Optical processor for the very large array (VLA) radiotelescope: encoding-dependent errors

J. R. Fienup, L. E. Somers*
Environmental Research Institute of Michigan
Radar and Optics Division, P.O. Box 8618, Ann Arbor, Michigan 48107

Abstract

An optical Fourier transform processor converts visibility data at its input to a sky brightness function at its output. Scattered light and spurious terms resulting from the method of encoding the complex visibility function as a real-valued and nonnegative transmittance result in errors in the processor output. In this paper, we discuss the degree to which each of these undesired terms degrades the signal-to-noise ratio of the output for different encoding methods and classes of images. We suggest alternative methods of encoding the input that greatly reduce the errors. Two particularly powerful methods of minimizing noise are "complementary weighting" and "bias equalization". Two data dependent parameters are found to be most important: (1) the ratio of the mean-squared visibility magnitude to the square of its maximum, and (2) the ratio of the squared maximum of the brightness function to its mean squared value.

1. Introduction

The very large array (VLA) radiotelescope measures the complex visibility function which, when properly formatted and Fourier transformed, yields the angular radio brightness distribution of a portion of the sky. If the Fourier transform operation is performed in a coherent optical processor, then the need arises to form at the input of the processor an optical wavefront of complex amplitude proportional to the complex visibility function. This wavefront is formed by illuminating a photographic transparency (or other spatial light modulator) on which the visibility function has been encoded as a spatial distribution of transmittance. Because of the need to encode a complex-valued function as a real, nonnegative transmittance, spurious terms are introduced resulting in noise in the detected brightness distribution at the output of the processor. This coupled with the impulse response due to the VLA aperture results in an unusual form of noise. In addition, scattered light due to film grain noise and imperfections in the optical components degrades the image. In this paper the nature and magnitude of these errors are discussed, and methods for decreasing them are described.

In general, it was found that within the output of the processor there are regions of high and regions of low signal-to-noise ratio. Proper encoding maximizes the extent of the regions of high quality, maximizes the use of the high quality region, and minimizes the use of low quality regions. In the output plane the area of highest quality output appears in an annular region about the optical axis. Near the inner boundary of the annulus, the defects are dominated by additive noise from an undiffracted beam of light; and near the outer boundary the data is distorted by phase errors introduced by the processor optics and the imperfect modulation of the visibility data on the input transducer. Novel methods for optimizing the signal-to-noise ratio of the output that are discussed include the "complementary weighting" of the input, the choice of the off-track bias level and the surrounding aperture, and the choice of the phase of the image. It was also found that the signal-to-noise ratio of the output is directly effected by two data-dependent quantities: the ratio of the mean-squared visibility magnitude to the square of its maximum, and the ratio of the square of the maximum of the brightness distribution to its mean squared value. For the worst case situation - an extended object for a brightness distribution and processing over the full A-array - the use of the novel methods referred to above was found to be an absolute necessity.

The remainder of this section reviews the characteristics of the visibility data, describes an optical Fourier transform processor, and defines an error criterion which the processor must satisfy. In Section 2, various encoding methods are compared, with emphasis on those aspects of each method that are critically important for this application. In Section 3, the effects of a most troublesome undiffracted beam, the on-axis VLA impulse response, are studied closely, and ways to reduce its effects are described. Section 4 describes the effects of two other undiffracted terms, including light scattered by the optical processor. Details on an optical processor system design and its aberrations are given in two companion papers[1,2].

*L.E. Somers is with Lawrence Livermore Laboratory, Livermore, California 94550.

The VLA radiotelescope complex visibility function

The characteristics of the VLA radiotelescope and the measured complex visibility data pertinent to the encoding problem are described below. More detailed descriptions can be found in References 1 and 3.

The VLA radiotelescope has a "Y" shaped array of 27 individual microwave antennas.[1,3] Each arm of the array has nine antennas, each of diameter 25 meters, which can be moved radially along the arms to achieve different configurations. The largest array configuration, the "A" array, is 36,000 meters across. The receivers associated with each antenna are connected in pairs to correlators. There are 27(27-1)/2=351 such unique pairs, each forming an interferometer. Thus, the VLA telescope operates as an array of 351 interferometers. The output of each interferometer, after proper normalization, is a sample of the complex visibility function V(u,v). The coordinates u and v are proportional to components of the vector baseline of an interferometer (i.e., the vector spacing between the two antennas) projected on a plane normal to a reference direction in the center of the portion of the sky being observed.

As given by the Van Cittert-Zernike theorem[4], the complex visibility function is related to the observed radio sky brightness distribution by a Fourier transform, which in simplified form is

$$B(x,y) = \iint_{-\infty}^{\infty} V(u,v) e^{-i2\pi(ux+vy)} du dv \qquad (1a)$$

where constant factors and scaling is ignored.

At any given instant in time, the telescope is sampling V(u,v) at 351 points in u-v space. As the earth rotates, the projected vector spacing of each antenna pair changes, and each pair scans out an elliptical path (or track) through u-v space.[1,3] A sample is obtained from each of the 351 interferometers once every ten seconds, or roughly 4,000 complex samples from each interferometer in eleven hours, or 1.4×10^6 total complex samples per observation. The resulting V(u,v) is sampled only along the elliptical tracks, providing a partially filled aperture, an example of which is shown in Figure 1. The coverage is dense in the central portion of the u-v plane and sparse near the edge. This partially filled aperture results in a VLA system impulse response having sidelobes that extend throughout the image plane as shown in Figure 2. For points away from its mainlobe, the sidelobe level of the VLA impulse response is about 25 dB down from (1/300) its peak for the A array.

V(u,v) is Hermetian since B(x,y) is real valued. The VLA radiotelescope does not generally measure V(0,0), that is to say, the "DC" value of V(u,v) is not usually available for processing. As a result of this and the existence of negative sidelobes in the impulse response of the VLA aperture, the resulting brightness images are real and bipolar; in order to accomodate bipolar output data, the optical processor must include an interferometric read-out.

The A-array telescope configuration generates complex visibility functions with space-bandwidth product requirements of about 3,000 cycles in each dimension. The telescope data is accurate to about 3°. The performance requirement put on the optical processor was that the radio brightness output function be accurate to within 1%. In this paper, the 1% criterion is taken to mean that the brightness function output not differ from the direct numerical Fourier transform (not the FFT) of the complex visibility function by more than 1% of the maximum of the brightness function. This requirement is dealt with further in Reference 2.

Optical processor

The aspects of the optical Fourier transform processor design for VLA radiotelescope data that are pertinent to the input encoding are described in this section; more detailed discussions of a processor design and its aberrations are found in References 1 and 2. As seen from the discussion above, the optical Fourier transform processor must be linear, have low scattered light, accept a complex-valued two-dimensional input and deliver a bipolar two-dimensional output. It is important to emphasize that the Fourier transform relationship exists between the complex wavefronts in the front and back focal planes of a Fourier transform lens, and so the amplitude rather than the intensity of the wavefront in the output plane must be detected. Note that the 1% criterion refers to the amplitude of the wavefront in the output plane; therefore the intensity of additive noise must be less than 10^{-4} of the peak of the intensity of the desired wavefront in the output plane.

For simplicity, we consider the optical Fourier transform processor configuration shown in Figure 3. A spatial light modulator (photographic transparency) illuminated by a plane

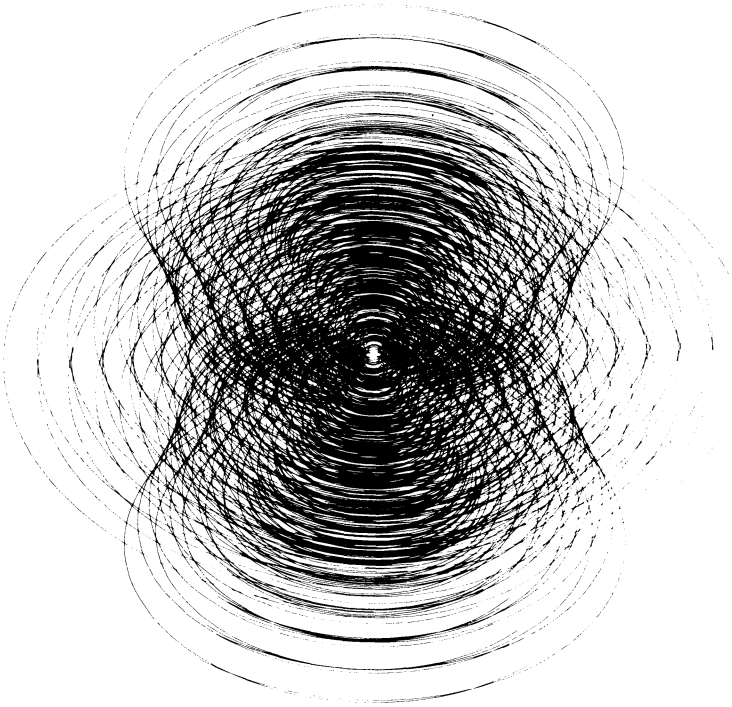

Figure 1. A simulation of the tracks in u-v space that define the VLA synthetic aperture.

