Noise in optical synthesis images. I. Ideal Michelson interferometer

Sudhakar Prasad

Center for Advanced Studies, University of New Mexico, Albuquerque, New Mexico 87131

Shrinivas R. Kulkarni

Palomar Observatory, California Institute of Technology, Pasadena, California 91125

Received April 25, 1989; accepted July 10, 1989

We study the distribution of noise in optical images produced by the aperture synthesis technique, in which the principal source of noise is the intrinsic shot noise of photoelectric detection. The results of our analysis are directly applicable to any space-based optical interferometer. We show that the signal-to-noise ratio of images synthesized by such an ideal interferometric array is essentially independent of the details of the beam-combination geometry, the degree of array redundancy, and whether zero-spatial-frequency components are included in image synthesis. However, the distribution of noise does depend on the beam-combination geometry. A highly desirable distribution, one of uniform noise across the entire image, is obtained only when the beams from the n primary apertures are subdivided and combined pairwise on $n(n-1)/2$ detectors.

INTRODUCTION

With the advent of ground-based optical interferometers[1] and nonredundant masking on large single telescopes[2,3] it is now possible to produce aperture synthesis images of astronomical objects in much the same way as radio astronomers synthesize images by using interferometers such as the Very Large Array (VLA). Owing to the turbulence caused by the atmosphere, the sensitivity of terrestrial interferometers is limited to sources probably no fainter than 10th magnitude. This limit, when contrasted with the truly large astrophysical advances in high-resolution imaging and astrometry of faint objects, makes a space-based optical interferometer highly desirable.

The technology to build a modest-length (10–30-m) space-based interferometer exists now, and with proper funding it will be possible to put such an instrument up in space within a decade. Indeed, several proposals with realistic goals have been submitted to the National Aeronautics and Space Administration for continued studies.

The fundamental basis of synthesis imaging is the van Cittert–Zernike theorem[4]: the image is the Fourier transform of the spatial coherence function. This statement is independent of the wavelength and thus at one stroke describes image synthesis at both radio and optical wavelengths.

Despite this commonality, there are many differences between radio and optical interferometers. These differences arise in part because of technology and in part because of the fact that the signals assume chiefly photonlike characteristics at optical wavelengths and wavelike features at radio wavelengths. The dominant source of noise in radio interferometers is the additive Gaussian noise generated by the receiving apparatus, whereas optical detectors are noise-free and the dominant source of noise in optical interferometers is the Poisson fluctuations of the signal itself. The distribution of noise in a radio synthesis image is a well-studied topic, and indeed descriptions of it can be found in standard textbooks.[5] However, for the reasons discussed above, this theory is not applicable to optical interferometers. In particular, the covariance properties of the fringe phasors are markedly different for optical and radio interferometers. Given this and the burgeoning interest in optical synthesis imaging, we believed it worthwhile to undertake a systematic investigation of noise in optical synthesis images.

This paper is the first of a three-paper series. In this paper we analyze the performance of an ideal interferometer. Specifically we assume that the rays reaching each of the n elements are not phase corrupted by the intervening medium. In such an ideal interferometer the Michelson fringe phasor, which is second order in the electric field, is the best estimator, and the van Cittert–Zernike theorem can be applied directly. Thus the analysis presented here is applicable to a space- or lunar-based optical interferometer. Owing to the corruption of the wave front by the atmosphere, ground-based interferometers must use a sixth-order estimator, referred to as the bispectrum or the triple product, whose phase is the closure phase. In subsequent papers we shall extend the analysis to images synthesized from bispectrum data.

A space-based interferometer, to the first order, is essentially an n-slit Young interferometer. The absence of the atmosphere permits the use of a large aperture for each element, and coherent integration times are limited only by the changes in the spatial-frequency plane (or the uv plane, in the jargon of radio astronomy) resulting from changes in the orientation of the interferometer with respect to the source. Every coherent integration time, $n_b \equiv n(n-1)/2$ complex fringe phasors z_{gh} $(g < h = 1, 2, \ldots, n)$ are obtained, and, when enough spatial frequencies have been measured, the van Cittert–Zernike theorem can be applied to yield the object intensity distribution.

0740-3232/89/111702-13$02.00

Although the above description summarizes the basis of all astronomical interferometry, many questions arise that are peculiar to optical interferometers. These peculiarities arise because, unlike the situation at radio wavelengths, it is not possible to amplify weak optical signals without a large degradation in the signal-to-noise ratio (SNR). The following questions must be answered before a practical design for a space-based optical interferometer is considered:

(1) Does the beam-combination geometry matter? There are many different ways of combining the n beams. The simplest is to bring all the beams together to one common focus, as in a telescope. We refer to this as an ${}^nC_n$ interferometer, for which the superscript represents the total number of primary beams and the subscript represents the number of beams per detector. This notation has the additional advantage that the number of detectors, which in this case is 1, is also given by the value of the combinatoric factor ${}^nC_n$. Another possible geometry is the ${}^nC_2$ interferometer, in which each of the n primary beams is split up $n-1$ ways and the resulting $n(n-1)$ subbeams are combined pairwise on $n_b = {}^nC_2$ detectors. Other beam-combination geometries (${}^nC_3$, ${}^nC_4$, etc.) lie between these two extremes.

(2) What is the SNR in a synthesis image? How does the variance vary across the image? Is it possible to trade off sensitivity for uniform variance?

Intuitively, it is clear that beam combination should not be a major factor for an ideal interferometer and that the SNR in the map should be proportional to $\sqrt{L}$, where L is the number of photoelectrons collected by the array within the total integration time. However, this is not the consensus view, as is clear from our experience at various meetings and workshops on high angular resolution.

In this paper we report exact calculations of the sensitivity as well as the distribution of variance of ideal interferometer arrays that are assumed to be limited only by the shot noise of the photoelectric detection process. Thus we assume implicitly that the detectors do not have any dark current and do not suffer from any readout penalty. Most modern cooled photoelectric detectors or the newer solid-state photomultipliers (SSPM's) satisfy this constraint, as a result of which they are the choice detectors for optical interferometry. These calculations enable us to answer decisively the questions raised above and should lay to rest the disagreement among astronomers as to the sensitivity of an ideal interferometer.

Specifically, we have obtained expressions for the distribution of variance for the two extreme cases of beam combination: ${}^nC_2$ (Section 1) and ${}^nC_n$ (Sections 2 and 3). The results for other beam-combination geometries should lie between the results for these two cases. We find that, to better than 40%, beam combination does not affect the SNR and that the sensitivity of an interferometer is equal to the square root of the total number of photoelectrons collected by the array. There are, of course, differences in details between different beam combination geometries. These differences and trade-offs are discussed in Section 4. We conclude by arguing, among other points, that, even if an ideal noise-free amplifier were available, the SNR of an optical interferometer could not exceed the SNR estimated in this paper.

1. ${}^nC_2$ INTERFEROMETER ARRAY

Let there be n identical principal apertures from which we derive n main beams. Each main beam is divided into $n-1$ identical subbeams by the use of beam splitters. The resulting $n(n-1)$ subbeams are combined pairwise on $n_b = {}^nC_2$ detectors. The detectors are assumed to be identical array detectors with pixels ranging from 1 to P. The intensity pattern on any one detector is then given by

$$\langle I_{gh}\rangle(\mathbf{x}) = 2\langle I_0\rangle[1 + \gamma_{gh}\cos(\kappa\mathbf{x}\cdot\mathbf{B}_{gh}/d + \phi_{gh})], \quad (1.1)$$

where $\langle I_0\rangle$ is the average intensity in each subbeam at the detector, $\mathbf{B}_{gh}$ is the vector or the baseline that connects apertures g and h ($g < h = 1, 2, \ldots, n$), κ is the light-wave vector, d is the distance between the aperture and detector planes, $\mathbf{x}$ is the spatial vector in the detector plane, and $\gamma_{gh}\exp(i\phi_{gh})$ is the complex visibility function (or the complex spatial coherence function) at the separation vector $\mathbf{B}_{gh}$. In deriving Eq. (1.1) we have assumed that the incident light is spectrally narrow so that the fringe visibility depends on only the spatial correlations in the field. In an effort to reduce the clutter in the equations we henceforth drop the vector notation (boldface), but bear in mind that spatial frequencies, pixel locations, etc. are really vectors.

According to the photoelectron-detection theory, the average photoelectron count $\langle k_{gh}(p)\rangle$ at the pixel location specified by the integer index p of the detector upon which subbeams from apertures g and h are incident is proportional to $\langle I_{gh}(x)\rangle$:

$$\langle k_{gh}(p)\rangle = 2\langle K_0\rangle[1 + \gamma_{gh}\cos(p\omega_{gh} + \phi_{gh})]. \quad (1.2)$$

Here, $\langle\ldots\rangle$ denotes averaging over the photoelectron-detection process. The product $p\omega_{gh}$ is understood to be the scalar product of the pixel position vector $\mathbf{p}$ and the spatial frequency $\omega_{gh} \equiv (\omega_g - \omega_h)$, expressed in inverse pixel units.

Let $\langle C\rangle$ be the average number of photoelectrons detected by the entire array in one integration period, and let $2\langle N\rangle$ be the average number of photoelectrons per detector per integration time. Clearly, then, $\langle C\rangle = 2\langle N\rangle n_b$, and thus $\langle N\rangle = \langle C\rangle/n(n-1)$. According to Eq. (1.2), the average number of photoelectrons per detector is equal to $2\langle K_0\rangle P$, and thus $\langle K_0\rangle P = \langle N\rangle$.

