

Cramér–Rao Sensitivity Limits for Astronomical Interferometry

J. Zmuidzinas

Division of Physics, Mathematics, and Astronomy

California Institute of Technology, 320–47

*Pasadena, CA 91125**

Abstract

Multiple telescope interferometry for high angular resolution astronomical imaging in the optical/IR/far-IR bands is currently a topic of great scientific interest. This paper reviews the fundamentals which govern the sensitivity of direct-detection interferometers, and discusses the sensitivity limits imposed by the Cramér–Rao theorem. This theorem is used to support the argument that interferometers which have more compact instantaneous beam patterns are more sensitive, since they extract more spatial information from each detected photon. This favors arrays with a larger number of telescopes, and it favors all-on-one beam combining methods as compared to pairwise combination.

©2002 Optical Society of America

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

*Electronic address: jonas@submm.caltech.edu; URL: <http://www.submm.caltech.edu>

1. Introduction

Astronomical spatial interferometry, which is the technique of interfering the radiation gathered by several separated telescopes, is of great current interest because of the exciting scientific potential of very high angular resolution observations, combined with the numerous technological advances that now make optical interferometry feasible. 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 have been given recently^{1,2}, which describe the scientific motivation and results, technical challenges and approaches, existing and planned facilities, and contain extensive bibliographies. Additional information is readily available on the internet; see <http://olbin.jpl.nasa.gov>.

Of course, interferometry is very well developed at radio wavelengths, where large arrays of telescopes such as the NRAO VLA (Very Large Array; <http://www.nrao.edu>) routinely provide high angular resolution synthetic aperture imaging. Ground-based interferometry at optical wavelengths is inherently more difficult due to the limitations imposed by atmospheric phase fluctuations; dealing with these fluctuations is perhaps the key issue for ground-based systems. This provides strong motivation to consider interferometry in space. As a result, NASA is pursuing the SIM mission³ (<http://sim.jpl.nasa.gov>), an optical astrometric interferometer with a 10m baseline, scheduled for launch at the end of the decade. SIM 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/submm spectrum. Looking further into the future, infrared interferometers are being considered for missions such as TPF⁴ (NASA; <http://planetquest.jpl.nasa.gov>) and Darwin⁵, (ESA; <http://sci.esa.int/darwin>) which have the ambitious goal of detecting and characterizing Earth-like planets around nearby stars. Direct-detection space interferometers at very long (far-IR/submm) wavelengths using cold telescopes have also been proposed, such as the NASA SPECS/SPIRIT concepts⁶⁻⁹. Such interferometers would give the angular resolution needed to break through the spatial confusion limit¹⁰, 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 redshifts¹¹⁻¹³.

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^{14–24}, there appears to be no general consensus on which design approaches give the best sensitivity, or even if there is much difference between them¹. 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 only discuss interferometers with nearly ideal characteristics.

Our approach will focus on the instantaneous (not synthesized) angular response function (or interferometric beam pattern) associated with a given detector (or collection of detectors) which receive 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 which have high occupation numbers for the photon modes. The connections and distinctions between the optical and radio cases will be discussed in more detail in a future paper²⁵.

The structure of the paper is as follows. We will first review the description of electromagnetic (including optical) systems in terms of scattering matrices. This formalism will be used extensively throughout the rest of the paper. Next, we will review the coherence properties of the radiation emitted by astronomical sources, and use this to calculate the

response of general optical systems to astronomical sources. This response can be given a very straightforward and natural interpretation, in which incoming photons have various probabilities to be absorbed in the different detectors in the instrument, depending on their frequency and the angular position from which they were emitted from the source. Using this simple model, we show that only instruments which can achieve ideal sensitivity are those which 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. In order to deal with this non-ideal case, we apply the Cramér–Rao limit to calculate the limiting sensitivities of non-ideal instruments. The remaining step is to calculate the actual response of interferometers, and their Cramér–Rao sensitivities, which we do for the simple but illustrative case of one-dimensional arrays. The paper closes with comments on these results and indicates areas for future work.

2. Scattering Matrix Description of Optical Systems

The response of any electromagnetic system, such as a collection of optical elements (mirrors, lenses, etc.) but excluding detectors, may be described by a classical scattering matrix S . Scattering matrices are common in electrical engineering^{26,27}, where they are used to describe linear N -port circuits:

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

The indices $1 \leq i, j \leq N$ label the ports. A set of N transmission lines attached to the ports carry 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, $P_{\text{inc}} = \sum_i |a_i|^2$. With this choice of normalization, it is straightforward to show that a lossless circuit has a unitary scattering matrix, $SS^\dagger = \mathbb{1}$. Reciprocity, which has its roots in time reversal symmetry and applies for most passive circuits, implies that $S^T = S$. By expressing the wave amplitudes 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 straightforward²⁵.

We can apply the scattering matrix concept to characterize an optical system, such as the collection of telescopes and beam-combining optics that comprise an interferometer.

This approach was developed for antenna problems during the WW2 radar effort³¹, and is particularly convenient at radio wavelengths since it allows circuit and antenna concepts to be treated in a unified fashion. While scattering matrices are now often applied to antenna problems^{32,33}, they are not used very often to describe optical systems (though they do find occasional use³⁴). 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. In order 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 $e^{+j\omega t}$ time dependence, the electric field of the incoming wave arriving at the telescope system (or antenna) can be expressed as

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

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

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

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

$$P_{\text{inc}} = \int d\Omega |\mathbf{a}(\Omega)|^2 \quad (4)$$

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, which 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, while “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 by 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 i th 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 which 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

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

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, etc.) 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:

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

The first block, $\mathbf{S}^{(\text{scat})}(\Omega, \Omega')$, describes the scattering of an incoming plane wave arriving from Ω' to an outgoing plane wave traveling toward Ω . The sans serif font reminds us that this is a 3×3 matrix to account for polarization. The second and third blocks are off-diagonal: $\mathbf{S}_{i\alpha}^{(\text{rec})}(\Omega')$ describes the (vector) receiving antenna pattern for the detector pixel mode $i\alpha$, while $\mathbf{S}_{j\beta}^{(\text{trans})}(\Omega)$, describes the transmitting antenna pattern. The fourth block is an ordinary matrix, $S_{i\alpha j\beta}^{(\text{refl})}$, which represents the scattering (reflection) of the telescope 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:

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

and

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

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

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

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

$$\mathbf{S}^{(\text{scat})}(\Omega, \Omega') = (\mathbf{S}^{(\text{scat})})^T(\Omega', \Omega) , \quad (11)$$

and that the pixel to pixel scattering matrix is reciprocal, $S^{(\text{refl})} = (S^{(\text{refl})})^T$.

The output power emanating from a passive optical system cannot exceed the input power, which imposes an important constraint on the scattering operator: $\mathbb{1} - S^\dagger S$ must be nonnegative definite. Combined with the reciprocity theorem, this can be used to demonstrate that the following matrix must also be nonnegative definite:

$$M_{i\alpha, j\beta} = \delta_{i\alpha, j\beta} - \int d\Omega (\mathbf{S}_{i\alpha}^{(\text{rec})}(\Omega))^* \cdot \mathbf{S}_{j\beta}^{(\text{rec})}(\Omega) - \sum_{k\gamma} (S_{k\gamma, i\delta}^{(\text{refl})})^* S_{k\gamma, j\beta}^{(\text{refl})} \quad (12)$$

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

$$\int d\Omega |\mathbf{S}_{i\alpha}^{(\text{rec})}(\Omega)|^2 \leq 1 - \sum_{j\beta} |S_{j\beta i\alpha}^{(\text{refl})}|^2 \leq 1 \quad (13)$$

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

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

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 themselves radiate into the optical system, so that $a_{i\alpha} = 0$. (The detectors are usually operated at a low enough temperature that the thermal radiation they emit is negligible.) The power absorbed by pixel α is

$$P_\alpha = \sum_i |b_{i\alpha}|^2 = \sum_i \left| \int d\Omega \mathbf{S}_{i\alpha}^{(\text{rec})}(\Omega) \cdot \mathbf{a}(\Omega) \right|^2 \quad (15)$$

3. Astronomical sources

Astronomical sources emit radiation which 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

$$\langle a_q(\Omega, \nu) a_{q'}^*(\Omega', \nu') \rangle = A_{qq'}(\Omega, \nu) \delta(\Omega - \Omega') \delta(\nu - \nu') \quad (16)$$

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 would not be diagonal in the mode indices, in general.