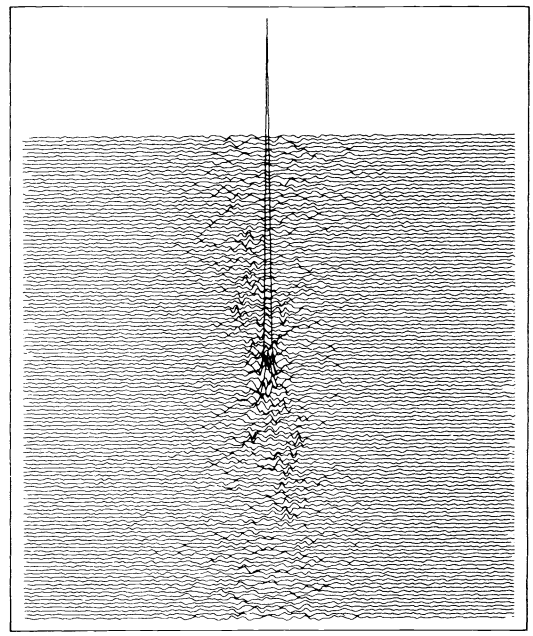

Figure 2. A simulation of the impulse response due to the VLA aperture.

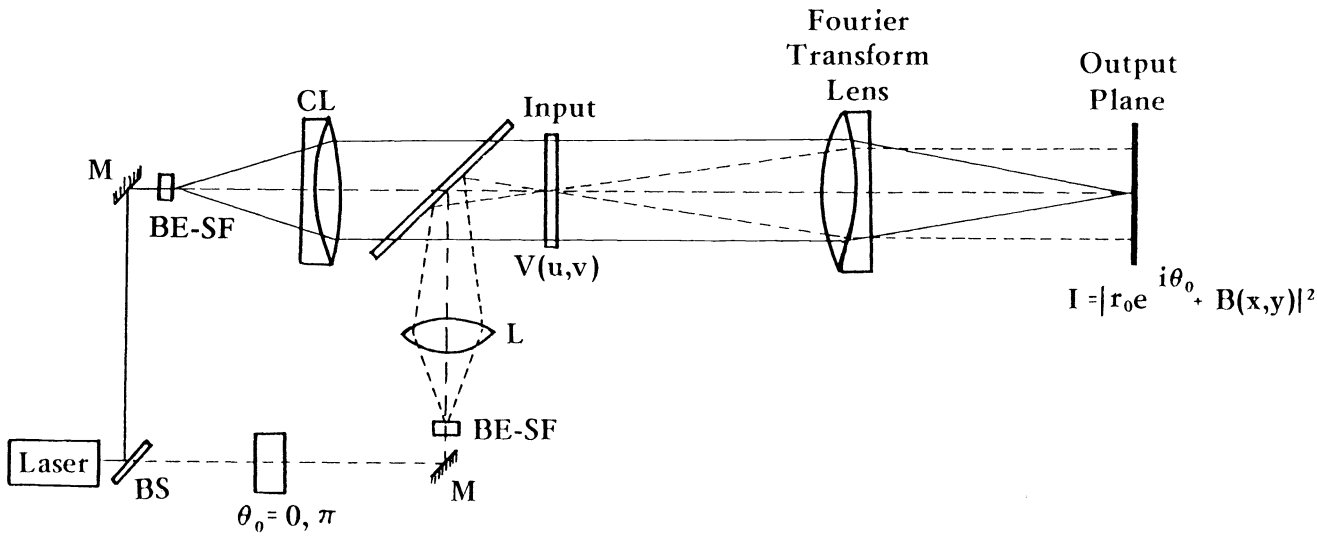

Figure 3. Optical processor for VLA radiotelescope visibility data.

wave of coherent light provides at the input plane a complex wavefront proportional to V(u,v). The Fourier transform lens causes a wavefront proportional to a spatially scaled version of B(x,y), the Fourier transform of V(u,v), to be formed in the output plane. Including the optical processor scale factors, the wavefront in the output plane is given by[5].

$$B(x,y) = \frac{1}{\lambda f} \iint_{-\infty}^{\infty} V(u,v) \exp[-j(2\pi/\lambda f)(ux + vy)]\, du\, dv \qquad (1b)$$

where λ is the wavelength of light used and f is the focal length of the Fourier transform lens. Eq. (1b) will be taken to be the Fourier transform relationship for the remainder of this paper, and u and x are scaled to have units of length in the input and output planes, respectively, of the optical processor. To detect the bipolar output B(x,y) with detectors sensitive only to intensity, it is necessary to add a coherent reference wave to the processor. For example, in Figure 3 an axial reference plane wave is provided at the output by bringing to a point focus a beam at the center of the input plane. The net intensity at the output is then

$$I_i(x,y) = \left| r_o e^{i\phi_o} + B(x,y) \right|^2 \qquad (2)$$

where r_o is the constant amplitude and ϕ_o is the constant phase of the reference wavefront. By detecting the output intensity twice, once with $\phi_o = 0$ and a second time with $\phi_o = \pi$, and taking the difference (digitally) we arrive at a quantity proportional to the real valued bipolar B(x,y):

$$\begin{aligned} I_+(x,y) + I_-(x,y) &= \left| r_o + B(x,y) \right|^2 - \left| - r_o + B(x,y) \right|^2 \\ &= 4\, r_o\, \text{Real}\left[B(x,y)\right] = 4\, r_o\, B(x,y) \end{aligned} \qquad (3)$$

In this detection scheme the imaginary component of any error would be suppressed since the only real component is detected.

An alternative detection scheme would be as follows. Set $\phi_o = 0$, and make r_o greater than the magnitude of the greatest negative value in the output plane, making $r_o + B(x,y)$ everywhere postive. Then detect $\left| r_o + B(x,y) \right|^2$, digitally take the square root, and subtract r_o to get the bipolar B(x,y). A disadvantage of this simpler detection scheme is that the imaginary components of error tems would not be suppressed as well as in the case of the double detection method described earlier.

2. Encoding methods

In this section a number of general observations on the encoding of V(u,v) are made, followed by a discussion of several different encoding methods that were considered. A wavefront of monochromatic light having a complex amplitude proportional to the complex visibility function is entered into the input of the Fourier transform optical processor by means of a spatial light modulator (SLM). A number of SLMs, such as Itek's PROM and Hughes' Liquid Crystal Light Value, are erasable and are capable of operating in real time[6]. However, for this application, it is preferable to use photographic film, which, although neither erasable nor real-time, is more easily controlled accurately and is capable of the high space-bandwidth product and signal-to-noise ratios required for this application. The spatial distribution of the effective transmittance of the film must be proportional to the complex valued visibility function, despite the fact that film is typically either pure amplitude (absorbing) or pure phase. The complex visibility function can be expressed as

$$V(u,v) = \left|V(u,v)\right| \exp\left[i\phi(u,v)\right] = R(u,v) + iQ(u,v) \qquad (4)$$

where $|V|$ and ϕ are the magnitude and phase, respectively, and R and Q are the real and imaginary components, respectively, of V(u,v). Complex-valued effective transmittances are achieved by the encoding methods of computer-generated holograms[7-9]. In effect, additional terms are added to the desired term in such a way that the sum is either pure amplitude (i.e., real and nonnegative) or pure phase, allowing the sum to be recorded on film. The extra terms are added in such a way that they can be spatially filtered from the desired term; that is, they do not overlap the desired term in the output of the optical processor. Alternatively, two or more wavefronts from separate film transparencies can be

added interferometrically in such a way as to produce a sum equal to the desired complex quantity.