Each detector yields two fringe phasors: z_{gh}, the spatial-frequency component corresponding to the vector or baseline connecting apertures g and h, and $z_{gh}{}^0$, the photoelectron count or zero-spatial-frequency component. These quantities are defined operationally as follows:

$$z_{gh} = \sum_{p=1}^{P} k_{gh}(p)\exp(-ip\omega_{gh}). \quad (1.3a)$$

As shown by Walkup and Goodman,[6] the quantity z_{gh} so defined is an optimum estimator of the actual fringe phasor under ideal photon-limited conditions such as the one that we are discussing here. We shall see below that the SNR in a map made by using this estimator has the desirable quality of being independent of the total number of pixels in the detector.

The average of the phasor z_{gh} over many coherent integration intervals is given by

$$Z_{gh} = \langle z_{gh}\rangle = \gamma_{gh}\langle N\rangle\exp(i\phi_{gh}). \quad (1.3b)$$

Throughout this paper we use uppercase symbols for the ensemble average of a random variable. Likewise,

$$z_{gh}{}^0 = \sum_{p=1}^{P} k_{gh}(p), \tag{1.3c}$$

the mean of which is

$$Z_{gh}{}^0 = 2\langle N\rangle. \tag{1.3d}$$

Note that the zero-spatial-frequency phasors are not complex but real numbers and represent the total photoelectron counts in the detectors.

There are two different ways by which the synthesized image can be constructed from the visibility data:

(1) Inversion without total counts. In this method, we use only the complex phasors z_{gh} in the Fourier inversion. The neglect of $z_{gh}{}^0$ means that the total number of photons in the map synthesized by this technique is zero! Note that, despite this unattractive feature, this is the standard method in radio astronomy.

(2) True inversion. This is a strict application of the van Cittert–Zernike theorem and uses the n_b complex phasors z_{gh} as well as the n_b zero-spatial-frequency components $z_{gh}{}^0$. The image produced by this technique has the desirable property of being nonnegative.

We now discuss the noise distribution in the maps produced by these two methods. The image produced by the application of the van Cittert–Zernike theorem suffers from two sources of noise:

(a) In practice we can obtain measurements over only a finite number of baselines or spatial frequencies, whereas the strict application of the van Cittert–Zernike theorem requires measurement of the entire continuum of spatial frequencies. This leads to errors that may be called sampling errors.

(b) The photoelectron detection process suffers from shot noise, which in turn limits the accuracy with which fringe phasors can be measured. Shot noise is governed by Poisson statistics, on account of which the variance in the photoelectron count in pixel p is equal to the average photoelectron count $\langle k(p)\rangle$. Modern deconvolution methods such as CLEAN and MEM, when applied to radio-interferometric data, appear to compensate for the sampling errors, and with some care radio synthesis images can be obtained that are limited only by the measurement uncertainties in the fringe visibility functions. (The reader is referred to Ref. 7 for a discussion of the current methods of imaging and deconvolution methods that are used commonly at radio frequencies.) In contrast to the sampling errors, there is no technique by which the effects of shot noise can be reduced. Thus in what follows we do not discuss sampling errors but restrict the study to the effects of shot noise on the maximum achievable SNR in the synthesized map. (Clearly, practical situations in which the photoelectron shot noise is small enough that sampling and other errors cannot be made smaller still lie beyond the scope of this paper.)

A note should be made here regarding the notation. Depending on our convenience, we sometimes use double indices to denote a spatial frequency such as ω_{gh}, with $g < h = 1, 2, \ldots, n$, and at other times use a single index such as ω_r, with $r = 1, 2, \ldots, n_b$.

A. Inversion without Total Counts

The synthesized image is the real portion of the Fourier transform of the spatial coherence function:

$$i_1(q) \equiv \mathrm{Re}\left[\sum_{r=1}^{n_b} z_r \exp(+iq\omega_r)\right]$$

$$= \sum_r [\mathrm{Re}(z_r)\cos(\omega_r q) - \mathrm{Im}(z_r)\sin(\omega_r q)]. \tag{1.4a}$$

Index q refers to pixels in the synthesized image; in particular, q ranges from $-Q/2$ to $+Q/2$, and $q = 0$ refers to the central pixel in the synthesized map. The variable $i_1(q)$ refers to the image obtained from one set of visibility data. Note that the sense of the Fourier transform used in Eq. (1.4a) is consistent with the definition of z_r [cf. Eq. (1.3)]. The mean map $I_1(q)$ is the average of $i_1(q)$ and is given by

$$I_1(q) = \mathrm{Re}\left[\sum_r Z_r \exp(+iq\omega_r)\right], \tag{1.4b}$$

which, by virtue of Eq. (1.3b), can be simplified to yield

$$I_1(q) = \langle N\rangle \sum_r \gamma_r \cos(\omega_r q + \phi_r). \tag{1.4c}$$

The image $I_1(q)$ is referred to as the dirty image in the parlance of radio astronomy. The dirty image is the convolution of the true image and the Fourier transform of the spatial-frequency-sampling function, which is also called the dirty beam. If the uv plane is sampled sufficiently well, then the dirty beam will be peaked with a full-width at half maximum given by the maximum extent of the uv coverage. A synthesized image can be obtained from the dirty image by any one of the popular deconvolution techniques (see, e.g., Ref. 7).

The variance $V[i_1(q)]$ in the synthesized map $i_1(q)$ [Eq. (1.4a)] is given by

$$\begin{aligned} V[i_1(q)] &\equiv \langle i_1(q)^2\rangle - \langle i_1(q)\rangle^2 \\ &= \sum_{r=1}^{n_b}\sum_{s=1}^{n_b} [\langle \mathrm{Re}(z_r)\mathrm{Re}(z_s)\rangle - \langle \mathrm{Re}(z_r)\rangle\langle \mathrm{Re}(z_s)\rangle] \\ &\quad \times \cos(\omega_r q)\cos(\omega_s q) - [\langle \mathrm{Re}(z_r)\mathrm{Im}(z_s)\rangle \\ &\quad - \langle \mathrm{Re}(z_r)\rangle\langle \mathrm{Im}(z_s)\rangle]\cos(\omega_r q)\sin(\omega_s q) \\ &\quad - [\langle \mathrm{Im}(z_r)\mathrm{Re}(z_s)\rangle - \langle \mathrm{Im}(z_r)\rangle\langle \mathrm{Re}(z_s)\rangle]\sin(\omega_r q) \\ &\quad \times \cos(\omega_s q) + [\langle \mathrm{Im}(z_r)\mathrm{Im}(z_s)\rangle - \langle \mathrm{Im}(z_r)\rangle \\ &\quad \times \langle \mathrm{Im}(z_s)\rangle]\sin(\omega_r q)\sin(\omega_s q). \end{aligned} \tag{1.5}$$

The problem of estimating the variance in the map thus reduces to one of estimating three types of covariance term: $\mathrm{cov}[\mathrm{Re}(z_r), \mathrm{Re}(z_s)]$, $\mathrm{cov}[\mathrm{Re}(z_r), \mathrm{Im}(z_s)]$, and $\mathrm{cov}[\mathrm{Im}(z_r), \mathrm{Im}(z_s)]$, where, for example,

$$\begin{aligned}\mathrm{cov}[\mathrm{Re}(z_r), \mathrm{Re}(z_s)] &\equiv \langle \mathrm{Re}(z_r)\mathrm{Re}(z_s)\rangle - \langle \mathrm{Re}(z_r)\rangle\langle \mathrm{Re}(z_s)\rangle \\ &= \sum_{p=1}^{P}\sum_{p'=1}^{P} [\langle k_r(p)k_s(p')\rangle \\ &\quad - \langle k_r(p)\rangle\langle k_s(p')\rangle]\cos(\omega_r p)\cos(\omega_s p'). \end{aligned} \tag{1.6a}$$

Similar expressions can be obtained for the other three covariance terms.

There is no correlation of the photoelectron shot noise between different detectors or between different pixels of the same detector. From the Poisson distribution of the shot noise, it then follows that

$$\langle k_r(p)k_s(p')\rangle - \langle k_r(p)\rangle\langle k_s(p')\rangle = \delta_{rs}\delta_{pp'}\langle k_r(p)\rangle. \tag{1.6b}$$

Thus Eq. (1.6a) simplifies, and only the covariance terms with $r = s$ are nonzero:

$$\begin{aligned}\mathrm{cov}[\mathrm{Re}(z_r), \mathrm{Re}(z_s)] &= V[\mathrm{Re}(z_r)]\delta_{rs} \\ &= \delta_{rs}\sum_p \langle k_r(p)\rangle\cos(\omega_r p)^2 \\ &= \langle N\rangle\delta_{rs}. \end{aligned} \tag{1.7a}$$

Using similar algebra, one can show that

$$\mathrm{cov}[\mathrm{Im}(z_r), \mathrm{Im}(z_s)] = \langle N\rangle\delta_{rs} \tag{1.7b}$$

and

$$\mathrm{cov}[\mathrm{Re}(z_r), \mathrm{Im}(z_s)] = 0. \tag{1.7c}$$

By symmetry, $\mathrm{cov}[\mathrm{Im}(z_r), \mathrm{Re}(z_s)] = 0$. Substituting the nonzero covariance elements into Eq. (1.5), we obtain

$$\begin{aligned}V[i_1(q)] &= \sum_r \{V[\mathrm{Re}(z_r)]\cos^2(\omega_r q) + V[\mathrm{Im}(z_r)]\sin^2(\omega_r q)\} \\ &= \sum_r \langle N\rangle = \frac{\langle C\rangle}{2}. \end{aligned} \tag{1.8}$$

Thus the variance is equal to half the total number of photoelectrons intercepted by the entire array, and furthermore the variance is independent of the pixel position as well as the object structure. This may be a highly desirable feature of an aperture synthesis technique, particularly when one is willing to sacrifice resolution in favor of sensitivity, e.g., in doing some kind of source count for which the uniformity of the background noise in the image plane may be the most useful requirement.