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 $\hat{\mathbf{s}}$:

$$F(\hat{\mathbf{p}}, \hat{\mathbf{s}}, \nu) = \frac{1}{\lambda^2} \sum_{qq'} \int d\Omega \hat{\mathbf{n}}(\Omega) \cdot \hat{\mathbf{s}} \left[\frac{\hat{\mathbf{p}}^* \cdot \hat{\mathbf{e}}_q(\Omega) A_{qq'}(\Omega, \nu) \hat{\mathbf{e}}_{q'}^*(\Omega) \cdot \hat{\mathbf{p}}}{|1 - \hat{\mathbf{n}}(\Omega) \cdot \hat{\mathbf{p}}|^2} \right] \quad (17)$$

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

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

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

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

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 using Eqs. (15) and (16):

$$\langle P_\alpha(\nu) \rangle = \sum_{qq'} \int d\Omega A_{qq'}(\Omega, \nu) \left[\sum_i (\hat{\mathbf{e}}_q \cdot \mathbf{S}_{i\alpha}^{(\text{rec})}(\Omega, \nu)) (\mathbf{S}_{i\alpha}^{(\text{rec})}(\Omega, \nu) \cdot \hat{\mathbf{e}}_{q'})^* \right] \quad (20)$$

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

$$\langle P_\alpha(\nu) \rangle = \int d\Omega A(\Omega, \nu) R_\alpha(\Omega, \nu) \quad (21)$$

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

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

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

$$m_\alpha(\nu) = \int d\Omega R_\alpha(\Omega, \nu) \quad (23)$$

According to Eq. (13), $m_\alpha(\nu) \leq \sum_i 1$, 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

$$\langle P_\alpha(\nu) \rangle = \frac{1}{2} F(\nu) \lambda^2 R_\alpha(\Omega^p, \nu) \quad (24)$$

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 $F(\nu) A_{\text{tel}}$. Thus,

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

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

$$m_\alpha \leq \frac{2A_{\text{tel}} \Delta\Omega_\alpha}{\lambda^2} \quad (26)$$

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.

Eq.(25) suggests that we renormalize our response functions:

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

so that they obey

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

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

$$\langle P_\alpha(\nu) \rangle = \frac{2A_{\text{tel}}}{\lambda^2} \int d\Omega A(\Omega, \nu) \rho_\alpha(\Omega, \nu) \quad (29)$$

We can discretize this integral by splitting up 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:

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

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

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

where the average of the response function over patch s is

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

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 :

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

where

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

is just *the maximum average number of photons (per unit bandwidth) which 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 Eq. (28).*

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

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

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

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

and the response function is now also averaged over frequency:

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

5. Photon counting statistics and ideal instruments

It is well established that the photons 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, which changes the counting statistics and can introduce correlations among the detectors. For now, we will discuss only the first case, $n \ll 1$, which applies to most astronomical observations at optical and infrared wavelengths.

The previous section (Eqs. 35, 36) reminds us that in the $n \ll 1$ limit, the description of *any photon direct-detection instrument*, regardless of spectral or spatial resolution, single telescope vs. 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 : $p_{\alpha c} = \bar{\rho}_{\alpha}(\Omega_s, \nu_f)$. 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 $\lambda_c = \langle N(\Omega_s, \nu_f) \rangle$ represent the mean number of “incident” photons from channel c arriving at the instrument. The mean number $\mu_{\alpha} = \langle N_{\alpha} \rangle$ of photons detected by detector α is given by

$$\mu_{\alpha} = \sum_c p_{\alpha c} \lambda_c \quad (38)$$

From Eq. (28), we have

$$\sum_{\alpha} p_{\alpha c} \leq 1 \quad (39)$$

so that the mean number of photons detected cannot exceed the mean number of incident photons, $\sum_{\alpha} \mu_{\alpha} \leq \sum_c \lambda_c$.

In most cases, we are interested in determining the quantities λ_c , which give the spatial and spectral distribution of the radiation emitted by the source. How well can this be done? We can imagine an “ideal” instrument in which photons arriving from channel c would be detected only by a single detector, α_c , with a one to one mapping between detectors and channels. In other words, this instrument does not “mix up” photons from different channels, and a photon detection event can be unambiguously assigned to the appropriate channel. The corresponding probability matrix $p_{\alpha c}$ is the identity matrix (or a permutation).

The spatial and spectral resolution needed are set by our choice of bin size. Does such an instrument exist, at least in principle? Achieving the required spectral resolution is not a fundamental difficulty; all one needs is a large enough grating. Achieving the required spatial resolution is a more subtle issue, since the parameters λ_c refer to a fixed telescope collecting area. 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) in order to keep the effective collecting area (and the parameters λ_c) constant. Of course, doing so would be foolish; this argument only serves as an existence proof.

For our “ideal” instrument, the number of photons N_c counted by the detectors are independent Poisson random variables:

$$\langle N_c \rangle = \lambda_c \quad (40)$$

$$\langle \delta N_c \delta N_{c'} \rangle = \lambda_c \delta_{cc'} \quad (41)$$

where $\delta N_c = N_c - \langle N_c \rangle$. Are there other instruments, which do not have a one to one mapping between detectors and channels, that can achieve the same sensitivity? *The answer is no, which we will demonstrate with a short proof.*

We represent the photon counts in the detectors by the vector \tilde{N} . Some sort of inversion procedure is necessary to convert the measured detector counts \tilde{N} into an estimate $\hat{\lambda}_c(\tilde{N})$ of the intensity of the source in the various channels c . This inversion procedure may or may not be linear. However, we assume that it is differentiable, so for small perturbations

$$\delta \hat{\lambda}_c = \sum_{\alpha} A_{c\alpha} \delta N_{\alpha} \quad (42)$$

where $A_{c\alpha} = \partial \hat{\lambda}_c / \partial N_{\alpha}$. We will restrict our attention to inversion procedures which faithfully extract small changes in the source distribution. To be precise, we will require

$$\delta \lambda_c = \langle \delta \hat{\lambda}_c \rangle = \sum_{\alpha} A_{c\alpha} \delta \mu_{\alpha} = \sum_{\alpha} A_{c\alpha} \sum_{c'} p_{\alpha c'} \delta \lambda_{c'} \quad (43)$$

for any type of small perturbation $\delta \lambda_c$. This means that the matrix A should be the inverse of p :

$$\sum_{\alpha} A_{c\alpha} p_{\alpha c'} = \delta_{cc'} \quad (44)$$

This equation shows that A must be independent of \tilde{N} , and so in fact only linear inversion procedures can meet the requirements that we have set. We dismiss the case that p has no

inverse, since such instruments are clearly inferior to our ideal instrument. Note that this means that there must be at least as many detectors as the number of spatial/spectral channels we wish to resolve (with proper interpretation, this statement also applies to aperture synthesis imaging; see section 8). We can now calculate the noise for some channel in our estimated source distribution:

$$\sigma_c^2 = \langle (\delta \hat{\lambda}_c)^2 \rangle = \sum_{\alpha c'} (A_{c\alpha})^2 p_{\alpha c'} \lambda_{c'} \quad (45)$$