Important considerations in choosing an encoding method are the diffraction efficiency η_m (the ratio of the light flux diffracted into the desired term to the light flux incident on the film), the relative strengths of spurious terms that overlap the desired term, other inherent errors, the type of optical recording device required (binary vs. continuous-tone), the spatial efficiency with which the recorder is utilized (the number of resolvable elements of the film recorder required to encode a single complex sampled value), the type of material required (amplitude or phase) and the material properties, and the practical difficulties in implementing an encoding method. As will be discussed in Section 3, the net diffraction efficiency includes a few factors, including η_m and other factors having to do with (a) the statistics of V(u,v), (b) the degree of filling of the partially-filled VLA aperture, and (c) the dynamic range of the film. η_m is the theoretical maximum diffraction efficiency of the encoding method, which would be obtained only for an input function of constant magnitude over a filled aperture.

For all the encoding methods considered, it is assumed that the writing beam of the film recorder can be accurately positioned to within a small fraction of the distance over which the phase of the visibility function, $\phi(u,v)$, typically changes by a cycle. For the full A-array and for points near the edge of the brightness map, this implies a writing beam positioning accuracy of one part in 30,000 in order to keep the phase errors to less than one-tenth wavelength. That distance would usually be considerably smaller than the width of the writing beam. The data is given on an irregular locus of points. It is assumed that the data is written onto the film either at the given sampled locations or else is reformatted to a very fine rectangular grid. If the recording beam were to address the SLM on a coarser rectangular grid of points, then phase errors would arise from the fact that the transmittance for a given grid point would be based on the value of the phase of a nearby point on a track rather than on the value of the phase at the given point. Interpolation of the phase to a grid point is not possible unless it lies on one of the ellipses or between two closely spaced ellipses. It is also assumed that the intensity of the writing beam is calibrated in such a way as to obtain the desired amplitude transmittance in the developed transparency.

There does exist a type of computer-generated hologram, the ROACH[10,11] that directly achieves complex-valued transmittance with η_m = 100%, has ideally efficient use of recorder resolution, reconstructs an image on the optical axis, and has no spurious terms or intrinsic errors. Unfortunately, high-quality recordings of ROACHs have not been demonstrated yet, since presently available two-layer (or multi-layer) materials (one layer each to control amplitude and phase, respectively) are not of sufficiently high quality[12], and accurate continuous-tone writing beam intensities are required to avoid phase errors. Should these technical difficulties be overcome, then the ROACH would ultimately be the best method of encoding the complex visibility function.

The simple carrier method encodes the complex visibility as the real, nonnegative transmittance[13]

$$H(u,v) = T_o + T_1 |V(u,v)| \cos [\omega u + \phi(u,v)]$$

$$= T_o + (1/2)T_1 V(u,v)\, e^{j\omega u} + (1/2)T_1 V^*(u,v) e^{-j\omega u} \qquad (5)$$

where T_o, T_1, and ω are constants. The corresponding Fourier transform output (neglecting the aperture) is

$$h(x,y) = \lambda f\, T_o\, \delta(x,y) + (1/2)T_1 B(x - x_o, y) + (1/2)T_1 B^*(-x - x_o, -y) \qquad (6)$$

where $x_o = \omega\lambda f/2\pi$. The values of the bias T_o and the modulation coefficient T_1 are chosen such that $0 \le H(u,v) \le 1$, and the carrier frequency ω is chosen large enough so that the desired term in the output, $(1/2)T_1 B(x-x_o,y)$, does not overlap the undiffracted term $\lambda f\, T_o \delta(x,y)$ or the conjugate term $(1/2)T_1 B^*(-x-x_o,-y)$. The image is reconstructed off the optical axis at a distance x_o. Although it suffers from low diffraction efficiency (η_m = 6.25%), the simple carrier encoding is attractive because it has (assuming a linear recording) no intrinsic errors and has no spurious terms except the two mentioned above, which in principle can be made to be nonoverlapping with the desired term. The effects of the undiffracted term will be discussed in detail in Section 3.

The simple carrier method can be considered as a modification to a hologram of V(u,v) recorded with a plane reference wave, except that it does not include the spurious term proportional to $|V(u,v)|^2$. Another modification that could be made to Eq. (5) is the substitution of the term $T_1|V(u,v)|$ for the bias term[14]; however, this is undesirable since

the extent of the Fourier transform of the spurious term $T_1|V(u,v)|$ is far greater than that of the bias term.

Still another possible modification of the simple carrier method would be to encode it as a pure phase function of the form

$$H(u,v) = \exp\{j\alpha|V(u,v)| \cos[\omega u + \phi(u,v)]\} \tag{7}$$

The primary advantage of this phase encoding is its higher diffraction efficiency (η_m = 33.9%). However, expanding the exponential above in a power series, one finds a large number of intrinsic spurious terms[15]. Although most of the terms (such as the term $-\alpha^2|V(u,v)|^2/4$) have Fourier transforms that can be made to not overlap the desired image term, overlap with the Fourier transform of the "intermodulation" term $(-j\alpha^3/16)V(u,v)|V(u,v)|^2 \exp(j\omega u)$ is unavoidable. This term results in an image-plane term $(-j\alpha^3/16)B(x-x_o,y) * [B(x,y) \bigstar B(x,y)]$, where $*$ denotes convolution and $\bigstar$ denotes autocorrelation. In order to keep this term below 1% of the desired term, it is necessary to keep α small and thereby keep the diffraction efficiency to well under the 6.25% efficiency of the amplitude version of the simple carrier method. Then for this application, one cannot take advantage of the phase encoding's chief advantage, its high diffraction efficiency. A similar conclusion can be reached for phase versions of the other encoding methods discussed below.

Because only binary transmittances [H(u,v) =0 or 1] are required, making it easy to fabricate, a very popular encoding method is the Lohmann binary detour-phase hologram[16-18]. It consists of an assembly of small rectangular apertures on an opaque background, where the area of an aperture is proportional to the local value of $|V(u,v)|$ and the translation of the aperture relative to a reference grid is proportional to $\phi(u,v)$. It is approximately a hard-clipped version of the simple carrier encoding method. Although it can be implemented in such a way as to eliminate film-grain noise[19], it unfortunately possesses inherent errors[17,18,20] that would prevent the 1% error criterion from being met for points away from the center of the brightness map. In addition, several resolvable elements of the writing beam are required to form one aperture to represent a single sampled value of V(u,v). The location of the samples of V(u,v) on the ellipses also presents a problem in using the Lohmann method. The proper position of the center of a given aperture is given by the position (u_p, v_p) determined by the equation[18,20]

$$[\omega u_p + \phi(u_p,v_p) = 2\pi L_p] \tag{8}$$

This corresponds to the locations for which $\cos[\omega u_p + \phi(u_p,v_p)] = 1$. Given four adjacent samples of $\phi(u,v)$ along a track, a cubic interpolating polynomial can be defined that can be used for $\phi(u_p,v_p)$ in the equation above. The resulting cubic equation can be solved for the position (u_p,v_p), which defines the center of the aperture. The area of the aperture is made proportional to $|V(u_p,v_p)|$ which is determined by a cubic interpolation of $|V(u,v)|$, using the four nearest samples of $|V(u,v)|$ on the track. A problem with this procedure is that where a track is roughly parallel to the fringes of $\cos[\omega u + \phi(u,v)]$, Eq. (8) above would not have a solution and those portions of the tracks would not contribute to the encoded transmittance. Either interpolation between tracks or simply using the nearest phase value and assuming the phase is locally constant would lead to unacceptably large phase errors for the less densely sampled regions of the u-v plane.

Other detour-phase encoding methods[21,22] are oriented toward samples on a rectangular grid and have other intrinsic errors, making them less desirable than the simple carrier method for this application. For example, the Lee method[21] can be shown to be a sampled half-wave-rectified version of the simple carrier method, but without the bias T_o. Since half-wave rectification results in a spectrum having spurious terms overlapping the desired term[23], the Lee method has noise terms not present in the simple carrier method.