Now we consider the specific case of a point source ($\gamma_r = 1$) at the phase center ($\phi_r = 0$) for which

$$I_1(q) = \frac{\langle C\rangle}{2}\delta_{q0} \tag{1.9a}$$

and $V[i_1(q)]$ is specified by Eq. (1.8). The use of the Kronecker δ in Eq. (1.9a) is valid only when one has a dense and infinite uv coverage. However, $I_1(0) = \langle C\rangle/2$ regardless of the uv coverage. The uv coverage or the lack of it gives rise to sidelobes, which are assumed to be removed by algorithms such as CLEAN. Since the source is located at the phase center, the SNR of the central pixel in the map can be considered to be indicative of the SNR in the map:

$$\frac{I_1(0)}{\{V[i_1(0)]\}^{1/2}} = \left(\frac{\langle C\rangle}{2}\right)^{1/2}. \tag{1.9b}$$

Indeed, apart from the factor of $\sqrt{1/2}$, this is the SNR that we expect from simple physical considerations.

We clarify here that the variance given in Eq. (1.8) is the variance in an image obtained by synthesizing one single set of measurements of the n_b phasors. If the measurements were repeated m times, then both the image [Eq. (1.4c)] and the variance [Eq. (1.8)] would be scaled up by m and the SNR in the resulting map would be $(\langle L\rangle/2)^{1/2}$, where $\langle L\rangle = \langle C\rangle m$ is the total number of photoelectrons intercepted by the array over the m coherent integration levels.

B. True Inversion

There are three reasons why one may wish to use the n_b zero-spatial-frequency components $z_{gh}{}^0$:

(1) The integral of a dirty image made without using any zero-spatial-frequency counts is zero. However, the astronomical source does put out some finite nonzero power.

(2) Some pixels of a dirty image made without using the zero-spatial-frequency components could be negative. This is clearly an artifact, since a true image, made, for example, by using a telescope, is nonnegative on a pixel-by-pixel basis.

(3) Given the n_b zero-spatial-frequency components, it would appear only logical to use these n_b measurements rather than simply to throw them away.

According to the van Cittert–Zernike theorem, all the spatial-frequency components must be used to construct the images. Thus the simplest method is to include the zero-spatial-frequency components in the Fourier inversion. In our inversion, we include only positive nonzero spatial frequencies, as is clear from an inspection of Eq. (1.4a). This is a valid procedure, since the corresponding negative-frequency components are merely their complex conjugates. Thus the zero-spatial-frequency phasor, which is its own complex conjugate, must be halved (or, equivalently, all the positive-frequency terms must be doubled) before it is included in such an inversion procedure, one that suppresses all nonzero spatial frequencies of one sign. Although we shall use this weighting for the zero-frequency phasor, our calculations may be modified trivially to include a different weighting. The synthesized image is then specified by

$$i_2(q) \equiv \sum_r [\mathrm{Re}(z_r)\cos(\omega_r q) - \mathrm{Im}(z_r)\sin(\omega_r q) + (1/2)z_r{}^0], \tag{1.10a}$$

the mean value of which is

$$I_2(q) = \langle N\rangle \sum_r [\gamma_r\cos(\phi_r + \omega_r q) + 1] = I_1(q) + \frac{\langle C\rangle}{2}. \tag{1.10b}$$

We note that $I_2(q) \geq 0$ for all q, since the term $\gamma_r\cos(\phi_r + \omega_r q) + 1$ in Eq. (1.10b) is always nonnegative for all r. The average of $I_2(q)$ over the map can be shown easily to be $\langle C\rangle/2$, which is greater than zero. Finally, we demonstrate below

that using the zero-spatial-frequency phasors results in an enhanced SNR. Thus all three of the points raised at the beginning of this subsection have been illustrated.

As above, we ignore the uncertainties introduced by the sampling process and estimate the variance that is due to the shot noise of the detection process:

$$V[i_2(q)] = V[i_1(q)] + \sum_{r=1}^{n_b}\sum_{s=1}^{n_b}\{(1/2)\mathrm{cov}[\mathrm{Re}(z_r), z_s^0]\cos(\omega_r q) - (1/2)\mathrm{cov}[\mathrm{Im}(z_r), z_s^0]\sin(\omega_r q) + (1/2)\mathrm{cov}[z_r^0, \mathrm{Re}(z_s)]\cos(\omega_s q) - (1/2)\mathrm{cov}[z_r^0, \mathrm{Im}(z_s)]\sin(\omega_s q) + (1/4)\mathrm{cov}(z_r^0, z_s^0)\}. \quad (1.11)$$

As in the treatment above [cf. Eq. (1.7)], only terms with $r = s$ are nonzero. This reduces the sum in Eq. (1.11) to a single sum, which we now evaluate. We note that

$$\mathrm{cov}[\mathrm{Re}(z_r), z_r^0] = \left\langle \sum_p k_r(p)\cos(p\omega_r) \sum_{p'} k_r(p') \right\rangle - \left\langle \sum_p k_r(p)\cos(p\omega_r) \right\rangle \left\langle \sum_{p'} k_r(p') \right\rangle$$

$$= \sum_p \langle k_r(p)\rangle \cos(\omega_r p). \quad (1.12a)$$

Similarly,

$$\mathrm{cov}[\mathrm{Im}(z_r), z_r^0] = -\sum_p \langle k_p(p)\rangle \sin(\omega_r p),$$

$$\mathrm{cov}(z_r^0, z_r^0) = \sum_p \langle k_r(p)\rangle = 2\langle N\rangle, \quad (1.12b)$$

and the remaining two terms are evaluated easily by noting that $\mathrm{cov}(A, B) = \mathrm{cov}(B, A)$. These results simplify Eq. (1.11) to

$$V[i_2(q)] = \sum_r [\langle N\rangle + (1/2)\langle N\rangle + \sum_p \langle k_r(p)\rangle \times \cos(\omega_r p - \omega_r q)]$$

$$= \langle N\rangle \sum_r [(3/2) + \gamma_r \cos(\omega_r q + \phi_r)]. \quad (1.13a)$$

Using Eq. (1.4c), we simplify the above equation to obtain

$$V[i_2(q)] = (3/4)\langle C\rangle + I_1(q) = \frac{\langle C\rangle}{4} + I_2(q). \quad (1.13b)$$

Thus, in contrast to the previous method, in this method the variance is no longer uniform across the map. In particular, the variance is composed of a fixed amount ($\langle C\rangle/4$) and a variable amount that is equal to the dirty image. Indeed, such behavior can be expected on physical grounds because the additional noise comes from including the zero-spatial-frequency components, which are highly correlated with the corresponding fringe phasors. That the variance follows the dirty image is a natural consequence of the statistics of Poisson noise, viz., the variance is equal to the mean.

Now we consider the specific case of a point source at the phase center, in which case $\gamma_r = 1$ and $\phi_r = 0$. Then

$$I_2(q) = \langle C\rangle, \quad (1.14a)$$

and the SNR in the map is

$$\frac{I_2(0)}{\{V[i_2(0)]\}^{1/2}} = \left(\frac{8}{5}\right)^{1/2}\left(\frac{\langle C\rangle}{2}\right)^{1/2}. \quad (1.14b)$$

Normalizing to the previous case, we note that the SNR is enhanced by a factor $F = (8/5)^{1/2}$. Henceforth we refer to F as the enhancement factor. We shall use F as a kind of a figure-of-merit indicator. Thus inclusion of the zero-spatial-frequency components improves the SHR, but the penalty that we pay is that the variance is no longer uniform across the image.

Why is it that in Subsection 1.A, in which we excluded the zero-spatial-frequency components, we obtained uniform variance across the image, whereas in Subsection 1.B, in which we included the zero-spatial-frequency components, we obtained a nonuniform variance? This difference arises because in Subsection 1.A we found that there is no covariance or cross talk between pairs of fringe phasors. Thus the variance must be white, i.e., uniform. On the other hand, in Subsection 1.B we found that there is a finite covariance that introduces structure into the variance distribution function. This is a general result, as can be seen by considering the variance distribution function in a radio synthesis image.[8] At radio frequencies, the large source of noise is an additive Gaussian noise, which comes from the receiving electronics, etc. Kulkarni[8] found that, for sources with a flux density (S) considerably weaker than the receiver additive noise (N), fringe phasors are pairwise uncorrelated, and the variance is indeed uniform across the synthesized image, a fact that is verified every day at the VLA. However, for S comparable with N, the fringe phasors start to become correlated, as a result of which the variance is predicted to be nonuniform across the synthesized image.