where we have used the fact that the photon counts are independent Poisson random variables, with $\langle \delta N_\alpha \delta N_{\alpha'} \rangle = \mu_\alpha \delta_{\alpha\alpha'}$. Since we do not expect that any instrument will be more sensitive than the “ideal” instrument, it should be possible to show that $\sigma_c^2 \geq \lambda_c$, or

$$\sum_{\alpha} (A_{c\alpha})^2 p_{\alpha c} \geq 1 \quad (46)$$

It is not difficult to demonstrate that this must in fact be true. Since $(A_{c\alpha} - 1)^2$ and $p_{\alpha c}$ are both nonnegative,

$$0 \leq T_c = \sum_{\alpha} (A_{c\alpha} - 1)^2 p_{\alpha c} \quad (47)$$

$$\begin{aligned} &= \sum_{\alpha} (A_{c\alpha})^2 p_{\alpha c} - 2 \sum_{\alpha} A_{c\alpha} p_{\alpha c} + \sum_{\alpha} p_{\alpha c} \\ &\leq \sum_{\alpha} (A_{c\alpha})^2 p_{\alpha c} - 2 + 1 \end{aligned} \quad (48)$$

which yields the desired inequality (46). Thus, the sensitivity of our ideal instrument cannot be exceeded – it does indeed live up to its name.

In fact, the sensitivity of an ideal instrument, $\sigma_c^2 = \lambda_c$, can only be obtained when the probability matrix $p_{c\alpha}$ is simply some permutation of the identity matrix. In other words, the instrument must be equivalent to the ideal instrument, perhaps with the detector labels shuffled around by some permutation; it cannot allow photons from any two different spatial/spectral channels to be detected by the same detector. To prove this, let us define $D_c = \{\alpha | p_{\alpha c} \neq 0\}$ to be the set of detectors which can receive photons from channel c . According to relations (45) and (46), an ideal instrument must obey

$$\sum_{\alpha} (A_{c\alpha})^2 p_{\alpha c'} = \delta_{cc'} \quad (49)$$

The $c = c'$ equation requires that $T_c = 0$, which according to (47, 48) means that $\sum_{\alpha} p_{\alpha c} = 1$ (no photons are lost) and that $A_{c\alpha} = 1$ for any $\alpha \in D_c$. On the other hand, the $c \neq c'$

equations then imply that $p_{\alpha c'} = 0$ for any $\alpha \in D_c$, since the terms in the sum (49) must vanish. In other words, if some detector α receives photons from some channel c , then it cannot receive photons from any other channel c' . This completes the proof.

6. The Cramér–Rao sensitivity limit

In the previous section, 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 which 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, in order to be able to compare various options in the design of an interferometer, such as the number of telescopes, their positions, the method of beam combining, etc. The Cramér–Rao theorem provides a strict lower limit for the variance of a quantity which is estimated from a set of noisy measurements, and can be applied to determine the minimum noise in the determination of the intensity in some spatial/spectral channel using an instrument with a nonideal response. 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 algorithm³⁷. In fact, the Cramér–Rao limit was applied to exactly this situation by Jakobsen *et al.*³⁸. 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 medicine³⁹. Finally, we point out that there is a fairly well developed literature which discusses the general role of information theory in image formation⁴⁰ in a more sophisticated

and formal manner; in addition, other types of lower bounds exist for errors in parameter estimation problems⁴¹.

We start by quickly reviewing the Cramér–Rao theorem; the papers cited above, and the references therein, should be consulted for more details. Let us consider an experiment which delivers a set of measurements, denoted by the vector \tilde{x} , but which has scatter due to measurement noise. The measurement process can be described by a probability distribution $f(\tilde{x}|\tilde{\theta})$ to obtain a result \tilde{x} . Here the vector $\tilde{\theta}$ 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 . In order to do this, we must construct some estimator $\hat{\theta}_i(\tilde{x})$, which uses the measured data to estimate θ_i . For simplicity, we assume that this estimator is unbiased; the results could be generalized to include bias. The Cramér–Rao theorem states that

$$\begin{aligned}\sigma_i^2 &= \langle (\hat{\theta}_i - \theta_i)^2 \rangle = \int d\tilde{x} f(\tilde{x}|\tilde{\theta}) (\hat{\theta}_i(\tilde{x}) - \theta_i)^2 \\ &\geq \left\langle \left(\frac{\partial \ln f}{\partial \theta_i} \right)^2 \right\rangle^{-1} \\ &= \left[\int d\tilde{x} f(\tilde{x}|\tilde{\theta}) \left(\frac{\partial \ln f(\tilde{x}|\tilde{\theta})}{\partial \theta_i} \right)^2 \right]^{-1}\end{aligned}$$

A somewhat more stringent limit can also be given. First, define the matrix

$$M_{ij} = \left\langle \frac{\partial \ln f}{\partial \theta_i} \frac{\partial \ln f}{\partial \theta_j} \right\rangle \quad (50)$$

This is known as the Fisher information matrix, and is symmetric and non-negative definite. In fact, it is positive-definite, unless there is some linear combination of parameters θ_i that the function $f(\tilde{x}|\tilde{\theta})$ 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. 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

$$\sigma_i^2 \geq [M^{-1}]_{ii} \quad (51)$$

It is instructive to apply this result to the familiar case of an N -dimensional multivariate Gaussian distribution of the form

$$f(\tilde{x}|\tilde{\theta}) = (2\pi)^{-N/2} (\det A)^{1/2} \quad (52)$$

$$\times \exp \left[-\frac{1}{2} \sum_{ij} (x_i - \theta_i) A_{ij} (x_j - \theta_j) \right]$$

for which one knows that the maximum likelihood estimators $\hat{\theta}_i = x_i$ are efficient, and $\langle (x_i - \theta_i)(x_j - \theta_j) \rangle = (A^{-1})_{ij}$. In fact, for this case one finds $M_{ij} = A_{ij}$, so the Cramér–Rao bound actually gives the true variance, $\sigma_i^2 = (A^{-1})_{ii}$.

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

$$f(\tilde{N}|\tilde{\lambda}) = \prod_{\alpha} \frac{\mu_{\alpha}^{N_{\alpha}}}{N_{\alpha}!} \exp(-\mu_{\alpha}) \quad (53)$$

so

$$\frac{\partial \ln f}{\partial \lambda_c} = \sum_{\alpha} \left(-p_{\alpha c} + \frac{N_{\alpha} p_{\alpha c}}{\mu_{\alpha}} \right) \quad (54)$$

and

$$\begin{aligned} M_{cc'} &= \left\langle \frac{\partial \ln f}{\partial \lambda_c} \frac{\partial \ln f}{\partial \lambda_{c'}} \right\rangle \\ &= \sum_{\alpha} \frac{p_{\alpha c} p_{\alpha c'}}{\mu_{\alpha}} = \sum_{\alpha} \frac{p_{\alpha c} p_{\alpha c'}}{\sum_{c''} p_{\alpha c''} \lambda_{c''}} \end{aligned} \quad (55)$$

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

$$\sigma_c^2 \geq (M^{-1})_{cc} \quad (56)$$

Eqs. (55) and (56) are key results, since they give us a quantitative way to set lower limits to the sensitivity of optical instruments. Alternatively, for simplicity we may wish to use the first (weaker) form of the Cramér–Rao bound, $\sigma_c^2 \geq (M_{cc})^{-1}$, which is the result quoted by Jakobsen *et al.*³⁸:

$$\sigma_c^2 \geq \left[\sum_{\alpha} \frac{p_{\alpha c}^2}{\sum_{c'} p_{\alpha c'} \lambda_{c'}} \right]^{-1} \quad (57)$$

It is easy to verify that either form gives $\sigma_c^2 = \lambda_c$ in the case of an ideal instrument, for which $p_{\alpha c} = \delta_{\alpha c}$. 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 , Eq. (57) gives $\sigma_c^2 = \lambda_c (\sum_{\alpha} p_{\alpha c})^{-1} \geq \lambda_c$; the equality holds if all of the photons are detected. The sensitivity limit given by the stronger bound (Eq. 56) 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 up all the photons counted by all the detectors; it doesn't matter much how the photons are distributed among the detectors. Thus, calculations which only compare the sensitivities of interferometers to point sources do not tell the whole story.