Quite a different approach is embodied in the real-imaginary encoding method[24]. The wavefronts from two transparencies with transmittances

$$H_R(u,v) = T_o + T_1 R(u,v) \tag{9a}$$

and

$$H_Q(u,v) = T_o + T_1 Q(u,v) \tag{9b}$$

respectively, are combined interferometrically, as shown in Figure 4. The wavefront from $H_Q(u,v)$ is phase delayed relative to that of $H_R(u,v)$ so that the sum of the two wavefronts is

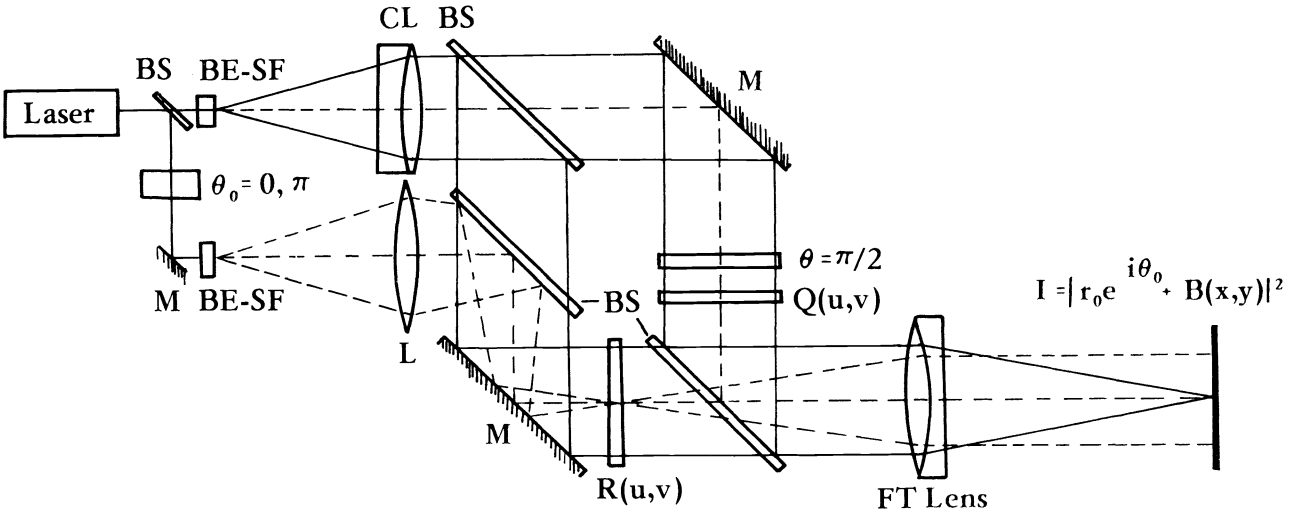

Figure 4. Optical processor for the case of the real-imaginary encoding method.

$$H_{total} = T_0 + T_1 R(u,v) + j[T_0 + T_1 Q(u,v)] = T_0(1 + j) + T_1 V(u,v) \qquad (10)$$

The resulting image $T_1 B(x,y)$ is reconstructed on axis; however, so is the undiffracted term $\lambda f(1 + j)T_0\delta(x,y)$, the sidelobes of which (when the aperture is considered) would wipe out a good portion of the center of the image. On the positive side, the real-imaginary method results in a single image term with no spurious terms other than the undiffracted term and has relatively high diffraction efficiency (η_m = 25% which includes a 50% loss at the final beamsplitter).

By decomposing the visibility function into three or more terms, the bias term can be eliminated. In general, N components taken along the N respective Nth roots of unity can be made into transparencies and combined interferometrically to produce the desired complex function. N = 3 or 4 are the only two cases of practical interest. For N = 4, for example, the four components R_+, Q_+, R_-, and Q_- are proportional to half-wave rectified versions of Real $[\exp(-jk\pi/4)V(u,v)]$, k = 0, 1, 2, and 3, respectively; and

$$V(u,v) = R_+(u,v) - R_-(u,v) + jQ_+(u,v) - jQ_-(u,v). \qquad (11)$$

The desired complex visibility function V(u,v) could be recovered exactly on-axis with no other terms present by the simultaneous illumination of all four transparencies and the interferometric combination of the four resulting wavefronts with the appropriate constant relative phase shifts between them. For this method η_m is only 6.25% (the average intensity transmittance of the four transparencies is 25% and only 25% of the light goes into the image from the wavefront-combining beamsplitters of the interferometer). The chief disadvantages of this method, which we refer to as the real+-imaginary+- method, are the requirement of a stable interferometer with very precisely controlled phase relationships between its various arms and the extra precision required of the intensity of the writing beam of the film recorder.

An alternative real+-imaginary+- method is to use the simpler processor configuration of Figure 3, process the four wavefronts serially, and add the results digitally. However, since each of the individual terms is a half-wave rectified version of a bipolar function, they each result in large on-axis delta-function-like terms[23]. Consequently, near the optical axis the brightness map will be arrived at by taking the difference of large numbers to yield a relatively small number. This is also true in the case of the interferometrically combined wavefronts discussed earlier. In that case, although the on-axis impulse response term is eliminated from the sum of all four component images, it is present in each component image. In both cases, very accurate transmittances of the four transparencies and a careful balancing of the relative intensities of the four wavefronts are required for points in the image near the optical axis.

Other variations of these encoding methods are possible and are too numerous to list here. One final example is a combination of the simple carrier method with the real-imaginary method. Wavefronts from two transparencies are combined interferometrically as in the real-imaginary method. The transmittance of one transparency is that of Eq. (5),

the simple carrier encoding, and the other is the same as Eq. (5) but replacing cos by sin. The relative phase of the two wavefronts are adjusted to result in the sum

$$T_o + T_1|V(u,v)|\cos[\omega u + \phi(u,v)] + jT_o + jT_1|V(u,v)|\sin[\omega u + \phi(u,v)]$$

$$= (1 + j)T_o + T_1 V(u,v)\exp[j\omega u] \qquad (12)$$

which results in $\lambda f(1 + j)T_o\delta(x,y) + T_1 B(x - x_o,y)$ in the output plane. This method eliminates the conjugate image present in the simple carrier method and yet produces an image off-axis away from the undiffracted term.

Irrespective of the encoding method used, it is possible to obtain a full-resolution image using only one half of the u-v plane. Using only half of the u-v plane is equivalent to multiplying V(u,v) by the unit step function (= 0 for u < 0; = 1 for u > 0). The result in the output plane is the convolution of the image with the Fourier transform of the unit step function:

$$B(x,y) * \left[\frac{\lambda f}{2}\delta(x)\delta(y) - \frac{j\lambda f}{2\pi x}\delta(y)\right] = \frac{1}{2}B(x,y) - \frac{j}{2\pi}B(x,y) * [\lambda f\delta(y)/x] \qquad (13)$$

Since the detection process allows the real part (1/2)B(x,y) to be separated out from the imaginary part, the image is recovered with full resolution although the aperture is only half the original width. This phenomenon can be exploited only when the image has constant phase; a partial explanation of this phenomenon is that since the addition of the second half-plane of a Hermetian function adds no new information that cannot be obtained from the first half-plane, then there should also be no additional information in the image (such as would be obtained if the resolution were improved by a factor of two). The desired image being obtainable with no loss in resolution, the reconstruction of only half the u-v plane is attractive since it reduces the space-bandwidth-product requirement on the optical recording device by a factor of two. However, the signal-to-noise ratio would decrease if only half the u-v plane were used; and since the signal-to-noise ratio is the most pressing demand on the system, it is advisable to use the entire u-v plane.

3. The on-axis VLA impulse response

In this section the most troublesome error term, an undiffracted term (the on-axis impulse response), is discussed in detail, and methods of combatting it are described. We begin by considering a more detailed expression for the input wavefront, then derive an expression for the amplitude of the maximum of the output wavefront (on which the 1% criterion is based), which is then used to compute the signal-to-noise ratio. The analysis given assumes the use of the simple carrier encoding method, but is readily applicable to other encoding methods as well.