2. NONREDUNDANT ^nC_n INTERFEROMETER

An ^nC_n interferometer is necessarily more complicated than an $^nC_2$ interferometer because in the former all the n beams interfere on one single detector. The n_b different fringes lie on top of one another. Application of Eq. (1.3a) with different spatial frequencies results in the extraction of the n_b fringe phasors, and the image may be synthesized in the usual fashion. One might think at the outset that image synthesis that requires the retrieval of individual spatial-frequency components is bound to be rather noisy with a single detector. However, our careful analysis proves otherwise and at the same time provides insight into improved schemes of imaging.

If the interferometer geometry allows two or more baselines to be exactly equal to one another, then the fringes in question will be indistinguishable from one another. In ground-based interferometers such a situation would be di-

sastrous, since the atmospheric phase aberrations would lead to a complete washout of the fringes on redundant baselines. In a space-based interferometer there is no aperture-dependent phase error, and thus redundancy in baseline is not a problem. On the other hand, redundancy inhibits rapid covergence of the uv plane. Thus, even in a space-based interferometer, nonredundant baseline geometry would be advantageous because it would reduce the sampling errors.

In this section we consider an ${}^nC_n$ interferometer with no redundancy of baselines, and in Section 4 we consider an ${}^nC_n$ interferometer with the maximum possible redundancy. We consider both cases because many analytical simplifications that are possible in the former case are invalid in the latter. However, we show that in either case the SNR in the map is roughly the same and, in fact, approximately equal to that of an ${}^nC_2$ interferometer.

Let the nonredundant mask consist of n identical apertures, not necessarily in a one-dimensional geometry, and let it be illuminated by a source. The classical intensity distribution of the interference pattern by the n apertures has the average value

$$\langle I(\mathbf{x})\rangle = \langle I_0'\rangle\left[n + 2\sum_{g<h}\gamma_{gh}\cos(\kappa\mathbf{x}\cdot\mathbf{B}_{gh}/d + \phi_{gh})\right], \quad (2.1)$$

where the various symbols have meanings similar to those in Eq. (1.1). Let $\langle k(\mathbf{p})\rangle$ denote the photoelectron count distribution due to $\langle I(\mathbf{x})\rangle$. As in Section 1 we discontinue the vector notation, assume that the total number of pixels is P, and note that $\langle k(p)\rangle$ is proportional to $\langle I(\mathbf{x})\rangle$:

$$\langle k(p)\rangle = \langle Q_0\rangle\left[n + 2\sum_{g<h}\gamma_{gh}\cos(p\omega_{gh} + \phi_{gh})\right]. \quad (2.2)$$

Here $\langle Q_0\rangle$ has approximately the same meaning as $\langle K_0\rangle$ in Section 1. However, since there is no beam splitting, $\langle Q_0\rangle = (n-1)\langle K_0\rangle$. In a typical setup, one can imagine integrating on a detector (usually a two-dimensional one) for a period equal to the coherent integration interval. A two-dimensional Fourier transform of the resulting image yields n_b peaks, which can be identified with the n_b fringe phasors corresponding to the n_b spatial-frequency components.

As in the treatment above, we must compute the means, variances, and covariances of the fringe phasors, z_{ij}, in order to estimate the variance in the synthesized image. Here i and j, like g and h, are aperture indices. The mean phasor on the ij baseline (i.e., the baseline connecting aperture i to aperture j) is given by

$$\begin{aligned} Z_{ij} &= \langle Q_0\rangle\sum_p \exp(-ip\omega_{ij})\left[n + 2\sum_{g<h}\gamma_{gh}\cos(p\omega_{gh} + \phi_{gh})\right] \\ &= \langle Q_0\rangle\sum_p\sum_{g<h}\gamma_{gh}\{\exp(i\phi_{gh})\exp[ip(\omega_{gh} - \omega_{ij})] \\ &\quad + \exp(-i\phi_{gh})\exp[-ip(\omega_{gh} + \omega_{ij})]\}. \end{aligned} \quad (2.3)$$

We now assume that the geometry of the interferometer is such that the baselines are nonredundant; i.e., $\omega_{ij} \neq \pm\omega_{gh}$ unless (ij) and (gh) refer to the same baseline. With this assumption, only the $g = i, h = j$ term remains in Eq. (2.3):

$$Z_{ij} = \langle M\rangle\gamma_{ij}\exp(i\phi_{ij}), \quad (2.4)$$

where we define $\langle M\rangle \equiv P\langle Q_0\rangle = \langle C\rangle/n$, $\langle C\rangle$ being, as before, the total number of photoelectrons intercepted by the array per coherent integration interval. The zero-spatial-frequency phasor, z_0, has the following average value:

$$Z_0 = \langle Q_0\rangle Pn = \langle C\rangle. \quad (2.5)$$

By using Eqs. (1.6) one can easily show that the covariance of the real components of a pair of fringe phasors is given by

$$\begin{aligned} \mathrm{cov}[\mathrm{Re}(z_{ij}), \mathrm{Re}(z_{kl})] &= \sum_p \langle k_p\rangle\cos(p\omega_{ij})\cos(p\omega_{kl}) \\ &= \langle Q_0\rangle\sum_p\left[n + 2\sum_{g<h}\gamma_{gh}\cos(p\omega_{gh} + \phi_{gh})\right] \\ &\quad \times\cos(p\omega_{ij})\cos(p\omega_{kl}). \end{aligned} \quad (2.6)$$

By writing every cosine in this equation as a sum of two exponential functions, we can see that we have terms that involve all possible combinations of two and three spatial frequencies: $\pm\omega_{ij} \pm \omega_{kl}$ and $\pm\omega_{gh} \pm \omega_{ij} \pm \omega_{kl}$. We obtain nonzero contributions from the pixel sum only when these spatial-frequency combinations vanish. For combinations of two frequencies $\pm\omega_{ij} \pm \omega_{kl}$, their nonvanishing is ensured for $i \neq k, j \neq l$ by the criterion of nonredundancy of baselines. We now impose an additional condition in order to simplify the calculations. This condition, hereafter referred to as the nonredundancy of triangles, concerns three-frequency combinations. Specifically, we assume that

$$\omega_{gh} \pm \omega_{ij} \pm \omega_{kl} \neq 0, \quad (2.7)$$

unless (gh), (ij), and (kl) form the sides of a triangle. Thus, whereas the first condition maximally constrains the baselines or vectors in any array, the second condition imposes the maximal nonredundancy condition on triangles.

By making use of the two nonredundancy conditions, it is easy to show that, for $i < j$ and $k < l$,

$$\mathrm{cov}[\mathrm{Re}(z_{ij}), \mathrm{Re}(z_{kl})] = \langle M\rangle\left(\frac{n}{2}\delta_{ik}\delta_{jl} + \frac{\gamma\cos\phi}{2}\Delta_{ij,kl}\right), \quad (2.8)$$

where the symbol $\Delta_{ij,kl} = 0$ unless the ij and kl baselines form two sides of a triangle, in which case it equals 1, and $\gamma e^{i\phi}$ is the phasor on the third side of that triangle. Similarly, one may compute the covariances of the imaginary parts of the z_{gh}'s:

$$\mathrm{cov}[\mathrm{Im}(z_{ij}), \mathrm{Im}(z_{kl})] = \langle M\rangle\left(\frac{n}{2}\delta_{ik}\delta_{jl} \mp \frac{\gamma\cos\phi}{2}\Delta_{ij,kl}\right). \quad (2.9)$$

Everywhere in this section, the upper sign (or expression) is the correct one when the sides (ij) and (kl) of the triangle meet at that vertex for which the label has a value intermediate to those of the two vertices, i.e., either when $i < j = k < l$ or when $k < l = i < j$. The lower sign (or expression) is the correct one otherwise, i.e., when $i = k$ or when $j = l$.

Mixed covariances are computed similarly. They vanish unless the two baselines have one common aperture. We obtain

$$\operatorname{cov}[\operatorname{Re}(z_{ij}), \operatorname{Im}(z_{kl})] = -\langle Q_0 \rangle \sum_p \langle k_p \rangle \cos(p\omega_{ij}) \sin(p\omega_{kl})$$

$$= -2\langle Q_0 \rangle \sum_p \sum_{g<h} \gamma_{gh} \cos(p\omega_{gh} + \phi_{gh}) \cos(p\omega_{ij}) \sin(p\omega_{kl})$$

$$= \langle M \rangle \frac{\gamma}{2} (\sin \phi) \Delta_{ij,kl} \times \begin{cases} +1 \\ \operatorname{sgn}(l-j) \text{ for } i = k, \operatorname{sgn}(i-k) \text{ for } j = l' \end{cases}, \quad (2.10)$$

where $\operatorname{sgn}(x)$ is $+1$ if $x > 0$ and -1 if $x < 0$. Furthermore,

$$\operatorname{cov}[\operatorname{Re}(z_{ij}), \operatorname{Im}(z_{kl})] = \pm \operatorname{cov}[\operatorname{Im}(z_{ij}), \operatorname{Re}(z_{kl})]. \quad (2.11)$$

Finally we need those covariances in which at least one of the two fringe phasors has zero spatial frequency. It is easy to show that

$$V(z_0) = \sum_p \langle k_p \rangle = Z_0 = \langle C \rangle, \quad (2.12)$$

$$\operatorname{cov}[z_0, \operatorname{Re}(z_{ij})] = \langle Q_0 \rangle \sum_p \cos(p\omega_{ij}) \times \left[n + 2 \sum_{g<h} \gamma_{gh} \cos(p\omega_{gh} + \phi_{gh}) \right]$$

$$= \langle M \rangle \gamma_{ij} \cos \phi_{ij} \quad (2.13a)$$

and

$$\operatorname{cov}[z_0, \operatorname{Im}(z_{ij})] = \langle M \rangle \gamma_{ij} \sin \phi_{ij}. \quad (2.13b)$$

This is a convenient point at which to pause and discuss the implications of the nonzero covariances that we found in Eqs. (2.8)–(2.11). In Subsection 1.B we found that $z_{gh}{}^0$ were correlated with z_{gh}. That was not at all surprising, since both z_{gh} and $z_{gh}{}^0$ were derived from the same data. However, here we find the surprising result that the covariance between different fringe phasors is not zero. This is unexpected, since it is assumed commonly that the fringe phasors are uncorrelated. Indeed, this is the case in radio interferometers, because of which the SNR of a radio synthesis image (which is, after all, a linear combination of the fringe phasors) is equal to $(n_b)^{1/2} \times S$, where S is the SNR of one fringe phasor. We show below that, despite the apparent nonzero covariances, the SNR in an optical synthesis map is equal to $(n_b)^{1/2} \times S \propto \langle C \rangle^{1/2}$. Thus we conclude that in the map-making process the covariance terms cancel each other.