7. Single-mode interferometers

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 $t = 1 \dots T$) 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

$$\mathbf{S}_t^{(\text{rec})}(\Omega, \nu) = \exp[+ik\hat{\mathbf{n}}(\Omega) \cdot \mathbf{r}_t] \mathbf{S}^{(\text{rec})}(\Omega, \nu) \quad (58)$$

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 :

$$b_t(\nu) = \int d\Omega \mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \cdot \mathbf{a}(\Omega, \nu) \quad (59)$$

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

$$\begin{aligned} & \int d\Omega \left[\mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \right]^* \cdot \mathbf{S}_{t'}^{(\text{rec})}(\Omega, \nu) \\ &= \int d\Omega |\mathbf{S}^{(\text{rec})}(\Omega, \nu)|^2 \exp[-ik\hat{\mathbf{n}}(\Omega) \cdot (\mathbf{r}_t - \mathbf{r}_{t'})] \\ &\approx 0 \end{aligned} \quad (60)$$

due to 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:

$$b_{i\alpha}(\nu) = \sum_t S_{i\alpha,t}^{(\text{comb})} b_t(\nu) \quad (61)$$

By combining Eqs. (59) and (61), we have

$$b_{i\alpha} = \sum_t S_{i\alpha,t}^{(\text{comb})}(\nu) \mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \cdot \mathbf{a}(\Omega, \nu) \quad (62)$$

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

$$R_\alpha(\Omega, \nu) = \sum_i \left| \sum_t S_{i\alpha,t}^{(\text{comb})}(\nu) \mathbf{S}_t^{(\text{rec})}(\Omega, \nu) \right|^2 \quad (63)$$

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

$$\rho_\alpha(\Omega, \nu) = \frac{\lambda^2}{A_T} R_\alpha(\Omega, \nu) \quad (64)$$

and we would also omit the corresponding factor of 2 in Eq. (36). Thus, the set of equations (36, 55, 56, 58, 63, 64) can be used to provide a sensitivity bound for any interferometer configuration and beam combining method. In fact, we will interpret this bound as a good estimate of the actual sensitivity, which is reasonable in the limit that many photons are detected, since the counting statistics become nearly Gaussian and maximum likelihood estimators are asymptotically efficient.

Although we have assumed identical telescopes, it is straightforward to generalize our expressions to include heterogeneous arrays. There is one small remaining issue, which has to do with the proper treatment of the various array configurations (including the effects of Earth rotation) that are used for aperture synthesis imaging. This will be tackled in the next section, where we describe the calculated results for simple uniform one-dimensional arrays.

8. Application to One-Dimensional Arrays

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. 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

$$S^{(rec)}(\theta) = \sqrt{\frac{L_1}{\lambda}} \operatorname{sinc}\left(\frac{\pi L_1}{\lambda}\theta\right) \quad (65)$$

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 FWHP 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,

$$S_t^{(rec)}(\theta) = \exp\left[i\frac{2\pi x_t}{\lambda}\theta\right] S^{(rec)}(\theta) \quad (66)$$

The response of the interferometer for a given choice of beam combination is still given by Eq. (63); however, the proper normalization for the detection probability function is

$$\rho_\alpha(\theta, \nu) = \frac{\lambda}{L_T} R_\alpha(\theta, \nu) \quad (67)$$

We compare two types of beam combination. The first case is the usual pairwise combination, in which we are focusing on measuring the $T(T-1)/2$ fringe visibilities and fringe phases that can be obtained from the various telescope “baseline” pairs. In order 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:

$$\begin{aligned} b_1(t, t') &= \frac{1}{2\sqrt{T-1}} (b_t + b_{t'}) \\ b_2(t, t') &= \frac{1}{2\sqrt{T-1}} (b_t - b_{t'}) \\ b_3(t, t') &= \frac{1}{2\sqrt{T-1}} (b_t + ib_{t'}) \\ b_4(t, t') &= \frac{1}{2\sqrt{T-1}} (ib_t + b_{t'}) \end{aligned}$$

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 $b_i(t, t')$ represent the various combinations of light from the two telescopes that are being detected to produce the corresponding photon counts $N_i(t, t')$. The total number of detectors in this scheme is $2T(T - 1)$. It is straightforward to verify that the first two beam combinations produce symmetric angular response functions, while the latter two produce antisymmetric response functions (apart from a constant offset term). Thus, both types of beam combination are needed in order to uniquely determine the image of a source. The latter two beam combinations are readily produced using a 50% beamsplitter; at microwave frequencies, the equivalent device is known as a “90° 3 dB hybrid”. The first two combinations can be obtained using the optical equivalent of a microwave “180° 3 dB hybrid”; such devices are currently being investigated for nulling interferometry^{42,43}. All four combinations can be gotten simultaneously by using a two-way power splitter (or a 50% beamsplitter), as shown in Fig. 1. It is straightforward to verify that for this beam combination scheme, all of 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.

As shown in Fig. 2, the second type of beam combination method we investigate is the standard “Butler matrix” beamforming network⁴⁴, 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 beamforming, Butler beamformers 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:

$$b_\alpha = \frac{1}{\sqrt{T}} \sum_{t=1}^T b_t \exp \left[i \frac{2\pi\alpha t}{T} \right] \quad (68)$$

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.

We consider array configurations of 3, 6, and 10 uniformly-spaced telescopes, as shown in Table 1. The source brightness distribution in θ is discretized into N_{bins} bins or “pixels”, extending across the field of view of an individual telescope. Since the maximum baseline of all of the array configurations shown in Table 1 is $B_{\text{max}} = 45L_1$, we use $N_{\text{bins}} = 41$ spatial pixels in the calculations. The pixel-averaged response functions, or detection probabilities, are calculated using a one-dimensional analog of Eq. (32). As shown in Table 1, the calculation also includes N_{spacings} different telescope spacings, ranging from close-packed to dilute arrays. In order to account for these various configurations, we imagine that there are actually $N_{\text{spacings}}N_{\text{detectors}}$ different “virtual” detectors, where $N_{\text{detectors}}$ is the physical number of detectors in any one configuration. (Note that we must have $N_{\text{bins}} < N_{\text{spacings}}N_{\text{detectors}}$ in order to reconstruct the discretized image of the source; otherwise, the Fisher information matrix (Eq. 55) will be singular.) The detection probabilities for any one configuration are reduced by the factor $(N_{\text{spacings}})^{-1}$. In essence, we are splitting up the total observing time into N_{spacings} 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 $(N_{\text{spacings}})^{-1}$. Of course, this ignores the time overhead that array reconfiguration would require in reality. 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 α was near unity: $\sum_{\alpha=1}^{(N_{\text{spacings}}N_{\text{detectors}})} p_{\alpha c} \approx 1$.

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

Figure 4 shows the results for a uniform source with Butler beam combination. The vertical axis is the normalized Cramér–Rao sensitivity bound, calculated using Eq. (56). This sensitivity would be unity for an ideal instrument, that is a single telescope with a large enough aperture 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, due to its inability to fully determine which spatial pixel each detected photon came from. 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 penalty for the Butler–combined interferometer is a factor of ~ 3 for a 10–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 line in Figure 4. The reason for this is that the Cramér–Rao sensitivity bound includes the effects of noise “crosstalk” between the spatial pixels that arise from the nonideal interferometer beam patterns.