A more complete expression for the amplitude transmittance of the input transparency must include the effects of the partially filled VLA aperture,

$$A_T(u,v) = \begin{cases} 1, & \text{in area of tracks} \\ 0, & \text{in non-track areas} \end{cases} \qquad (14)$$

(an example of which is shown in Figure 1). If the amplitude transmittance, T_b, in the non-track area is nonzero, then it and the filled aperture $A_o(u,v)$, defining the extent of the input plane, must be considered. The Fourier transform of the VLA aperture $A_T(u,v)$ is the VLA impulse response $a_T(x,y)$ (an example of which is shown in Figure 2), which is sometimes referred to as the synthetic beam or the "dirty beam"; and the Fourier transform of $A_o(u,v)$ is $a_o(x,y)$. The amplitude transmittance of the input transparency is

$$H(u,v) = A_o(u,v)\left(A_T(u,v)\{T_o + T_1|V(u,v)|\cos[\omega u + \phi(u,v)]\} + T_b[1 - A_T(u,v)]\right)$$

$$= A_o(u,v)\left(T_b + A_T(u,v)\{(T_o - T_b) + T_1|V(u,v)|\cos[\omega u + \phi(u,v)]\}\right) \qquad (15)$$

The resulting image is given by

$$h(x,y) = T_b\, a_o(x,y) + (T_o - T_b)\, a_o(x,y) * a_T(x,y)$$

$$+ (T_1/2)a_o(x,y) * a_T(x,y) * B(x - x_o,y) + \text{conjugate image} \qquad (16)$$

where $x_o = \omega\lambda f/2\pi$. Assuming that $A_o(u,v)\, A_T(u,v) = A_T(u,v)$, then $a_o(x,y) * a_T(x,y) = a_T(x,y)$. The image is convolved with the VLA impulse response $a_T(x,y)$, which is inherent in the sampled data and occurs in digital as well as optical processing. Depending on the choice of T_b, there is a tradeoff between the on-axis impulse response terms $T_b\, a_o(x,y)$ and $(T_o - T_b)a_T(x,y)$, of which more will be said later.

Signal-to-noise ratio

The amount of light going into the brightest point in the image is determined as follows. The maximum and minimum possible amplitude transmittance in the track areas of the transparency are $T_o + T_1 |V|_{max}$ and $T_o - T_1 |V|_{max}$, respectively, where $|V|_{max} = \mathrm{Max}[A_T(u,v)\, V(u,v)]$. At most, the difference of these two terms is unity, and they will generally be somewhat less than unity according to the dynamic range capability of the film. Let the difference of the maximum and minimum allowable film transmittance be $\sqrt{\eta_f} = 2T_1|V|_{max}$. Then $T_1 = \sqrt{\eta_f}/(2\,|V|_{max})$, and the desired image term in Eq. (16) can be written

$$(T_1/2)\, a_T(x,y) * B(x - x_o, y) = \sqrt{\eta_f \eta_m}\; a_T(x,y) * B(x - x_o, y)/|V|_{max} \tag{17}$$

where $\sqrt{\eta_m} = 1/4$ for the simple carrier method. Define

$$B_{max} = \mathrm{Max}\left[a_T(x,y) * B(x,y)\right] \tag{18}$$

$$\overline{B^2} = W_B^{-2} \iint \left|a_T(x,y) * B(x,y)\right|^2 dx\, dy \tag{19}$$

where W_B^2 is the area of the image,

$$\eta_c = \frac{B_{max}^2}{\overline{B^2}} \tag{20}$$

$$\overline{|V|^2} = W_T^{-2} \iint \left|A_T(u,v)\, V(u,v)\right|^2 du\, dv \tag{21}$$

where

$$W_T^2 = \iint \left|A_T(u,v)\right|^2 du\, dv = \iint A_T(u,v)\, du\, dv \tag{22}$$

is the area of the tracks, and

$$\eta_V = \frac{\overline{|V|^2}}{|V|^2_{max}} \tag{23}$$

By Parseval's theorem (conservation of flux for the optical system),

$$\overline{B^2} = \overline{|V|^2}\,(W_T^2/W_B^2) \tag{24}$$

Combining this with Eqs. (18)-(20), we have

$$B_{max}^2 = \eta_c \overline{B^2} = \eta_c \overline{|V|^2}\; W_T^2/W_B^2 \tag{25}$$

Therefore, using Eqs. (23) and (25) the peak of the amplitude of the image wavefront from Eq. (17) is

$$S_P = \sqrt{\eta_f \eta_m}\, B_{max}/|V|_{max} = \sqrt{\eta_f \eta_m \eta_c \eta_V}\; W_T/W_B \tag{26}$$

The amount of light going into the sidelobes of the on-axis VLA impulse response (the noise) is determined as follows. We start by computing the peak of the on-axis VLA impulse response which is, using Eq. (22),

$$a_T(0,0) = \frac{1}{\lambda f} \iint_{-\infty}^{\infty} A_T(u,v)\, du\, dv = W_T^2/\lambda f \tag{27}$$

If the width of the VLA aperture is W_o, then the fractional track area is $\eta_T = W_T^2/W_o^2$. A filled square aperture of width W_o would result in an image-plane impulse response of width (peak to first null) $\lambda f/W_o$. The width of $a_T(x,y)$ exceeds $\lambda f/W_o$ by an amount that depends on the array configuration. If we let the ratio of the width of $a_T(x,y)$ to $\lambda f/W_o$ be d_T, then there are

$$N_T = W_B W_o/\lambda f d_T \tag{28}$$

resolvable elements over the image in each dimension. For the A-array, $d_T \simeq 1.8$ and $T \simeq 0.16$. Then Eq. (27) can be expressed as

$$a_T(0,0) = \frac{W_T^2}{\lambda f} = \frac{W_T W_o \sqrt{\eta_T}}{\lambda f} = \frac{W_T N_T d_T \sqrt{\eta_T}}{W_B} \tag{29}$$

Therefore, the amplitude ratio of the peak of the signal to the peak of the on-axis VLA impulse response is

$$\frac{S_P}{(T_o - T_b)a_T(0,0)} = \frac{\sqrt{\eta_f \eta_m \eta_c \eta_V/\eta_T}}{(T_o - T_b)N_T d_T} \tag{30}$$

The sidelobes of $a_T(x,y)$, as shown in Figure 2, die down quickly to about the 3×10^{-3} level, but then stay at that level even for large distances from the peak. That is, for a given point in the image the noise term $a_T(x,y) \lesssim a_T(0,0)/300$, and the signal-to-noise ratio in amplitude is

$$\frac{\text{signal}}{\text{noise}} = \frac{S_P}{(T_o - T_b)a_T(x,y)} \sqrt{\frac{300\, \eta_f \eta_m \eta_c \eta_V/\eta_T}{(T_o - T_b)N_T d_T}} \tag{31}$$

which must be greater than 100 in order to satisfy the 1% criterion.

Table 1 shows values of Eq. (31) and its most important factors for the A-array for four different cases: (a) a single point star; (b) 100 weighted stars (having brightness values

TABLE 1.

Signal-to-noise (amplitude) ratios for various astronomical objects for the noise term due to the sidelobes on the on-axis VLA impulse response for the A-array.

Here it is assumed that $\eta_f = 0.5$, $\eta_m = 0.0625$ (the simple carrier method), $T_o = 0.5$, $T_b = 0$, $\eta_T = 0.16$ and $d_T = 1.8$. The final row shows the degree to which the on-axis impulse response term must be reduced in order to just satisfy the 1% criterion.

	Single Star	100 Weighted Stars	100 Equal Stars	Cas-A
$\eta_c N_T^2$	1.0	0.355	10^{-2}	1.44×10^{-4}
η_V	1.0	10^{-1} to 10^{-2}	10^{-1} to 10^{-2}	4.09×10^{3}
$\frac{S_P}{0.5\, A_T(0,0)}$	0.49	$(0.93 \text{ to } 0.29)\times10^{-1}$	$(1.6 \text{ to } 0.2)\times10^{-2}$	0.38×10^{-3}
$\frac{\text{signal}}{\text{noise}} = \frac{S_P}{0.5\, a_T(x,y)}$	147	28 to 8.8	4.7 to 1.5	0.113
reduction of $a_T(x,y)$ required for 1% criterion	(none)	3.2 to 11x	21 to 68x	880x

$10^{-0.1(k-1)}$ for k = 1, ..., 12 and $10^{-1-0.01K}$ for k = 13, ..., 100); (c) 100 equal stars; and (d) Cas-A, an extended source. As seen from Table 1, the 1% criterion is satisfied only for the easiest case (a single star); for the worst-case Cas-A, the image is buried far below the sidelobes of the on-axis VLA impulse response.

The factor η_c/N_T^2 is the fraction of the total light flux going into the brightest image point (on which the signal-to-noise ratio is based). Since only a fixed amount of light flux is spread over the image, the more it is concentrated into a small number of points, the greater will be the signal-to-noise ratio at those points. $\eta_V = \overline{|V|^2}/|V|^2_{max}$ enters since the entire visibility function must be scaled down in order to keep its maximum value within the dynamic range of the film. Brightness distributions of small extent (and large η_c) tend to have more nearly constant visibility magnitudes (large η_V). On the other hand, broad extended distributions (having small η_c) tend to have sharply peaked visibility functions (small η_V). That is, the two critical parameters η_c and η_V tend to get small together as the brightness distribution becomes more extended and less point-like. Thus the signal-to-noise ratio, which is proportional to $\sqrt{\eta_V \eta_c}/N_T$, is strongly object-dependent.