A. Inversion without Total Counts

By following the formulation in Subsection 1.A we find the mean synthesized image to be

$$\langle I_3(q) \rangle = \langle M \rangle \sum_{i<j} \gamma_{ij} \cos(q\omega_{ij} + \phi_{ij}). \quad (2.14)$$

To evaluate the variance, we first expand it in terms of the covariances of the individual fringe phasors. The expression is

$$V[i_3(q)] = \sum_{i<j} \sum_{k<l} \{ \operatorname{cov}[\operatorname{Re}(z_{ij}), \operatorname{Re}(z_{kl})] \cos(q\omega_{ij}) \cos(q\omega_{kl}) + \operatorname{cov}[\operatorname{Im}(z_{ij}), \operatorname{Im}(z_{kl})] \sin(q\omega_{ij}) \sin(q\omega_{kl}) - \operatorname{cov}[\operatorname{Re}(z_{ij}), \operatorname{Im}(z_{kl})] \cos(q\omega_{ij}) \sin(q\omega_{kl}) - \operatorname{cov}[\operatorname{Im}(z_{ij}), \operatorname{Re}(z_{kl})] \sin(q\omega_{ij}) \cos(q\omega_{kl}) \}. \quad (2.15)$$

The first two terms in the sum can be evaluated easily by using Eqs. (2.8) and (2.9). The remaining two terms are simplified best by noting that they are nonzero only if they involve baselines that form sides for a triangle, for which we have discussed two possibilities, summarized in Eq. (2.10). We display these terms explicitly in the following expression:

$$V[i_3(q)] = \frac{\langle M \rangle}{2} \Bigg\{ n \sum_{i<j} [\cos^2(q\omega_{ij}) + \sin^2(q\omega_{ij})] + \sum_{i<j} \sum_{k<l} \Delta_{ij,kl} \cos[q(\omega_{ij} \pm \omega_{kl})] \gamma \cos \phi - \sum_{i<j=k<l} \sin[q(\omega_{ij} + \omega_{jl})] \gamma_{il} \sin \phi_{il} - \sum_{k<l=i<j} \sin[q(\omega_{lj} + \omega_{kl}) \gamma_{kj} \sin \phi_{kj} - \sum_{j<l} (n - l + j - 1) \sin[q(\omega_{il} - \omega_{ij}) \gamma_{jl} \sin \phi_{jl} - \sum_{l<j} (n - j + l - 1) \sin[q(-\omega_{il} + \omega_{ij})] \gamma_{lj} \sin \phi_{lj} \Bigg\}. \quad (2.16)$$

The triple sums in the third and fourth terms on the right-hand side of Eq. (2.16) are equal to each other, as are the double sums in the last two terms. This can be seen clearly if the sum indices are relabeled. Furthermore, the jth sum of the third term (and the lth sum in the fourth term) is trivial, since $\omega_{ij} + \omega_{jl} = \omega_{il}$ is independent of j. It thus merely produces a factor of $l - i - 1$. We shall also now interpret, in the second sum above, the symbol $\Delta_{ij,kl}$ explicitly for the two cases for which it does not vanish. The resulting expression for $V[i_3(q)]$ is then

$$V[i_3(q)] = \frac{\langle M\rangle}{2}\Bigg\{nn_b + 2\sum_{i<j=k<l}\cos(q\omega_{il})\gamma_{il}\cos\phi_{il} + 2\sum_{j<l}(n-l+j-1)\cos(q\omega_{jl})\gamma_{jl}\cos\phi_{jl} - 2\sum_{i<l}(l-i-1)\sin(q\omega_{il})\gamma_{il}\sin\phi_{il} - 2\sum_{j<l}(n-l+j-1)\sin(q\omega_{jl}\gamma_{jl}\sin\phi_{jl}\Bigg\}. \quad (2.17)$$

By relabeling the indices slightly and combining the various sums, using simple trigonometric identities, we obtain the following final expression:

$$V[i_3(q)] = \frac{\langle M\rangle}{2} \times\left[nn_b + 2(n-2)\sum_{i<j}\gamma_{ij}\cos(q\omega_{ij}+\phi_{ij})\right]. \quad (2.18)$$

The variance is seen to consist of a constant component $n_b\langle C\rangle/2$ and a comparable variable component, equal to the dirty image, which is reminiscent of the situation considered in Subsection 1.B. Note the curious fact that the nonconstant component disappears for $n = 2$.

Consider a point source at the phase center, for which $\gamma_{ij} = 1$ and $\phi_{ij} = 0$. Then the SNR of the central pixel may easily be evaluated to be

$$\frac{I_3(0)}{\{V[i_3(0)]\}^{1/2}} = \left(\frac{\langle C\rangle}{2}\right)^{1/2}\left(\frac{2n-2}{3n-4}\right)^{1/2}. \quad (2.19)$$

The enhancement factor $F = [(2n-2)/(3n-4)]^{1/2}$ is unity for $n = 2$ and decreases steadily to $\sqrt{2/3}$ as the number of apertures increases. Thus this interferometer is not quite so efficient as the ${}^nC_2$ interferometer.

B. True Inversion

The mean and the variance of the map constructed by including z_0 are given by appending to Eqs. (2.14) and (2.15) terms that arise from the inclusion of z_0 in the Fourier inversion. Using Eqs. (2.12) and (2.13), one has

$$I_4(q) = \langle M\rangle\left[\frac{n}{2} + \sum_{i<j}\gamma_{ij}\cos(q\omega_{ij}+\phi_{ij})\right] \quad (2.20)$$

and

$$V[i_4(q)] = V[i_3(q)] + \frac{1}{4}V(z_0) + \sum_{i<j}\mathrm{cov}[z_0, \mathrm{Re}(z_{ij})]\cos(q\omega_{ij}) - \sum_{i<j}\mathrm{cov}[z_0, \mathrm{Im}(z_{ij})]\sin(q\omega_{ij}) = \frac{\langle M\rangle}{2}\left[n\left(\frac{1}{2}+n_b\right) + 2(n-1)\sum_{i<j}2(j-1)\gamma_{ij} \times\cos(q\omega_{ij}+\phi_{ij})\right]. \quad (2.21)$$

Clearly, even for $n = 2$, a single nontrivial baseline, the variance is not uniform throughout the map. However, the SNR at the map center for a point source ($\gamma_{ij} = 1$, $\phi_{ij} = 0$) is

$$\frac{I_4(0)}{\{V[i_4(0)]\}^{1/2}} = \left(\frac{\langle C\rangle}{2}\right)^{1/2}\left(\frac{2n^2}{3n^2-5n+3}\right)^{1/2}, \quad (2.22)$$

which is larger by a factor of $\sqrt{8/5}$, for $n = 2$, than for the case in Subsection 2.A, in which z_0 is excluded. However, as in Subsection 2.A, for large n the enhancement factor F attains an asymptotic value of $\sqrt{2/3}$.

3. MAXIMALLY REDUNDANT ${}^nC_n$ INTERFEROMETER

In Section 2 we considered an ${}^nC_n$ interferometer using nonredundant baselines. We imposed an additional condition of nonredundant triangles [see relation (2.7)] that simplified the calculation of the covariances. Intuitively it is clear that neither of these conditions should affect the final sensitivity of the interferometer in an essential way. In order to demonstrate this point, we now consider an interferometer for which these two conditions are not satisfied, viz., an array with maximal redundancy.

Specifically, we consider here an array of n regularly spaced apertures in a one-dimensional geometry. For this mask there are $(n-1)$ distinct spatial frequencies $\omega_0, 2\omega_0, \ldots, (n-1)\omega_0$, where ω_0 is the fundamental frequency corresponding to any two successive apertures. Clearly the spatial frequency $r\omega_0$ ($1 \le r \le n-1$) is $(n-r)$-fold redundant.