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 . The comparison between the Butler–combined and the pairwise–combined 10–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, 8, and 9. The normalized Cramér–Rao sensitivities for the point source itself are quite comparable in all cases, which we expect, since we can in essence sum up all of the photons detected in order 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, about an order of magnitude for the 10–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 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 telescopes systems due to scattered light.)

9. 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 arrays which 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.

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 of any configuration and which use any type of beam combination method. Our approach, which is based on scattering matrices, is especially well suited to describe beam combination using guided–wave (integrated) 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, due to thermal emission from the telescope or atmosphere, detector dark current and/or read noise, etc. Although these effects are important in real applications and should be included in more realistic calculations, we have ignored them in order to limit the parameter space and to focus on the fundamental issues involved. In closing, we encourage other workers who are involved in interferometer design to investigate the applicability of the Cramér–Rao approach for determining the sensitivity tradeoffs for real arrays, and for a range of astrophysical problems.

Acknowledgments

The author wishes to thank W. C. Danchi, S. H. Moseley, and T. G. Phillips for useful discussions.

References

- [1] A. Quirrenbach, “Optical Interferometry,” *Annual Review of Astronomy and Astrophysics* **39**, 353–401 (2001).
- [2] S. K. Saha, “Modern optical astronomy: technology and impact of interferometry,” *Rev. Mod. Phys.* 74 (2002), in press.

- [3] M. Shao, “SIM: the space interferometry mission,” In *Proc. SPIE Vol. 3350*, p. 536-540, *Astronomical Interferometry*, Robert D. Reasenberg; Ed., **3350**, 536–540 (1998).
- [4] C. A. Beichman, “Terrestrial Planet Finder: the search for life-bearing planets around other stars,” In *Proc. SPIE Vol. 3350*, p. 719-723, *Astronomical Interferometry*, Robert D. Reasenberg; Ed., **3350**, 719–723 (1998).
- [5] A. J. Penny, A. Leger, J. Mariotti, C. Schalinski, C. Eiroa, R. J. Lurance, and M. Fridlund, “DARWIN interferometer,” In *Proc. SPIE Vol. 3350*, p. 666-671, *Astronomical Interferometry*, Robert D. Reasenberg; Ed., **3350**, 666–671 (1998).
- [6] D. Leisawitz, J. C. Mather, S. Harvey Moseley, and X. Zhang, “The Submillimeter Probe of the Evolution of Cosmic Structure (SPECS),” *Astrophysics and Space Science* **269**, 563–567 (1999).
- [7] M. Shao *et al.*, “Space-based interferometric telescopes for the far infrared,” In *Proc. SPIE Vol. 4006*, p. 772-781, *Interferometry in Optical Astronomy*, Pierre J. Lena; Andreas Quirrenbach; Eds., **4006**, 772–781 (2000).
- [8] D. T. Leisawitz *et al.*, “Scientific motivation and technology requirements for the SPIRIT and SPECS far-infrared/submillimeter space interferometers,” In *Proc. SPIE Vol. 4013*, p. 36-46, *UV, Optical, and IR Space Telescopes and Instruments*, James B. Breckinridge; Peter Jakobsen; Eds., **4013**, 36–46 (2000).
- [9] J. C. Mather and D. Leisawitz, “The SPIRIT and SPECS Far-Infrared/Submillimeter Interferometry Missions,” The Institute of Space and Astronautical Science Report SP No. 14, p. 219-224. **14**, 219–224 (2000).
- [10] A. Blain, R. Ivison, and I. Smail, “Observational limits to source confusion in the millimetre/submillimetre waveband,” *MNRAS* **296**, L29–L33 (1998).
- [11] I. Smail, R. J. Ivison, and A. W. Blain, “A Deep Submillimeter Survey of Lensing Clusters: A New Window on Galaxy Formation and Evolution,” *ApJ* **490**, L5 (1997).
- [12] D. Hughes *et al.*, “High-redshift star formation in the Hubble Deep Field revealed by a submillimetre-wavelength survey,” *Nature* **394**, 241–247 (1998).
- [13] 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).
- [14] F. Roddier, “Pupil plane versus image plane in Michelson stellar interferometry,” *J. Opt. Soc.*

- Am. A **3**, 2160–2166 (1986).
- [15] W. Traub, “Combining Beams from Separated Telescopes,” *Applied Optics* **25**, 528–532 (1986).
- [16] S. Prasad and S. R. Kulkarni, “Noise in optical synthesis images. I. Ideal Michelson interferometer,” *J. Opt. Soc. Am. A* **6**, 1702–1714 (1989).
- [17] 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; Proceedings of the Meeting, Tucson, AZ, Feb. 14-16, 1990 (A91-30676 12-89)*. Bellingham, WA, SPIE, 1990, p. 206-217., **1237**, 206–217 (1990).
- [18] 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).
- [19] S. Prasad, “Sensitivity limits on an image-plane fiber interferometer at low-light levels,” In *Proc. SPIE, Amplitude and Intensity Spatial Interferometry II, James B. Breckinridge; Ed.*, **2200**, 51–59 (1994).
- [20] 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).
- [21] F. Roddier and S. T. Ridgway, “Filling Factor and Signal-to-Noise Ratios in Optical Interferometric Arrays,” *Pub. Astron. Soc. Pac.* **111**, 990–996 (1999).
- [22] M. L. Cobb, “A Comparison of Michelson and Fizeau Beam Combiners for Optical Interferometry,” American Astronomical Society Meeting (2000).
- [23] T. Nakajima, “Sensitivity of a Ground-based Infrared Interferometer for Aperture Synthesis Imaging,” *Pub. Astron. Soc. Pac.* **113**, 1289–1299 (2001).
- [24] T. Nakajima and H. Matsuhara, “Sensitivity of an imaging space infrared interferometer,” *Applied Optics* **40**, 514–526 (2001).
- [25] J. Zmuidzinas, (the author) is preparing a manuscript to be called “Thermal Noise and Correlations in Photon Detection” (unpublished).
- [26] S. W. Wedge, Ph.D. thesis, California Institute of Technology, 1991.
- [27] S. W. Wedge and D. B. Rutledge, “Wave Computations for Microwave Education,” *IEEE Trans. Education* **36**, 127–131 (1993).
- [28] H. Bosma, “On the Theory of Linear Noisy Systems,” *Phillips Res. Repts. Suppl.* **10**, 1–190 (1967).

- [29] S. W. Wedge and D. B. Rutledge, “Noise Waves and Passive Linear Multiports,” *IEEE Microwave Guided Wave Lett.* **MGL-1**, 117–119 (1991).
- [30] S. W. Wedge and D. B. Rutledge, “Wave Techniques for Noise Modeling and Measurement,” *IEEE Trans. Microwave Theory Tech.* **MTT-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., (New York: McGraw–Hill, 1948), Chap. 9, pp. 317–333.
- [32] W. K. Kahn and W. Wasyliwskyj, “Coupling, Radiation, and Scattering by Antennas,” In *Proceedings of the Symposium on Generalized Networks*, Microwave Research Institute Symposia Series **XVI**, 83–114 (New York: Polytechnic Press, 1966).
- [33] A. C. G. amd 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. Anten. Prop.* **46**, 1651–1659 (1998).
- [35] *Integrated Optics for Astronomical Interferometry*, P. Kern and F. Malbet, eds., (Grenoble, France: Bastinelli–Guirimand, 1996).
- [36] J. Berger, P. Haguenaer, 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] J. G. Cohen, “Tests of the photometric accuracy of image restoration using the maximum entropy algorithm,” *Astron. J.* **101**, 734–737 (1991).
- [38] P. Jakobsen, P. Greenfield, and R. Jedrzejewski, “The Cramér–Rao lower bound and stellar photometry with aberrated HST images,” *Astron. Astrophys* **253**, 329–332 (1992).
- [39] 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. Nuclear Science* **46**, 587–593 (1999).
- [40] J. A. O’Sullivan, R. E. Blahut, and D. L. Snyder, “Information–Theoretic Image Formation,” *IEEE Trans. Information Theory* **44**, 2094–2123 (1998).
- [41] E. Weinstein and A. J. Weiss, “A General Class of Lower Bounds in Parameter Estimation,” *IEEE Trans. Information Theory* **34**, 338–342 (1988).
- [42] 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, 2000), pp. 275–292.
- [43] E. Serabyn and M. M. Colavita, “Fully Symmetric Nulling Beam Combiners,” *Appl. Opt.* **40**, 1668–1671 (2001).
- [44] J. Butler and R. Lowe, “Beam-forming matrix simplifies design of electronically scanned antennas,” *Electron. Des.* **9**, 170–173 (1961).