Reduction of the on-axis VLA impulse response

Several possible methods of greatly reducing or eliminating the on-axis VLA impulse response $a_T(x,y)$ are as follows: (1) subtract it, since its form and position are known; (2) make the brightness distribution and the reference beam imaginary; (3) transform it into a well-behaved impulse response by setting the bias T_b in the non-track areas equal to the bias T_o in the track areas; (4) boost the signal by increasing η_V by clipping $|V|_{max}$; (5) complementary weighting of the visibility data. These methods of increasing the signal-to-noise ratio, some of which can be combined together, are discussed below.

1. Subtract it. Since the exact form and position of $a_T(x,y)$ will be known, it can be digitally subtracted from the brightness data after detection. This is exactly analogous to subtracting the sidelobes of an extremely bright star that is just outside of the field of view. For the 100 equal stars case, the sidelobes of $a_T(x,y)$ would be comparable to the star brightness, and the subtraction process would cause a moderate loss in accuracy. However, for Cas-A, the sidelobes of $a_T(x,y)$ would be an order of magnitude brighter than the brightest stars, and the subtraction process would cause a great loss in accuracy, since the difference of two large numbers would be taken in order to arrive at a relatively small number.

2. Make the brightness distribution imaginary. If in Eq. (15) cos is replaced by sin, then the resulting image term will be pure imaginary instead of pure real. However, $a_T(x,y)$ remains a pure real function. Then the detection process would consist of the subtraction of the intensities $I_+ - I_-$ which are obtained by mixing with the brightness distribution reference waves that are of relative phase $\pi/2$ and $-\pi/2$, respectively. In this case only the imaginary component of the output wavefront is detected, and the real-valued $a_T(x,y)$ is suppressed. In the practical execution of this method, the phase of the reference beam must be accurately controlled, and the factor by which $a_T(x,y)$ is suppressed is approximately equal to the error of the phase of the reference beam in radians. Therefore, to achieve by this method alone the 880 factor necessary for Cas-A, phase control accurate to 1/880 radians = 1.8×10^{-4} wavelengths would be required, which would be exceedingly difficult to achieve. In addition, as with method (1), the high sidelobe levels would present a problem in the detection process.

3. Equalize the bias levels. By Eq. (16), setting the off-track bias level T_b equal to the on-track bias T_o would eliminate the term $(T_o - T_b)a_T(x,y)$. If the visibility data were to be formatted onto a fine rectangular grid for raster-scan recording, then the non-track areas could be exposed along with the track areas, and having $T_b = T_o$ would be straightforward. On the other hand, if the visibility data is to be recorded one track at a time using a random-access film writing beam, then the off-track areas would have to be exposed in a second step. This might be accomplished, for example, by a uniform exposure through a mask blocking off the track areas. If this is done, then accurate sensitometry and mask registration would become problems. Note that the problem is much the same for the pure-phase version of any encoding method, since the bias phase of the off-track areas would have to closely match the bias phase of the on-track areas.

4. Clip $|V|_{max}$ to boost η_V. Since the signal-to-noise ratio is proportional to η_V, which is inversely proportional to $|V|^2_{max}$, reducing $|V|_{max}$ increases the signal-to-noise ratio (if η_f is kept constant). Reducing $|V|_{max}$ would be accomplished by clipping $|V(u,v)|$, which would destroy some low spatial-frequency information since $|V(u,v)|$ is maximum for low spatial frequencies. In practice some clipping occurs naturally since V(0,0) is not measured by the interferometers.

5. Complementary weighting. The most interesting method of increasing the signal-to-noise ratio is by complementary weighting of the visibility data. We replace Eq. (15) by the film transmittance

$$H(u,v) = A_o(u,v)\left(A_T(u,v)\left\{T_o+T_1\frac{|V(u,v)|}{A_1(u,v)}\cos[\omega u+\Phi(u,v)]\right\} + T_b[1 - A_T(u,v)]\right) \quad (32)$$

Then the input plane is illuminated by an apodized wavefront of the form $A_1(u,v)$, and the resulting wavefront is the product

$$A_1(u,v)H(u,v) = A_1(u,v)A_o(u,v)[T_b+A_T(u,v)(T_o-T_b)]$$
$$+ A_o(u,v)A_T(u,v)T_1|V(u,v)|\cos[\omega u+\phi(u,v)] \quad (33)$$

where $A_1(u,v)$ is the complementary weighting function. Now the undiffracted terms are multiplied by the weighting function $A_1(u,v)$, which results in a reduction of the sidelobes of those terms. The image-forming term, on the other hand, is precompensated for $A_1(u,v)$ and is therefore unaffected by it.

In order to not exceed the available dynamic range of the film, $A_1(u,v)$ must be limited to the range

$$\frac{|V(u,v)|}{|V|_{max}} \le A_1(u,v) \le 1 \text{ for all } (u,v) \quad (34)$$

Since the visibility magnitude $|V(u,v)|$ is typically very much smaller at high spatial frequencies than at low frequencies, $A_1(u,v)$ can be an effective apodization for the on-axis terms. Consequently, the sidelobes from the on-axis terms are greatly reduced.

Combination of methods. If the bias levels are equalized in a one-step raster scan recording, then the on-axis VLA impulse response term $a_T(x,y)$ can be made to vanish, and no other method would be required. Any one of the other methods by itself would be inadequate to reduce $a_T(x,y)$ by the factor of 880 for the worst-case situation. However, most of the methods can be combined, with their effects being cumulative, with the following exceptions. Little is gained by using both clipping of $|V|_{max}$ and complementary weighting simultaneously. Subtraction of $a_T(x,y)$ cannot be accomplished while equalizing the bias levels or while making the brightness and reference beams imaginary. When equalizing the bias levels, the sensitivity to mask misregistration is much greater when the brightness is imaginary than when it is real.

Factors of improvement that can be expected from the methods are as follows (the factors are data-dependent and are only approximate). (1) The analysis needed to predict the performance of the method of subtracting the known sidelobes of $a_T(x,y)$ has not been performed. It might reduce the error by as much as a factor of 10 or 100. (2) Assuming the phase of the reference beam can be controlled to within 10^{-3} wavelengths, $a_T(x,y)$ could be reduced by a factor of $1/(2\pi\times10^{-3}) \simeq 160$ by making the brightness imaginary. (3) For the two-step bias equalization method, assuming an ability to register the mask to within 1/25 of a track width and match the on-track and off-track bias levels to within an optical density of 0.01, it can be shown that $a_T(x,y)$ could be reduced by a factor of 160 for a real-valued brightness image and by a factor of 40 for an imaginary brightness. (4) Clipping $|V|_{max}$ could increase the signal-to-noise ratio by a factor of 2.5 to 25 for Cas-A, depending on the clipping level. (5) The reduction of the sidelobes of $a_T(x,y)$ by complementary weighting is limited by the smallest value of a smooth $A_1(u,v)$ allowed by Eq. (34). For Cas-A complementary weighting would allow an improvement by a factor of about 25.

The factor of 880 reduction of $a_T(x,y)$ for the worst case situation could be achieved by a one-step bias equalization, or by combining complementary weighting with a two-step bias equalization, or by combining complementary weighting with making the brightness imaginary. For the less demanding cases (100 weighted stars), the 1% criterion can be met by complementary weighting alone, or bias equaliation alone, or making the brightness imaginary alone. It should be cautioned that the improvement quoted in making the brightness imaginary depends on a very accurate detection process, whereas complementary weighting and bias equalization do not depend on the detection process.

4. Other noise terms

In this section two other on-axis noise terms are described: the term $T_b a_o(x,y)$ in Eq. (16) and noise due to light scattering in the optical system; and the conjugate image is discussed. Finally, the weighting of the output plane is described.

The filled aperture impulse response

When bias equalization is used to combat the on-axis VLA impulse response $(T_o-T_b)a_T(x,y)$, then $T_b \neq 0$, and the filled aperture impulse response term $T_b a_o(x,y)$ appears

in place of the VLA impulse response term in Eq. (16). Table 2 summarizes the important characteristics of the two most familiar filled aperture shapes, square and circular.