Owing to its complexity, we restrict the sensitivity analysis to the case of a point source at the phase center. The average photoelectron count is given by

$$\langle k(p)\rangle = \langle Q_0\rangle\left[n + 2\sum_{r=1}^{n-1}(n-r)\cos(pr\omega_0)\right]. \quad (3.1)$$

The constant term in Eq. (3.1) is the zero-spatial-frequency component. The fringe phasor z_r for spatial frequency $r\omega_0$ is defined as in Eq. (1.3b):

$$z_r = \sum_{p=1}^{P}\langle k(p)\rangle\exp(-ipr\omega_0) \qquad (0 \le r \le n-1). \quad (3.2a)$$

Its mean value is

$$\langle z_r\rangle = \langle M\rangle(n-r), \quad (3.2b)$$

to derive which we made use of the fact that, for $P \gg 1$,

$$\sum_p \exp(ipm\omega_0) \simeq P\delta_{m,0}. \quad (3.3)$$

Here, as in Section 2, $\langle C\rangle = n\langle M\rangle$ is the average number of photoelectrons in one coherent integration time across the entire detector. The mean source flux or the mean value of the central pixel is given by the sum of all the $\langle z_r\rangle$'s.

We must calculate the covariances of the real and imaginary parts of z_r in order to estimate the variance in the image. As in Section 2,

$$\operatorname{cov}[\operatorname{Re}(z_r), \operatorname{Re}(z_s)] = \sum_p \sum_{p'} \operatorname{cov}[k(p), k(p')]\cos(pr\omega_0) \times \cos(p's\omega_0)$$

$$= \sum_p \langle k(p)\rangle \cos(pr\omega_0)\cos(ps\omega_0), \quad (3.4)$$

in which we have used the statistical independence of the photoelectron counts at two different pixels and the fact that these counts are Poisson distributed.

By substituting the expression in Eq. (3.1) for $\langle k(p)\rangle$ into Eq. (3.4) and making use of relation (3.3), one may show that

$$\operatorname{cov}[\operatorname{Re}(z_r), \operatorname{Re}(z_s)] = \frac{\langle M\rangle}{2}[n\delta_{rs} + (n - |r-s|)(1-\delta_{rs}) + (n-r-s)\Theta(n-r-s)], \quad (3.5)$$

where the symbol $\Theta(i)$ vanishes for all $i \le 0$ and equals 1 for $i \ge 1$. In a similar fashion one may show that

$$\operatorname{cov}[\operatorname{Im}(z_r), \operatorname{Im}(z_s)] = \frac{\langle M\rangle}{2}[n\delta_{rs} + (n - |r-s|)(1-\delta_{rs}) - (n-r-s)\Theta(n-r-s)]. \quad (3.6)$$

Finally, all the mixed covariances, namely, $\operatorname{cov}[\operatorname{Re}(z_r), \operatorname{Im}(z_s)]$, may be shown to vanish identically. We are now ready to calculate the variace in the map. However, in order not to detract from the physical discussions of imaging, we relegate the details of the long calculation to Appendix A.

A. Inversion without Total Counts

The calculations are straightforward, and we merely summarize the results obtained in Appendix A. The mean is found to be

$$I_5(q) = \langle M\rangle \left\{ \frac{n}{2}\cos\left(\frac{nx}{2}\right)\frac{\sin\left[(n-1)\frac{x}{2}\right]}{\sin\left(\frac{x}{2}\right)} - \sin\left(\frac{nx}{2}\right)\frac{d}{dx}\frac{\sin\left[\frac{(n-1)x}{2}\right]}{\sin\left(\frac{x}{2}\right)} \right\}_{x=q\omega_0}, \quad (3.7)$$

and the variance is

$$V[i_5(q)] = \frac{\langle M\rangle}{2}\left\{ \frac{n\sin^2\left(\frac{n-1}{2}x\right)}{\sin^2\left(\frac{x}{2}\right)} + \frac{n(n-2)}{2} \times \cos\left(\frac{nx}{2}\right)\frac{\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} + \frac{n}{2}\sin\left(\frac{nx}{2}\right)\frac{d}{dx}\frac{\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} + \frac{d}{dx}\frac{\cos\left(\frac{n-1}{2}x\right)\sin\left(\frac{n-1}{2}x\right)}{\sin^2\left(\frac{x}{2}\right)} - (n-1)\frac{d}{dx}\cot\left(\frac{x}{2}\right) - \frac{n-2}{2}\frac{d}{dx}\frac{\sin\left(\frac{nx}{2}\right)\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} + \cos\left(\frac{nx}{2}\right)\frac{d^2}{dx^2}\frac{\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} \right\}_{x=q\omega_0}. \quad (3.8)$$

At the phase center,

$$I_5(0) = \langle C\rangle\frac{(n-1)}{2}, \quad (3.9)$$

$$V[i_5(0)] = \frac{\langle C\rangle}{12}(5n^2 - 9n + 4), \quad (3.10)$$

leading to a SNR at the phase center of

$$\frac{I_5(0)}{\{V[i_5(0)]\}^{1/2}} = F\left(\frac{\langle C\rangle}{2}\right)^{1/2}, \quad (3.11)$$

where

$$F = \left[\frac{6(n-1)}{5n-4}\right]^{1/2} \quad (3.12)$$

is our enhancement factor. For $n = 2$ we find $F = 1$, and the value of F in the limit of large n is $\sqrt{6/5}$.

B. True Inversion

As in Subsection 3.A, we merely summarize the results. The reader is referred to Appendix A for the details of the calculations. Including the zero-spatial-frequency component in the Fourier inversion, we obtain the following mean and variance at pixel location q:

$$I_6(q) = \langle N\rangle \left(\frac{n}{2} + \frac{n}{2}\cos\left(\frac{nx}{2}\right)\frac{\sin(n-1)\frac{x}{2}}{\sin\left(\frac{x}{2}\right)} - \sin\frac{nx}{2}\frac{\mathrm{d}}{\mathrm{d}x}\left\{\frac{\sin(n-1)\frac{x}{2}}{\sin\left(\frac{x}{2}\right)}\right\}\right)_{x=q\omega_0} \tag{3.13}$$

and

$$V[i_6(q)] = V[i_5(q)] - \frac{3}{4}\langle M\rangle n + \langle M\rangle \left[n\cos\left(\frac{n-1}{2}\right)x \times \frac{\sin\left(\frac{nx}{2}\right)}{\sin\left(\frac{x}{2}\right)} - \frac{\mathrm{d}}{\mathrm{d}x}\sin\left(\frac{n-1}{2}\right)x \frac{\sin\left(\frac{nx}{2}\right)}{\sin\left(\frac{x}{2}\right)}\right]_{x=q\omega_0}, \tag{3.14}$$

where $V[i_5(q)]$ is given by Eq. (3.8).

These expressions are rather opaque for general values of q, but at the phase center of the map, $q = 0$, they take quite simple forms:

$$I_6(0) = \langle C\rangle \frac{n}{2} \tag{3.15}$$

and

$$V[i_6(0)] = \frac{\langle C\rangle}{12}(5n^2 - 3n + 1). \tag{3.16}$$

Thus the SNR at the phase center is

$$\frac{I_6(0)}{\{V(i_6(0))\}^{1/2}} = F\left(\frac{\langle C\rangle}{2}\right)^{1/2}, \tag{3.17}$$

where F, the enhancement factor, is given by

$$F = \left(\frac{6n^2}{5n^2 - 3n + 1}\right)^{1/2}. \tag{3.18}$$

For $n = 2$, when z_0 is included in the reconstruction process, F is enhanced from 1 to $\sqrt{8/5}$. The limiting value of F for large n is $\sqrt{6/5}$.

4. DISCUSSION AND CONCLUSION

In this paper we have analyzed the distribution of noise in the image synthesized by an ideal n-element optical interferometer. By an ideal interferometer we mean one for which the rays reaching the elements are assumed not to be corrupted by local agents such as the atmosphere. Thus the analysis reported here is applicable to space- or lunar-based optical interferometers. We have also assumed that the detectors are photon-counting detectors without significant dark current or readout noise.

At radio wavelengths the availability of low-noise amplifiers makes it possible to make multiple copies of the signal at each element and to detect the fringes on a pair-by-pair basis. Unfortunately, at optical wavelengths, amplification without substantial degradation in the SNR of the signal is not possible. This is most important for low photon numbers per coherence volume, a condition that characterizes faint optical sources. For each mode of the input field, if $G - 1$ copies of the input photons are generated by an ideal, phase-preserving amplifier, then at least $G - 1$ noise (spontaneously emitted) photons are added incoherently by the amplifier as well.[9,10] These noise photons will in general overwhelm the amplified signal. Thus one must address the issue of how to combine beams without being able to amplify them faithfully.

There are many possibilities, and we have studied two extreme cases: (1) the so-called ${}^nC_2$ interferometer, in which the beam from each element is split into $n - 1$ subbeams and the resulting $n(n - 1)$ subbeams are combined pairwise onto $n_b = {}^nC_2$ detectors, and (2) an ${}^nC_n$ interferometer, in which all the beams are combined on one detector. We have performed a complete covariance analysis and evaluated the variance across the synthesized image. Our calculations neglect any errors caused by an incomplete sampling of the spatial frequencies.

Our most important result is that the SNR in the synthesized image, defined operationally as the SNR for detecting a point source, for either kind of array is equal to $F(\langle L\rangle/2)^{1/2}$, where $\langle L\rangle$ is the total number of photoelectrons collected by the array. The values of F, the enhancement factor, is approximately unity for both the geometries. By including the zero spatial frequencies (i.e., the total counts in the detectors) one can increase the value of F to a maximum of $\sqrt{8/5}$.

