 8. Similar to Fig. 7 but with pairwise beam combination assumed; there is no sensitivity improvement with array size. The sensitivity of the LS method is indistinguishable from the Cramér–Rao bound, so the plot gives the actual sensitivity.
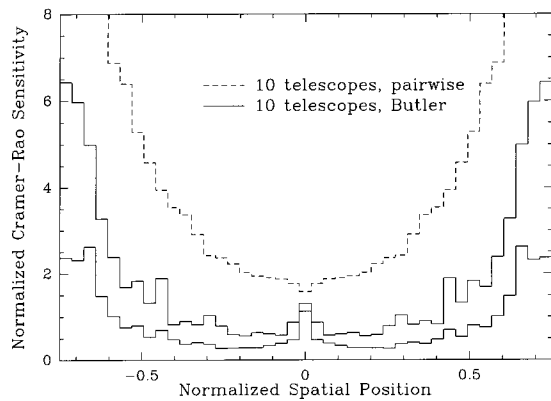
Fig. 9.   Comparison of normalized point-source sensitivities for Butler combination and pairwise combination for a ten-element array.   Although the sensitivities to the point source itself are comparable, Butler combination is an order of magnitude more sensitive for off-source pixels.   The upper solid curve gives the LS upper sensitivity limit for the Butler-combined array, and it lies well below the Cramér–Rao lower bound for the pairwise-combined array.

nearby object.   An ideal instrument, such as a large single-aperture telescope, would, in fact, have $\sigma_c = 0$ for the off-source pixels, since the detectors corresponding to these pixels do not receive any photons.   (Of course, this is not entirely true for real telescope systems owing to scattered light.)

8.   CONCLUDING REMARKS

In this paper we have made the case that the instantaneous angular response functions of an interferometer govern its sensitivity: Interferometers with more compact and localized response functions are more sensitive. The physical reason for this is simple and clear: Such interferometers obtain more spatial information per photon detected.   We have demonstrated this effect by numerically calculating the Cramér–Rao sensitivity limits for the simple case of homogeneous, equally spaced, one-dimensional aperture synthesis arrays that use either Butler or pairwise beam combining.   These calculations show that Butler beam combining, which is analogous to the image-plane combination used in Fizeau interferometry, is substantially more sensitive, which we expect, since the response functions are more compact.   In addition, our example shows that when telescopes are added to the array the sensitivity of a Butler-combined interferometer improves much more rapidly than that of the total collecting area.   However, it is important to remember that the imperfect beam patterns of sparse-aperture interferometers extract a sensitivity penalty as compared with filled-aperture telescopes, even after accounting for the differences in collecting areas.

The Cramér–Rao bound appears to be a very interesting and useful tool for the study and optimization of interferometer designs.   The approach presented in this paper can readily be applied to two-dimensional arrays that are of any configuration and that use any type of beam combination method.   Our approach, which is based on scattering matrices, is especially well suited to describe beam combination by use of guided-wave (inte-

grated) optics.   Although we have discussed only homogeneous arrays, in which all telescopes are identical, the formalism can readily be adapted to handle heterogeneous arrays.   The Cramér–Rao sensitivity bounds were obtained for very idealized circumstances, in which we have included only the counting statistics of the photons arriving from the source.   However, it is again not difficult to extend our results to include effects such as background noise, which is due to thermal emission from the telescope or atmosphere and from detector dark current or read noise or both.   Although these effects are important in real applications and should be included in more realistic calculations, we have ignored them to limit the parameter space and to focus on the fundamental issues involved.   In closing, we encourage other researchers who are involved in interferometer design to investigate the applicability of the Cramér–Rao approach for determining the sensitivity trade-offs for real arrays and for a range of astrophysical problems.   As an example, it would be interesting to study the exoplanet detection problem, under realistic constraints such as a fixed total collecting area.

REFERENCES

1.   A. Quirrenbach, "Optical interferometry," Annu. Rev. Astron. Astrophys. **39**, 353–401 (2001).
2.   S. K. Saha, "Modern optical astronomy: technology and impact of interferometry," Rev. Mod. Phys. **74**, 551–600 (2002).
3.   See http://olbin.jpl.nasa.gov.
4.   See http://www.nrao.edu.
5.   M. Shao, "SIM: the space interferometry mission," in *Astronomical Interferometry*, R. D. Reasenberg, ed., Proc. SPIE **3350**, 536–540 (1998), see also http://sim.jpl.nasa.gov.
6.   C. A. Beichman, "Terrestrial Planet Finder: the search for life-bearing planets around other stars," in *Astronomical Interferometry*, R. D. Reasenberg, ed., Proc. SPIE **3350**, 719–723 (1998), see also http://planetquest.jpl.nasa.gov.
7.   A. J. Penny, A. Leger, J. Mariotti, C. Schalinski, C. Eiroa, R. J. Laurance, and M. Fridlund, "Darwin interferometer," in *Astronomical Interferometry*, R. D. Reasenberg, ed., Proc. SPIE **3350**, 666–671 (1998), see also http://sci.esa.int/darwin.
8.   D. Leisawitz, J. C. Mather, S. Harvey Moseley, and X. Zhang, "The Submillimeter Probe of the Evolution of Cosmic Structure (SPECS)," Astrophys. Space Sci. **269**, 563–567 (1999).
9.   M. Shao, W. Danchi, M. J. DiPirro, M. Dragovan, L. D. Feinberg, M. Hagopian, W. D. Langer, C. R. Lawrence, P. R. Lawson, D. T. Leisawitz, J. C. Mather, S. H. Moseley, M. R. Swain, H. W. Yorke, and X. Zhang, "Space-based interfero-