Since

$$a_o(0,0) = \frac{W_o^2}{\lambda f} = \frac{W_T^2}{\lambda f} \cdot \left(\frac{W_o^2}{W_T^2}\right) = a_T(0,0)/\eta_T \tag{35}$$

we have

$$\frac{S_P}{0.5\ a_o(0,0)} = \frac{S_P\ \eta_T}{0.5\ a_T(0,0)} \tag{36}$$

where $\eta_T \simeq 0.16$ for the A-array less the outermost track. Assuming $T_b = 0.5$, $S_P/[0.5\ a_o(0,0)]$ is η_T times the values in the third row of Table 1, or values ranging from 0.08 for the single star case to 6×10^{-5} for Cas-A. Therefore, for Cas-A the image should be located beyond the point where the sidelobes of $a_o(x,y)$ are a factor of $100/(6 \times 10^{-5}) = 1.7 \times 10^6$ below its peak. For the circular aperture, this distance would be $r = (x^2 + y^2)^{1/2} = (1.7 \times 10^6/\pi^{1/4})^{2/3}\ (\lambda f/W_o) = 1.2 \times 10^4\ (\lambda f/W_o)$, which is unacceptably large. For the square aperture, most of the sidelobe energy is concentrated along the x and y axes, where the noise is quite large. However, in a direction half way between the two axes (at 45 degrees) the sidelobes die off very fast. Letting $y = x$, the envelope of the square aperture sidelobes goes below the 6×10^{-5} level at a distance of only $x = y = (1.7 \times 10^6/\pi^2)^{1/2}\ (\lambda f/W_o) = 410\ (\lambda f/W_o)$. This distance is only a small fraction of the width of the image, which is about $3000\ (\lambda f/W_o)$. The image must be moved away from the optical axis by this additional distance, which is accomplished by increasing the carrier frequency ω, which is done at the expense of a minor increase in the space-bandwidth requirement on the optical recorder.

Three possible ways of orienting the sinc x sinc y sidelobes and the brightness map with respect to one another are as follows, as illustrated in Figure 5.

a. Use a carrier ωu and orient the square aperture $A_o(u,v)$ at a 45 degree angle with respect to the u-v axes;

b. Use a carrier $\omega(u+v)$ and orient the square aperture along the u-v axes;

c. Use a carrier ωu, and have the orientation of the square aperture track the sweep of a one-dimensional detector array.

In the above, complementary weighting, which would greatly reduce the sidelobes of $a_o(x,y)$, was not taken into account. Even without complementary weighting, however, the sidelobes of $a_o(x,y)$ would not limit the performance of the system as long as a square aperture is used.

TABLE 2
Filled Square and Circular Apertures

	Square (area W_o^2)	Circular (area W_o^2)
$A_o(u,v)$	$\text{rect}(u/W_o)\text{rect}(V/W_o)$	$\text{rect}[(u^2 + v^2)^{1/2}/(2W_o/\sqrt{\pi})]$
$a_o(x,y)$	$\left(\frac{W_o^2}{\lambda f}\right) \text{sinc}\left(\frac{W_o x}{\lambda f}\right) \text{sinc}\left(\frac{W_o y}{\lambda f}\right)$	$(W_o/\sqrt{\pi}\, r) J_1(2\pi r W_o/\sqrt{\pi}\, \lambda f)$
$a_o(0,0)$	$W_o^2/\lambda f$	$W_o^2/\lambda f$
envelope large x,y	$\left(\frac{W_o^2}{\lambda f}\right) \cdot \frac{1}{\pi^2} \left(\frac{\lambda^2 f^2}{W_o^2 xy}\right)$	$\left(\frac{W_o^2}{\lambda f}\right) \frac{1}{\pi^{1/4}} \left(\frac{\lambda f}{W_o r}\right)^{3/2}$

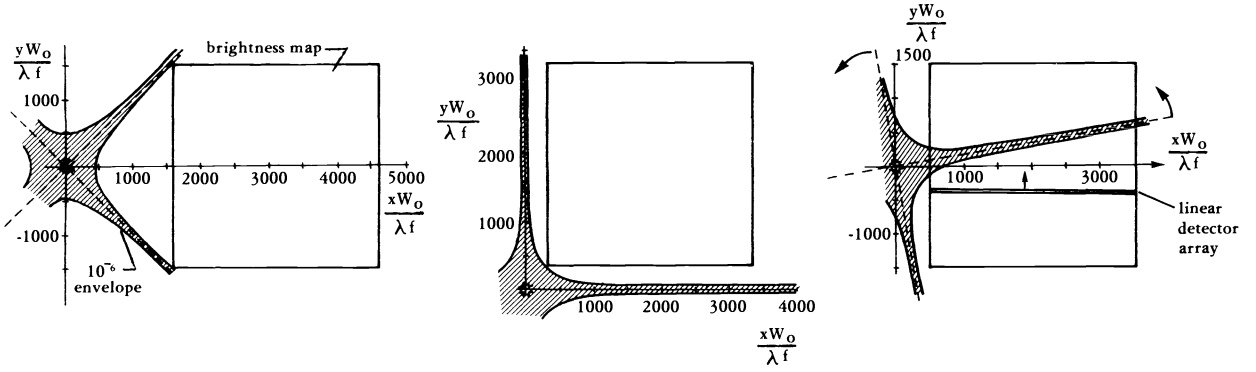

Figure 5. Orientation of the brightness map in order to avoid the sidelobes from a filled square aperture (see text).

Film and optical system noise

Scattered light arises from a number of different sources in the optical system including lens and beamsplitter surfaces and bubbles within the glass, as well as the input transparency. However, experience has shown that if a serious effort is made to reduce scattered light in the optical system, then the primary remaining source of scattered light will be film grain noise in the input transparency. Biederman[25], found that the scattered flux spectrum for a number of holographic materials is given by the formula

$$\frac{\Phi(\nu)}{\Phi_o} = a \cdot \nu^{-b} \tag{37}$$

where Φ_o is the transmitted flux, equal to the incident flux times T_b^2 and $\nu = r/\lambda f$ in units of mm^{-1}. For Kodak 649F plates, $a = 2.6 \times 10^{-4}$, $b = 2.26$, and the equation above is valid only for spatial frequencies in the range $6 \leq \nu \leq 80$ cycles/mm. For lower spatial frequencies, the noise is less than that predicted by this equation. The scattered flux $\Phi(\nu)$ is normalized for an area equivalent to $(1 \text{ cycle/mm})^2$. In terms of the noise amplitude $A_n(\nu)$, we have

$$\frac{A_n(\nu)}{T_b a_o(0,0)} = \left(\frac{1 \text{ mm}}{W_o}\right) \sqrt{a}\, \nu^{-b/2} \tag{38}$$

assuming the worst case for scattered light, namely, when bias equalization is used. Assuming the use of 649-F plates and $W_o = 50$ mm, this becomes

$$\frac{A_n(\nu)}{T_b a_o(0,0)} = 2.7 \times 10^{-2} \, (r/\Delta x)^{-1.13} \tag{39}$$

where $\Delta x = \lambda f/W_o$. Again for the worst case, Cas-A, we have, as in the previous section, $S_P/[0.5\, a_o(0,0)] = 6 \times 10^{-5}$. In this case the equation above would have to equal 6×10^{-7} in order to satisfy the 1% criterion. To achieve this would require much too large a carrier frequency. If by complementary weighting the noise were reduced by a factor of 25, then the 1% criterion would be satisfied outside a radius of $r/\Delta x = (2.7 \times 10^{-2} \times 0.04/6 \times 10^{-7})^{1/1.13} = 760$, compared with a total image width of about 3000. This implies a 50% increase of the carrier frequency and a 25% increase of the space-bandwidth product of the optical recorder compared with the amount necessary to just separate the image from the optical axis. It should be emphasized that this is for the worst case: Cas-A imaged with the A-array and satisfying the 1% criterion. For most cases of interest film grain noise and scattered light in the optical system would not mandate a substantial increase in the carrier frequency beyond that required to separate the image from the optical axis. Scattered light, being random and complex-valued in amplitude, cannot be subtracted from the image, nor does making the image imaginary-valued reduce its effects.

The conjugate image

If the image $B(x,y)$ were strictly equal to zero outside a given width W_B, then separation of the image term $B(x-x_o,y)$ from the optical axis would also ensure separation from the conjugate image term $B^*(-x-x_o,-y)$. For the VLA radiotelescope, the extent of the image

is determined by the far-field pattern of the individual antennas, which may have extended sidelobes, depending on the weighting of the antenna aperture. In addition, there are high frequency random noise terms in the visibility data, which result in random noise terms in the computed brightness distribution that go beyond the width of the image. Those noise terms in the conjugate image would extend into the desired image, resulting in up to twice the noise level that would occur in digital processing of the data.