This point is illustrated graphically in Fig. 1, in which we display our results for the enhancement factor F of the SNR as a function of the number of array elements for all six interferometers considered in the present study. That the SNR of an ${}^nC_2$ array (the horizontal dashed lines labeled F_1 and F_2), for a given total number of photoelectrons detected by the entire array, is independent of n arises from the fact that individual fringe phasors are detected on independent detectors. What is most striking about the graph is that the SNR is more or less independent of the details of the array, whether it is ${}^nC_2$ or ${}^nC_n$ or whether it is redundant. The sensitivity of ideal Michelson interferometers is limited solely by the total number of photoelectrons detected by the entire array and not by how individual beams are combined on the detectors. Thus, if detectors are limited only the photoelectron-counting noise, then the sensitivity of an ${}^nC_r$ array should be qualitatively independent of r, the number of subbeams per detector. Thus our most important conclusion is that the beam-combination geometry should not be a critical issue in the design of a space interferometer.

We find that the ${}^nC_2$ array with the zero-spatial-frequency components excluded yields uniform variance across synthesized images. Including the zero-spatial-frequency components for the ${}^nC_2$ array results in the variance's mimicking the dirty image. The variance is not uniform for the ${}^nC_n$ case regardless of whether the zero-spatial-frequency components are included. Overall, we favor the ${}^nC_2$ array for its constant-variance feature.

Some detectors, such as charge-coupled detectors (CCD's), have a small but nonneglible readout noise. Improvements in detector technology may make CCD's, which have a high quantum efficiency compared with the modern cooled pho-

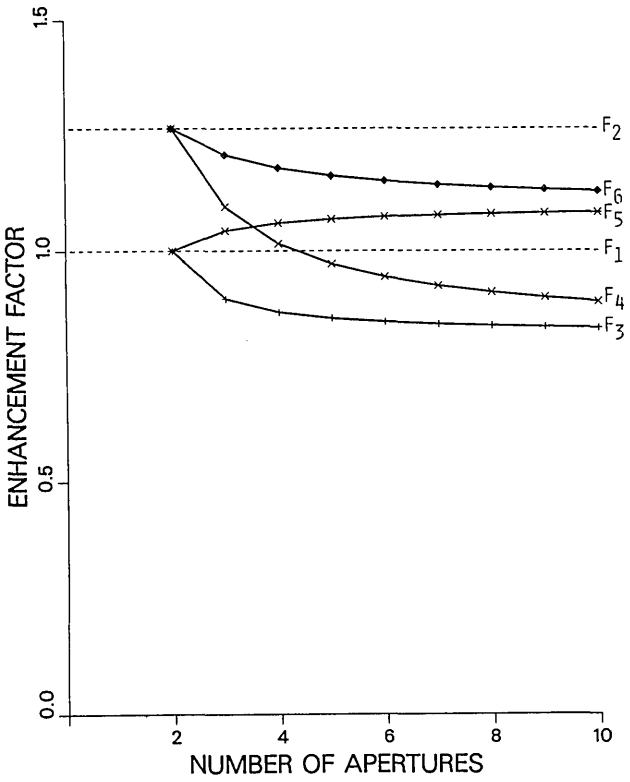

Fig. 1. Enhancement factor F of SNR versus the number n of apertures in the array. F_1 and F_2 refer to the ${}^nC_2$ array without and with the zero frequency, F_3 and F_4 refer to the nonredundant ${}^nC_n$ array, and F_5 and F_6 refer to the maximally redundant ${}^nC_n$ array.

toelectric detectors, suitable detectors for interferometry.[11] For such detectors the ${}^nC_2$ array is not suitable because beam splitting reduces the single strength. The performance of the interferometer would degrade rapidly when $\langle K_0 \rangle$ became comparable with the read noise. Thus in this case we advocate an ${}^nC_n$ detector.

Why is it that the ${}^nC_2$ and ${}^nC_n$ interferometers are more or less equally sensitive? This is by no means an obvious result. In an ideal ${}^nC_2$ interferometer each of the two beams incident upon any one detector is weaker by a factor of $(n-1)$ than the original beam from each aperture. By contrast, in an ${}^nC_n$ array all the beams are incident upon a single detector and therefore do not suffer from beam splitting. Thus it may appear that the ${}^nC_n$ array should be superior to the ${}^nC_2$ array. However, the compensating factor is that the fringe visibilities, say, for a point source, are equal to unity for the ${}^nC_2$ interferometer, whereas they are $2/n$ for the ${}^nC_n$ interferometer. Thus the SNR of the synthesized image, which, after all, a linear combination of the fringe phasors, is essentially the same for both geometries. On the other hand, were we to use an estimator such as the bispectrum, this would no longer be true, and in that case beam splitting would degrade the sensitivity of the interferometer.

Our analysis shows that the limiting sensitivity of an ideal optical interferometer is a result of the counting fluctuations in the signal itself. Thus when the signal is so weak that we are detecting only a few photons, then the SNR is of order 1, and nothing can be done to improve it. Specifically, one might consider using an amplifier to improve the signal, but since any real amplifier would add quantum noise, the SNR of the amplified signal would in fact be lower than that of the input signal. This argument should settle the issue of the (nonexistent) role of amplifiers for optical interferometry.

APPENDIX A: DERIVATION OF EQS. (3.7), (3.8), (3.13), AND (3.14)

We shall first derive Eqs. (3.7) and (3.8) for the mean and the variance of the map at pixel q and then use those results to deduce Eqs. (3.13) and (3.14). Without the zero-frequency phasor, the map $i_5(q)$ defined by the relation

$$i_5(q) = \mathrm{Re}\left[\sum_{r=1}^{n-1} z_r \exp(iqr\omega_0)\right] \tag{A1}$$

takes the following mean value:

$$I_5(q) = \langle M \rangle \mathrm{Re}\left[\sum_{r=1}^{n-1} (n-r)\exp(iqr\omega_0)\right], \tag{A2}$$

where use is made of Eq. (3.2b). By writing this sum in terms of the sum $\sum_r \exp(iqr\omega_0)$ and its derivative, we may evaluate it easily to obtain the result of Eq. (3.7).

To derive Eq. (3.8), we note that we may write $V[i_5(q)]$ as

$$V[i_5(q)] = \sum_{r=1}^{n-1}\sum_{s=1}^{n-1} \{\mathrm{cov}[\mathrm{Re}(z_r), \mathrm{Re}(z_s)]\cos(qr\omega_0)\cos(qs\omega_0) + \mathrm{cov}[\mathrm{Im}(z_r), \mathrm{Im}(z_s)]\sin(qr\omega_0)\sin(qs\omega_0)\}, \tag{A3}$$

since all mixed covariances are zero. We rewrite Eq. (A3), by separating the $r = s$ terms and the $r \neq s$ terms, as

$$\begin{aligned} V[i_5(q)] = {} & \sum_{r=1}^{[(n-1)/2]} \{V[\mathrm{Re}(z_r)]\cos^2(qr\omega_0) + V[\mathrm{Im}(z_r)]\sin^2(qr\omega_0)\} \\ & + \sum_{r=[(n+1)/2]}^{n-1} \{V[\mathrm{Re}(z_r)]\cos^2(qr\omega_0) \\ & + V[\mathrm{Im}(z_r)]\sin^2(qr\omega_0)\} + 2\sum_{r>s=1}^{n-1} \{\mathrm{cov}[\mathrm{Re}(z_r), \\ & \mathrm{Re}(z_s)]\cos(qr\omega_0)\cos(qs\omega_0) + \mathrm{cov}[\mathrm{Im}(z_r), \mathrm{Im}(z_s)] \\ & \times \sin(qr\omega_0)\sin(qs\omega_0)\}, \end{aligned} \tag{A4}$$

where, for the summation indices, $[x]$ is the greatest integer function, which is to say that $[x]$ = the largest integer that is $\leq x$. The single sums in Eq. (A4) are obtained easily by using Eqs. (3.5) and (3.6) for $r = s$. They add up to the expression

$$\begin{aligned} SS &= \frac{\langle M \rangle}{2}\left[\sum_{r=1}^{n-1} n + \sum_{r=1}^{[(n-1)/2]} (n-2r)\cos(2qr\omega_0)\right] \\ &= \frac{\langle M \rangle}{2}\left\{n(n-1) + \left[n\,\mathrm{Re} - \frac{\partial}{\partial(q\omega_0)}\mathrm{Im}\right] \times \sum_{r=1}^{[(n-1)/2]} \exp(2iq\omega_0)\right\}. \end{aligned} \tag{A5}$$

The double sums in Eq. (A4) may be simplified in terms of the following simpler double sums:

$$S_1 = \sum_{r>s=1}^{n-1} \exp[iq(r-s)\omega_0] \tag{A6}$$

and

$$S_2 = \sum_{r>s=1}^{n-1} \exp[iq(r+s)\omega_0]\Theta(n-r-s). \tag{A7}$$

In terms of SS and

$$S \equiv S_1 + S_2, \tag{A8}$$

we may write $V[i_5(q)]$ as

$$V[i_5(q)] = SS + \langle M\rangle\left[n\,\mathrm{Re} - \frac{\partial}{\partial(q\omega_0)}\,\mathrm{Im}\right]S. \tag{A9}$$

We now turn to an evaluation of S_1 and S_2.