Table 1. Array Configurations

T^a	N_{spacings}^b	S_{min}^c	S_{max}^c	ΔS^c	B_{max}^d
3	44	1.0	22.5	0.5	45.
6	17	1.0	9.0	0.5	45.
10	9	1.0	5.0	0.5	45.

^a The number of telescopes.

^b The number of element spacings.

^c Minimum, maximum, and step size for the element spacings.

^d The maximum baseline.

All dimensions are scaled to the telescope size L_1 .

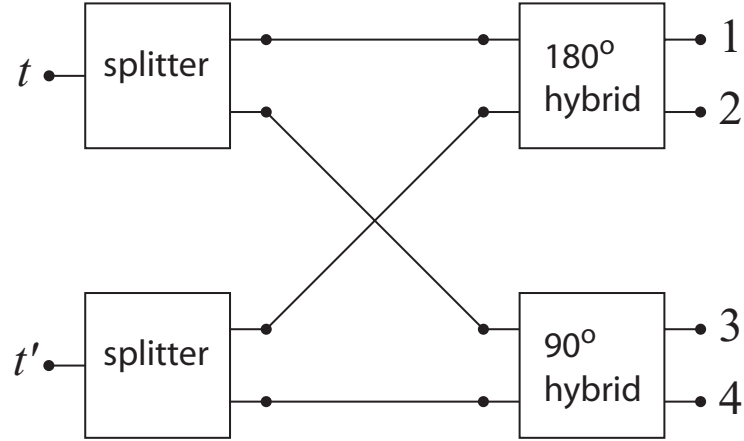


FIG. 1: A schematic diagram of the pairwise beam combination scheme for a single baseline between telescopes t and t' . The inputs t and t' on the left represent the single-mode beams from the two telescopes, after division $T - 1$ ways. The four outputs on the right are sent to photon-counting detectors.

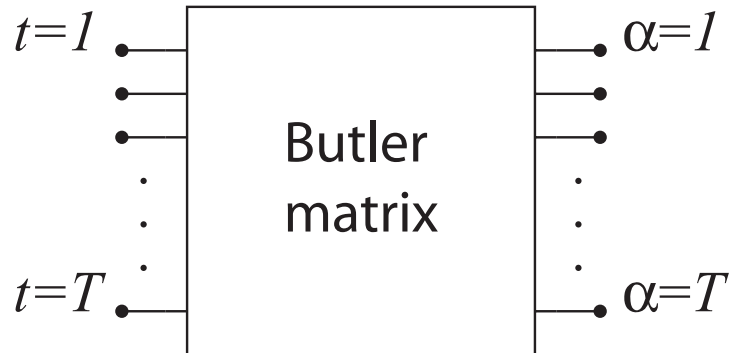


FIG. 2: A schematic diagram of the “Butler matrix” all-on-one beam combination scheme. The inputs on the left represent the single-mode beams from all T telescopes; the T outputs on the right, which are sent to photon counting detectors, each contain some contribution from all T telescope inputs.

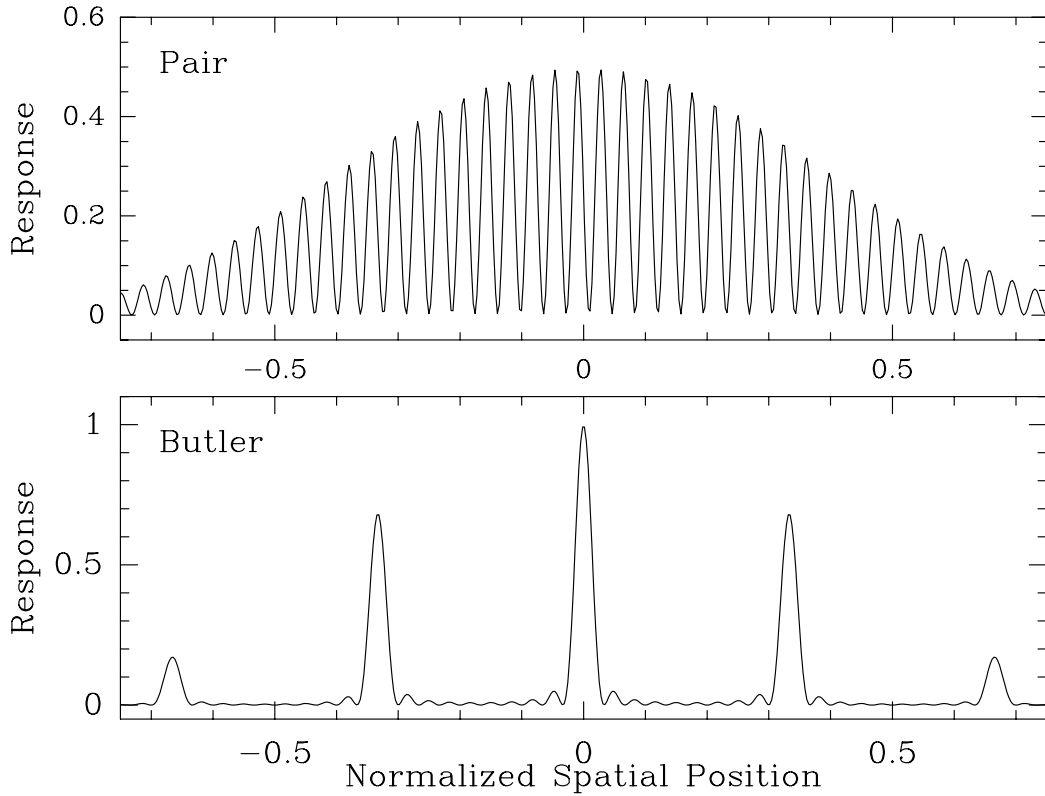


FIG. 3: This figure compares the angular response functions for pairwise and Butler beam combining. The top panel shows the typical response corresponding to two telescopes in a pair-combined array; the separation between the telescopes is $27L_1$ for this example. The bottom panel shows the typical 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 $27L_1$.