metric telescopes for the far infrared," in *Interferometry in Optical Astronomy*, P. J. Lena and A. Quirrenbach, eds., Proc. SPIE **4006**, 772–781 (2000).

10. D. T. Leisawitz, W. C. Danchi, M. J. DiPirro, L. D. Feinberg, D. Y. Gezari, M. Hagopian, W. D. Langer, J. C. Mather, S. H. Moseley, M. Shao, R. F. Silverberg, J. Staguhn, M. R. Swain, H. W. Yorke, and X. Zhang, "Scientific motivation and technology requirements for the SPIRIT and SPECS far-infrared/submillimeter space interferometers," in *UV, Optical, and IR Space Telescopes and Instruments*, J. B. Breckinridge and P. Jakobsen, eds., Proc. SPIE **4013**, 36–46 (2000).

11. A. Blain, R. Ivison, and I. Smail, "Observational limits to source confusion in the millimetre/submillimetre waveband," Mon. Not. R. Astron. Soc. **296**, L29–L33 (1998).

12. I. Smail, R. J. Ivison, and A. W. Blain, "A deep submillimeter survey of lensing clusters: a new window on galaxy formation and evolution," Astrophys. J. **490**, L5–L8 (1997).

13. D. Hughes, S. Serjeant, J. Dunlop, M. Rowan-Robinson, A. Blain, R. G. Mann, R. Ivison, J. Peacock, A. Efstathiou, W. Gear, S. Oliver, A. Lawrence, M. Longair, P. Goldschmidt, and T. Jenness, "High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey," Nature **394**, 241–247 (1998).

14. A. Barger, L. Cowie, D. Sanders, E. Fulton, Y. Taniguchi, Y. Sato, K. Kawara, and H. Okuda, "Submillimetre-wavelength detection of dusty star-forming galaxies at high redshift," Nature **394**, 248–251 (1998).

15. F. Roddier, "Pupil plane versus image plane in Michelson stellar interferometry," J. Opt. Soc. Am. A **3**, 2160–2166 (1986).

16. W. Traub, "Combining beams from separated telescopes," Appl. Opt. **25**, 528–532 (1986).

17. S. Prasad and S. R. Kulkarni, "Noise in optical synthesis images. I. Ideal Michelson interferometer," J. Opt. Soc. Am. A **6**, 1702–1714 (1989).

18. M. Faucherre, B. Delabre, P. Dierickx, and F. Merkle, "Michelson versus Fizeau type beam combination—Is there a difference?" in *Amplitude and Intensity Spatial Interferometry*, Proc. SPIE **1237**, 206–217 (1990).

19. S. R. Kulkarni, S. Prasad, and T. Nakajima, "Noise in optical synthesis images. II. Sensitivity of an $^nC_2$ interferometer with bispectrum imaging," J. Opt. Soc. Am. A **8**, 499–510 (1991).

20. S. Prasad, "Sensitivity limits on an image-plane fiber interferometer at low-light levels," in *Amplitude and Intensity Spatial Interferometry II*, J. B. Breckinridge, ed., Proc. SPIE **2200**, 51–59 (1994).

21. L. Mugnier, G. Rousset, and F. Cassaing, "Aperture configuration optimality criterion for phased arrays of optical telescopes," J. Opt. Soc. Am. A **13**, 2367–2374 (1996).

22. F. Roddier and S. T. Ridgway, "Filling factor and signal-to-noise ratios in optical interferometric arrays," Publ. Astron. Soc. Pac. **111**, 990–996 (1999).

23. T. Nakajima, "Sensitivity of a ground-based infrared interferometer for aperture synthesis imaging," Publ. Astron. Soc. Pac. **113**, 1289–1299 (2001).

24. T. Nakajima and H. Matsuhara, "Sensitivity of an imaging space infrared interferometer," Appl. Opt. **40**, 514–526 (2001).