By making use of the simple geometric-sum formula,

$$\sum_{s=1}^{r-1} \exp(-isq\omega_0) = \frac{\exp(-iq\omega_0) - \exp(-irq\omega_0)}{1-\exp(-iq\omega_0)}, \tag{A10}$$

we may express S_1 as

$$S_1 = \frac{1}{1-\exp(-iq\omega_0)} \sum_{r=2}^{n-1} \{\exp[i(r-1)q\omega_0] - 1\}. \tag{A11}$$

We now extend the lower limit on the preceding sum to $r = 1$ without changing the result and use Eq. (A10) again to obtain

$$S_1 = \frac{-i\exp[i(n-1)q\omega_0/2]\sin[q(n-1)\omega_0/2]}{2\sin^2(q\omega_0/2)} + \frac{i(n-1)\exp(iq\omega_0/2)}{2\sin(q\omega_0/2)}. \tag{A12}$$

We evaluate S_2 by noting that the restricted double sum is over terms symmetric under the interchange of r and s. We may therefore transform it into an unrestricted double sum minus a single sum obtained with only the $r = s$ terms:

$$S_2 = \frac{1}{2}\sum_{r=1}^{n-1}\sum_{s=1}^{n-1} \exp[iq(r+s)\omega_0]\Theta(n-r-s) - \frac{1}{2}\sum_{r=1}^{[(n-1)/2]} \exp(2iqr\omega_0). \tag{A13}$$

By now transforming the dummy variable pair (r, s) to the pair $(r, R = r + s)$ in the double sum, we may easily reduce it to a single sum. The result of this manipulation is

$$S_2 = \frac{1}{2}\sum_{R=2}^{n-1} \exp(iqR\omega_0)(R-1) - \frac{1}{2}\sum_{r=1}^{[(n-1)/2]} \exp(2iqr\omega_0). \tag{A14}$$

The two single sums above can be evaluated easily. We note that, in the first sum, extending the lower limit of R to 1 leaves the sum unchanged. Then, when R is replaced by $R + 1$, the sum becomes a derivative of a simple geometric-series sum. The second single sum is of course itself a simple geometric-series sum, but we need not evaluate it explicitly, since its contribution to Eq. (A9) cancels exactly with an identical term in SS. We finally have the expression

$$S_2 = \frac{1}{2}\exp(inq\omega_0/2)\left[\frac{n-2}{2} - i\frac{\partial}{\partial(q\omega_0)}\right] \times \frac{\sin[q(n-1)\omega_0/2]}{\sin(q\omega_0/2)} - \frac{1}{2}\sum_{r=1}^{[(n-1)/2]} \exp(2iqr\omega_0). \tag{A15}$$

By using Eqs. (A5), (A8), (A12), and (A15) in Eq. (A9), we may rewrite $V[i_5(q)]$ as

$$V[i_5(q)] = \frac{\langle M\rangle}{2}n(n-1) + \langle M\rangle\left[n\,\mathrm{Re} - \frac{\partial}{\partial(q\omega_0)}\,\mathrm{Im}\right] \times \left\{\frac{-i\exp[i(n-1)q\omega_0/2]\sin[q(n-1)\omega_0/2]}{2\sin^2(q\omega_0/2)} + \frac{i(n-1)\exp(iq\omega_0/2)}{2\sin(q\omega_0/2)} + \frac{1}{2}\exp(inq\omega_0/2) \times \left[\frac{n-2}{2} - i\frac{\partial}{\partial(q\omega_0)}\right]\frac{\sin[q(n-1)\omega_0/2]}{\sin(q\omega_0/2)}\right\}. \tag{A16}$$

It is now only a matter of straightforward, although tedious, algebra to express this equation in the following explicit form:

$$V[i_5(q)] = \frac{\langle M\rangle}{2}\left\{\frac{n\sin^2\left(\frac{n-1}{2}x\right)}{\sin^2\frac{x}{2}} + \frac{n(n-2)}{2}\cos\left(\frac{nx}{2}\right) \times \frac{\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} + \frac{n}{2}\sin\left(\frac{nx}{2}\right)\frac{d}{dx}\frac{\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} + \frac{d}{dx}\frac{\cos\left(\frac{n-1}{2}x\right)\sin\left(\frac{n-1}{2}x\right)}{\sin^2\left(\frac{x}{2}\right)} - (n-1)\frac{d}{dx}\cot\left(\frac{x}{2}\right) - \frac{n-2}{2}\frac{d}{dx}\frac{\sin\left(\frac{nx}{2}\right)\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)} + \cos\left(\frac{nx}{2}\right)\frac{d^2}{dx^2}\frac{\sin\left(\frac{n-1}{2}x\right)}{\sin\left(\frac{x}{2}\right)}\right\}_{x=q\omega_0}. \tag{A17}$$

This was the equation to be derived.

We now consider the case of imaging, in which the zero-frequency phasor is included in the image. The map $i_6(q)$ is defined to be

$$i_6(q) = \mathrm{Re}\left[\sum_{r=1}^{n-1} z_r \exp(iqr\omega_0)\right] + \frac{1}{2} z_0$$

$$= i_5(q) + \frac{1}{2} z_0. \tag{A18}$$

The average map then is given by Eq. (3.13):

$$I_6(q) = I_5(q) + \frac{1}{2}\langle C\rangle. \tag{A19}$$

Similarly, the variance $V[i_6(q)]$ of this map differs from $V[i_5(q)]$ only by terms that contain the variance and covariances of z_0:

$$V[i_6(q)] = V[i_5(q)] + \frac{1}{4} V(z_0) + \sum_{r=1}^{n-1} \mathrm{cov}(z_0, \mathrm{Re}\, z_r)\cos(qr\omega_0)$$

$$= V[i_5(q)] + \frac{1}{4}\langle M\rangle n + \langle M\rangle \sum_{r=1}^{n-1} (n-r)\cos(qr\omega_0)$$

$$= V[i_5(q)] - \frac{3}{4}\langle M\rangle n + \langle M\rangle \left[n\,\mathrm{Re} - \frac{\partial}{\partial(q\omega_0)}\mathrm{Im}\right] \times \sum_{r=0}^{n-1} \exp(iqr\omega_0)$$

$$= V[i_5(q)] - \frac{3}{4}\langle M\rangle n + \langle M\rangle \left[n\cos\left(\frac{n-1}{2}x\right) \times \frac{\sin\left(\frac{nx}{2}\right)}{\sin\left(\frac{x}{2}\right)} - \frac{\mathrm{d}}{\mathrm{d}x}\sin\left(\frac{n-1}{2}x\right)\frac{\sin\left(\frac{nx}{2}\right)}{\sin\left(\frac{x}{2}\right)}\right]_{x=q\omega_0}, \tag{A20}$$

where $V[i_5(q)]$ is given by Eq. (3.8). This completes the derivation of Eq. (3.14).

ACKNOWLEDGMENTS

Shrinivas R. Kulkarni gratefully acknowledges financial support from the W. M. Keck Foundation. The work of Sudhakar Prasad was supported partially by the Sandia National Laboratories under a Sandia University Research Program contract. Shrinivas R. Kulkarni is a National Science Foundation Presidential Young Investigator and an Alfred P. Sloan Fellow.

Sudhakar Prasad is also with the Department of Physics and Astronomy, University of New Mexico. Shrinivas Kulkarni is also with the Owens Valley Radio Observatory, California Institute of Technology.

REFERENCES

1. See, e.g., D. Mozurkewich, D. J. Hutter, K. J. Johnston, R. S. Simon, M. Shao, M. M. Colavita, D. H. Staelin, B. E. Hines, J. L. Hershey, J. A. Hughes, and G. H. Kaplan, "Preliminary measurements of star positions with Mark III stellar interferometer," Astron. Astrophys. **193**, 1269–1277 (1988).
2. C. A. Haniff, C. D. Mackay, D. J. Titterington, D. Sivia, and P. J. Warner, "The first images from optical aperture synthesis," Nature **328**, 694–696 (1987).
3. T. Nakajima, S. R. Kulkarni, P. W. Gorham, A. M. Ghez, G. Neugebauer, B. J. Oke, T. A. Prince, and A. C. S. Readhead, "Diffraction-limited imaging II: optical aperture synthesis imaging of two binary stars," Astron. J. (to be published).
4. See, e.g., J. W. Goodman, *Statistical Optics* (Wiley, New York, 1985), Chap. 5.
5. See, e.g., A. R. Thompson, J. M. Moran, and G. W. Swenson, *Interferometry and Aperture Synthesis* (Wiley, New York, 1986).
6. J. F. Walkup and J. W. Goodman, "Limitations of fringe-parameter estimation at low light levels," J. Opt. Soc. Am. **63**, 399–407 (1973); see also Ref. 4, Chap. 9.
7. R. A. Perley, F. R. Schwab, and A. H. Bridle, *Synthesis Imaging* (National Radio Astronomy Observatory, Socorro, N. Mex., 1985).
8. S. R. Kulkarni, "Self-noise in interferometers: radio and infrared," submitted to Astron. J.
9. K. Shimoda, H. Takahashi, and C. H. Townes, "Fluctuations in amplification of quanta with application to maser amplifiers," J. Phys. Soc. Jpn. **12**, 686–700 (1957).
10. H. Steinberg, "The use of a laser amplifier in a laser communication system," Proc. IEEE **51**, 943 (1963).
11. M. Shao, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109 (personal communication, 1989). The National Aeronautics and Space Administration is working currently to improve the readout noise of CCD's with the aim of making them more attractive for space interferometry than they are now.