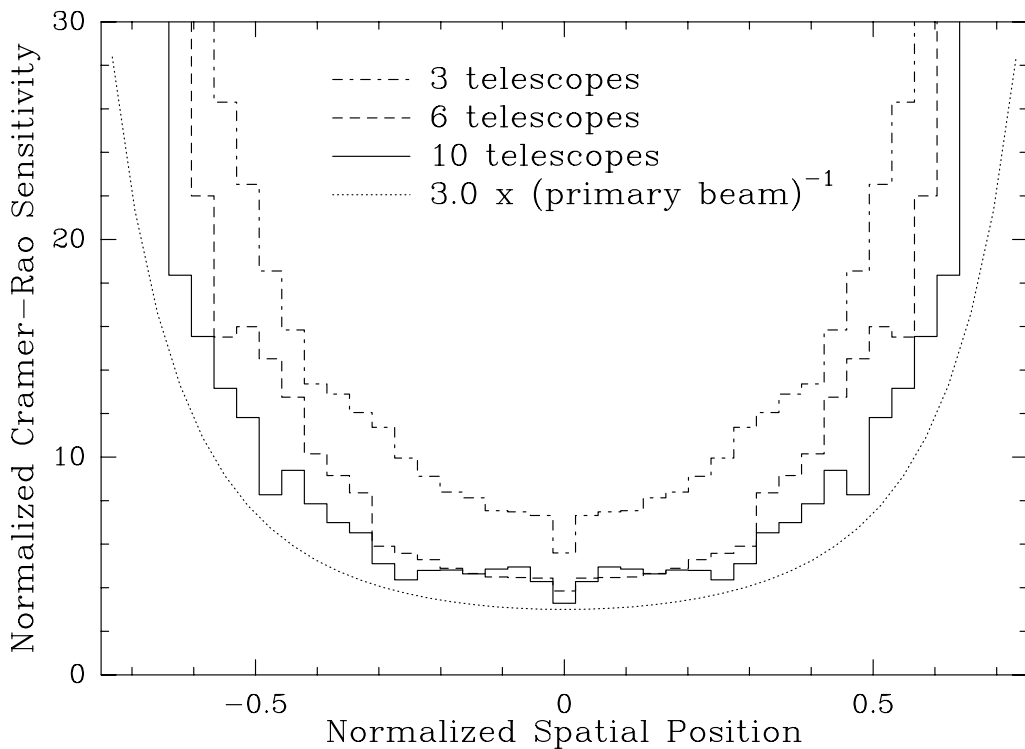


FIG. 4: Variation of the normalized sensitivity with telescope array size for the case of Butler beam combination. The source is assumed to have a uniform spatial distribution. The horizontal axis gives the spatial position θ in units of λ/L_1 ; note that the FWHP of the single-element pattern is $0.886\lambda/L_1$. 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 line shows that the sensitivity degradation toward the edges of the field of view scales as the reciprocal of the single-element beam pattern.

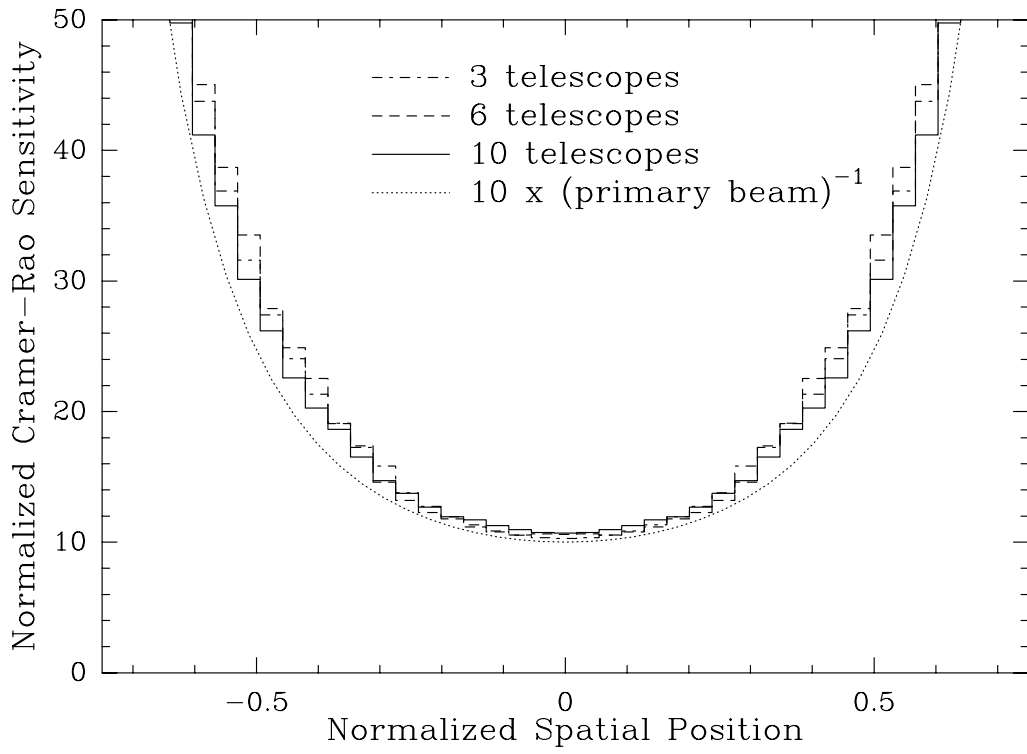


FIG. 5: This figure is similar to Fig. 4, but is calculated for the case of pairwise beam combination. The normalized sensitivity is essentially independent of array size in this case.

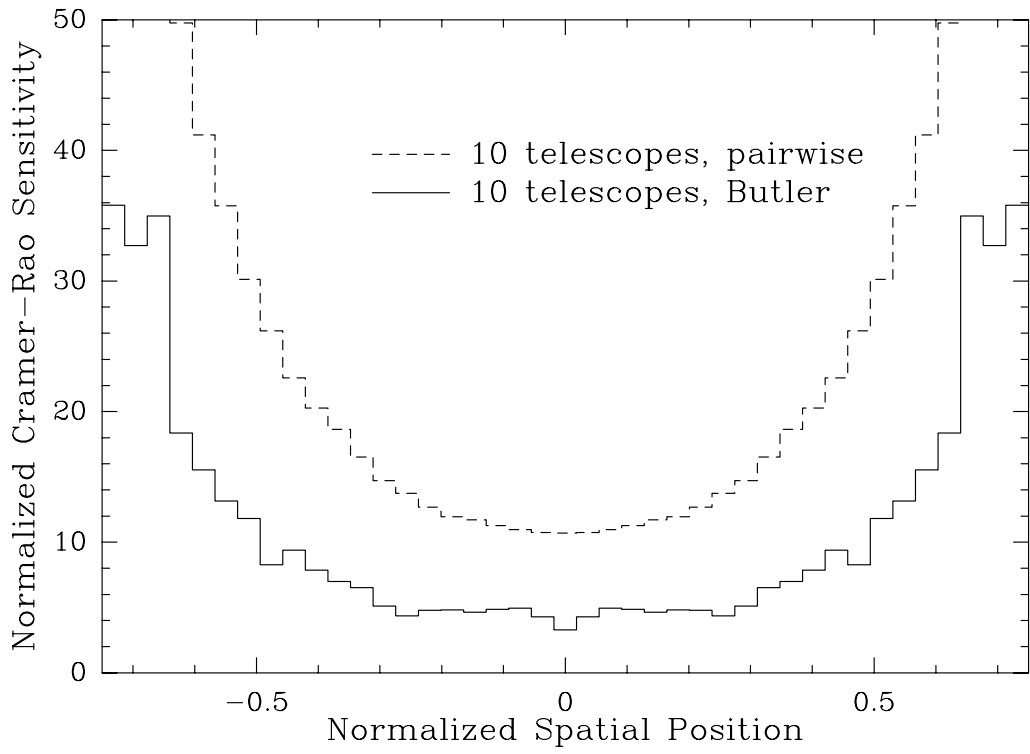


FIG. 6: Comparison of normalized sensitivities for Butler vs. pairwise beam combining for a 10-telescope array when observing uniform sources. For this case, the sensitivity advantage for Butler combining is more than a factor of 3.