25. The author is preparing a manuscript to be called "Thermal noise and correlations in photon detection."

26. S. W. Wedge, "Computer-aided design of noise microwave circuits," Ph.D. thesis (California Institute of Technology, Pasadena, Calif., 1991).

27. S. W. Wedge and D. B. Rutledge, "Wave computations for microwave education," IEEE Trans. Educ. **36**, 127–131 (1993).

28. H. Bosma, "On the theory of linear noisy systems," Philips Res. Rep. Suppl. **10**, 1–190 (1967).

29. S. W. Wedge and D. B. Rutledge, "Noise waves and passive linear multiports," IEEE Microwave Guid. Wave Lett. **1**, 117–119 (1991).

30. S. W. Wedge and D. B. Rutledge, "Wave techniques for noise

31. modeling and measurement," IEEE Trans. Microwave Theory Tech. **40**, 2004–2012 (1992).

31. in *Principles of Microwave Circuits*, Vol. 8 of M.I.T. Radiation Laboratory Series, C. G. Montgomery, R. H. Dicke, and E. M. Purcell, eds. (McGraw-Hill, New York, 1948), Chap. 9, pp. 317–333.

32. W. K. Kahn and W. Wasylkiwskyj, "Coupling, radiation, and scattering by antennas," in *Proceedings of the Symposium on Generalized Networks*, Vol. 16 of Microwave Research Institute Symposia Series (Polytechnic, New York, 1966), pp. 83–114.

33. A. C. Gately and D. J. R. Stock, and B. R.-S. Cheo, "A network description for antenna problems," Proc. IEEE **56**, 1181–1193 (1968).

34. S. Withington and J. Murphy, "Modal analysis of partially coherent submillimeter-wave quasi-optical systems," IEEE Trans. Antennas Propag. **46**, 1651–1659 (1998).

35. P. Kern and F. Malbet, eds., *Integrated Optics for Astronomical Interferometry* (Bastinelli-Guirimand, Grenoble, France, 1996).

36. J. Berger, P. Haguenauer, P. Kern, K. Perraut, F. Malbet, I. Schanen, M. Severi, R. Millan-Gabet, and W. Traub, "Integrated optics for astronomical interferometry. IV. First measurements of stars," Astron. Astrophys. **376**, L31–L34 (2001).

37. L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," IEEE Trans. Med. Imaging **1**, 113–122 (1982).

38. R. Molina, J. Núñez, F. J. Cortijo, and J. Mateos, "Image restoration in astronomy—a Bayesian perspective," IEEE Signal Process. Mag. **18**, 11–29 (2001).

39. A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time-delay estimation—Part I: narrow-band systems," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-32**, 472–486 (1983).

40. E. Weinstein and A. J. Weiss, "Fundamental limitations in passive time-delay estimation—Part II: wide-band systems," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-32**, 1064–1078 (1984).

41. K. L. Bell, Y. Ephraim, and H. L. Van Trees, "Explicit Ziv–Zakai lower bound for bearing estimation," IEEE Trans. Signal Process. **44**, 2810–2824 (1996).

42. J. G. Cohen, "Tests of the photometric accuracy of image restoration using the maximum entropy algorithm," Astron. J. **101**, 734–737 (1991).

43. P. Jakobsen, P. Greenfield, and R. Jedrzejcwski, "The Cramer–Rao lower bound and stellar photometry with aberrated HST images," Astron. Astrophys. **253**, 329–332 (1992).

44. C. H. Hua, N. H. Clinthorne, S. J. Wilderman, J. W. LeBlanc, and W. L. Rogers, "Quantitative evaluation of information loss for Compton cameras," IEEE Trans. Nucl. Sci. **46**, 587–593 (1999).

45. H. Cramér, *Mathematical Methods of Statistics* (Princeton U. Press, Princeton, N. J., 1946).

46. C. R. Rao, "Information and accuracy attainable in the estimation of statistical parameters," Bull. Calcutta Math. Soc. **37**, 81–91 (1945).

47. H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I* (Wiley, New York, 1968).

48. P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," IEEE Trans. Signal Process. **49**, 87–90 (2001).

49. E. Serabyn, "Nulling interferometry and planet detection," in *Principles of Long Baseline Stellar Interferometry*, Course Notes from the 1999 Michelson Summer School, August 15–19, 1999, P. R. Lawson, ed., NASA/JPL Publication 00-009 Rev. 1 03-01 (NASA/Jet Propulsion Laboratory, Pasadena, Calif. 2000), pp. 275–292.

50. E. Serabyn and M. M. Colavita, "Fully symmetric nulling beam combiners," Appl. Opt. **40**, 1668–1671 (2001).

51. J. Butler and R. Lowe, "Beam-forming matrix simplifies design of electronically scanned antennas," Electron. Des. **9**, 170–173 (1961).