A solution to this problem is to replace the $\cos[\omega u+\phi(u,v)]$ factor in Eqs. (15) and (32) by $\cos[\omega u+\phi(u,v) + \pi/4]$, in which case the image terms become

$$(T_1/2)a_T(x,y) \ast [e^{j\pi/4}B(x - x_o, y) + e^{-j\pi/4} B^*(-x - x_o, -y)] \qquad (40)$$

The detection process would then employ reference beams of phase $\pi/4$ and $5\pi/4$, respectively, which would cause the conjugate image term and the noise associated with it to be suppressed. This would, of course, take away the option of suppressing the on-axis terms (which would still be real valued) by making the brightness imaginary.

Image weighting

Three image-plane weighting factors must be compensated in order to arrive at an accurate measure of $B(x,y)$. One is the weighting of $B(x,y)$ due to the effective far-field pattern of the individual antennas, which is maximum at the center of the image and, for the case of the full A-array, dies off to one-half its maximum value for $[(x-x_o)^2+y^2]^{1/2} = 1500(\lambda f/W_o)$, i.e., at what has been considered the edge of the image. The second and third weighting factors are those due to the optical recorder and input film. $V(u,v)$ is effectively convolved with the distribution of the writing beam spot, causing the output plane to be multiplied by the Fourier transform of the writing beam spot. Note that the writing spot should not exceed the diameter of the autocorrelation of the antenna aperture scaled to the input plane of the optical processor; e.g., for a 50 mm input plane, the writing spot should be smaller than (2 x 25 m/72km)50mm = 35 μm. Further weighting is caused by the modulation transfer function of the input film. Both of these latter weighting functions are maximum on the optical axis and are smallest for the corners of the image farthest from the optical axis. All the weighting functions can be determined, allowing for compensation after $B(x,y)$ has been detected.

Summary and conclusions

The encoding method most likely to succeed is the simple carrier method. Of the major contending methods, the real-imaginary method fails near the center of its image, and so does the real+-imaginary+- method, but to a lesser degree. These two methods, requiring an additional interferometric combination of wavefronts, are also more demanding on the accuracy of the recording devices and of the optical processor. The Lohmann method tends to fail near the edges of the image field due to approximations inherent to that encoding method. The ideal encoding method, the ROACH, which avoids nearly all of the error terms described in this paper, is not practical since a suitable recording material of sufficiently high quality is not available.

The signal-to-noise ratio of the output of the optical processor was found to be strongly dependent on the characteristics of the visibility function and of the brightness distribution. The signal-to-noise ratio of the extended source Cas-A would be three orders of magnitude lower than the signal-to-noise ratio for a single point-like star.

The sidelobes associated with the on-axis VLA impulse response can be reduced sufficiently to meet the 1% criterion even for the worst case of an extended source such as Cas-A, but only if special techniques are employed. For maps of extended sources, complementary weighting (or clipping of $|V|_{max}$) is an absolute necessity, and a combination of complementary weighting with another of the techniques such as bias equalization may be required. Complementary weighting is particularly attractive because it reduces the effects of film grain noise as well as the sidelobes of the on-axis impulse response. Film grain noise is a problem for maps of extended sources and requires that greater demands be made on the film recorder for the worst case. The phase of the image can be chosen to be either $\pi/4$ in order to suppress the conjugate image term or $\pi/2$ in order to suppress the on-axis impulse response terms.

The 1% accuracy criterion appears to be achievable even for the worst case (Cas-A, full A-array, 3000 picture elements), although it is very demanding on the optical processor.

Acknowledgment

This work[26] was sponsored by the National Radio Astronomy Observatory, Charlottesville, Va., operated by Associated Universities Incorporated for the National Science Foundation.

References

1. I. Cindrich, C.C. Aleksoff, J.R. Fienup, and L.E. Somers, "Optical Processor for the VLA Radiotelescope: System Concept", Proc. SPIE 231-06 (April 1980).
2. C.C. Aleksoff and L.E. Somers, "Optical Processor for the VLA Radiotelescope: Aberration Analysis", Proc. SPIE 231-07 (April 1980).
3. E.B. Fomalont, "Earth-Rotation Aperture Synthesis", Proc. IEEE 61, 1211-18 (1973).
4. M. Born and E. Wolf, Principles of Optics, (Pergamon Press, Oxford, 1975).
5. J.W. Goodman, Introduction to Fourier Optics, (McGraw Hill, San Francisco, 1968).
6. Special issue on spatial light modulators, Opt. Eng. 17, 307-84, July/August 1978.
7. T.S. Huang, "Digital Holography", Proc. IEEE 59, 1335-46 (1971).
8. R.J. Collier, C.B. Burkhardt, and L.H. Lin, Optical Holography (Academic Press, New York, 1971).
9. W.-H. Lee, "Computer-Generated Holograms: Techniques and Applications", in E. Wolf, ed., Progress in Optics Vol. 16 (North-Holland, 1978).
10. D.C. Chu, J.R. Fienup, and J.W. Goodman, "Multiemulsion On-Axis Computer Generated Hologram", Appl. Opt. 12, 1386-88 (1973).
11. D.C. Chu and J.R. Fienup, "Recent Approaches to Computer-Generated Holograms", Opt. Eng. 13, 189-95 (1974).
12. J.R. Fienup, "Kodachrome as a Holographic Material", J. Opt. Soc. Am. 65, 1220(1975) (Abstract); J.R. Fienup, "Improved Synthesis and Computational Methods for Computer-Generated Holograms", Ph.D. Thesis, Stanford University, May 1975, University Microfilms No. 75-255-23, Chapter 3.
13. J.J. Burch, "A Computer Algorithm for the Synthesis of Spatial Frequency Filters", Proc. IEEE 55, 599-601 (1967).
14. T.S. Huang and B. Prasada, "Considerations on the Generation and Processing of Holograms by Digital Computers", MIT/RLE Quarterly Progress Report 81, April 15, 1966, pp. 199-205.
15. J.C. Urbach and R.W. Meier, "Properties and Limitations of Hologram Recording Materials", Appl. Opt. 8, 2269-81 (1969).
16. B.R. Brown and A.W. Lohmann, "Complex Spatial Filtering with Binary Masks", Appl. Opt. 5, 967-69 (1966).
17. A.W. Lohmann and D.P. Paris, "Binary Fraunhofer Holograms, Generated by Computer", Appl. Opt. 6, 1739-48 (1967).
18. B.R. Brown and A.W. Lohmann, "Computer-Generated Binary Holograms", IBM J. Res. Develop. 13, 160-68 (1969).
19. R.E. Williams, K. vonBieren, and M. Morales, "Wide Aperture Optical Correlator - Spectrum Analyzer: Evaluating of a New Design", Appl. Opt. 14, 2947 (1975).
20. W.-H. Lee, "Binary Synthetic Holograms", Appl. Opt. 13, 1677-82 (1974).
21. W.-H. Lee, "Sampled Fourier Transform Hologram Generated by Computer", Appl. Opt. 9, 639-43 (1970); C.B. Burckhardt, "A Simplification of Lee's Method of Generating Holograms by Computer", Appl. Opt. 9, 1949 (1970).
22. R.E. Haskell, "Computer-Generated Holograms with Minimum Quantization", J. Opt. Soc. Am. 63, 504 (1973) (Abstract); B.D. Hansche, "Generalized Binary Computer Generated Holograms: Noise Processes", Ph.D Thesis, University of Michigan, 1976.
23. W.B. Davenport, Jr. and W.L. Root, An Introduction to the Theory of Random Signals and Noise (McGraw-Hill, New York, 1958), pp. 267-73.
24. P.L. Ransom, "Synthesis of Complex Optical Wavefronts", Appl. Opt. 11, 2554-61 (1972).
25. K. Biedermann, "The Scattered Flux Spectrum of Photographic Materials for Holography", Optik 31, 367-89 (1970).
26. I. Cindrich, C. Aleksoff, J. Fienup, A. Klooster, and R. Dallaire, "Study for an Optical Processor System for VLA Radiotelescope Data", Final Report to NRAO, ERIM Report No. 123400-8-F (April 1977).