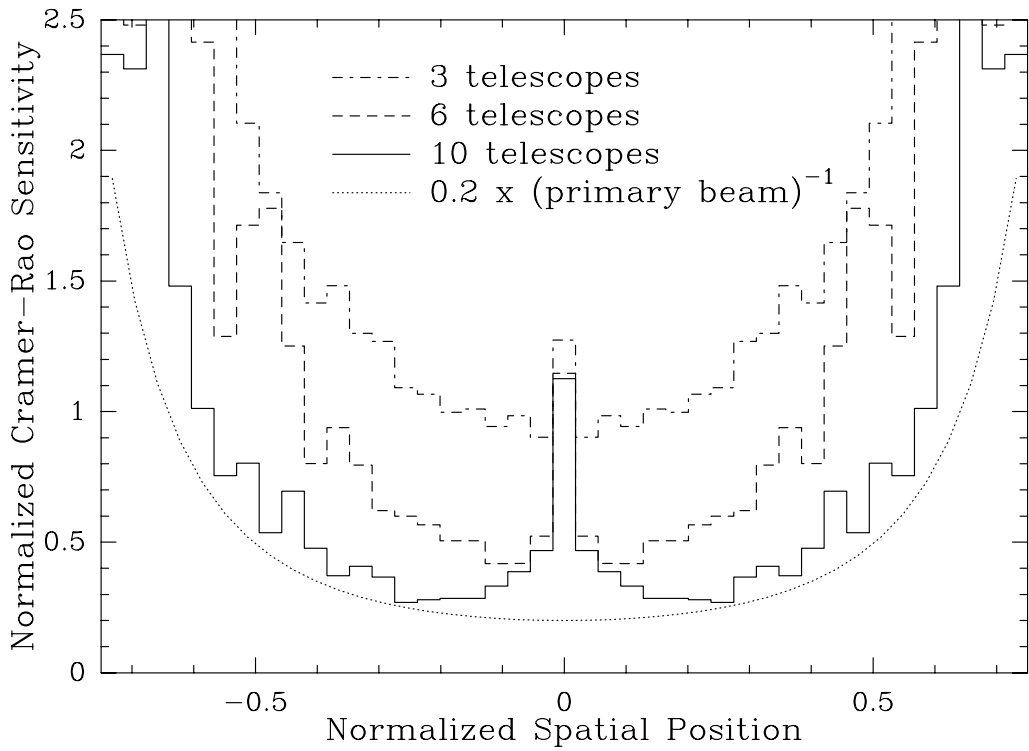


FIG. 7: This figure is similar to Fig. 4, but is calculated for the case of a point source in the center of the field instead of for a uniform source. Note that although the normalized sensitivity to the point source is nearly constant, the sensitivity for off-source pixels improves substantially with array size.

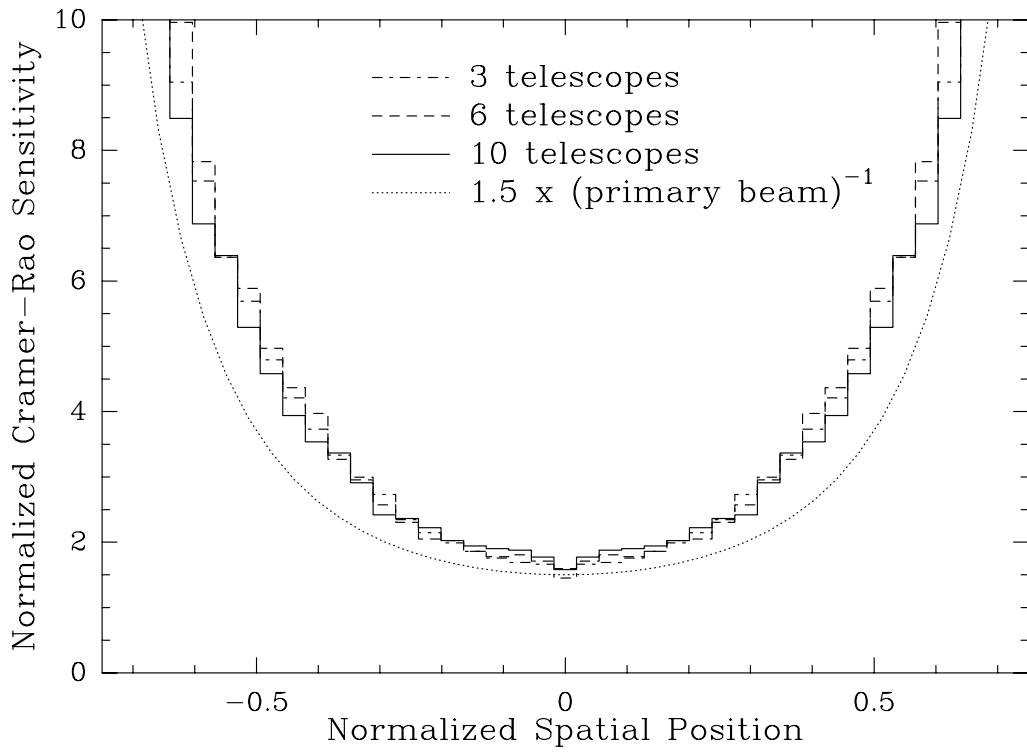


FIG. 8: This figure is similar to Fig. 7, but now pairwise beam combination is assumed. For this case, there is no sensitivity improvement with array size.

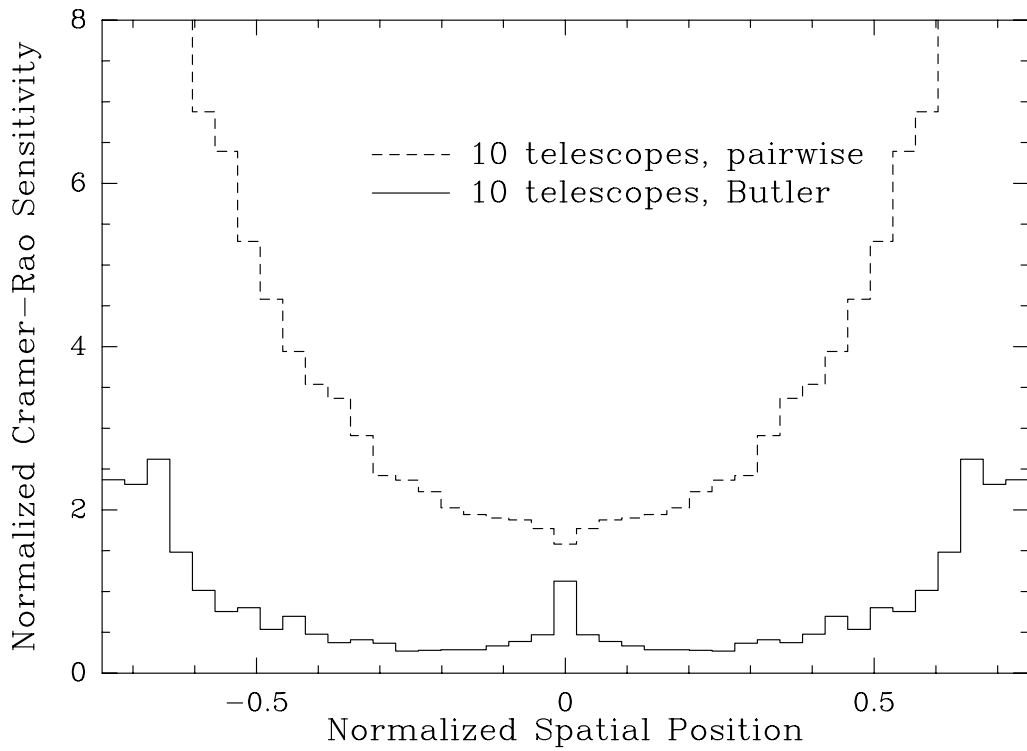


FIG. 9: This figure compares normalized point source sensitivities for Butler combination and pairwise combination for a 10-element array. Although the sensitivities to the point source itself are comparable, Butler combination is an order of magnitude more sensitive for off-source pixels.