A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television

By PIERRE MERTZ and FRANK GRAY

By the use of a two-dimensional Fourier analysis of the transmitted picture a theory of scanning is developed and the scanning system related to the signal used for the transmission. On the basis of this theory a number of conclusions can be drawn:

1. The result of the complete process of transmission may be divided into two parts, (*a*) a reproduction of the original picture with a blurring similar to that caused in general by an optical system of only finite perfection, and (*b*) the superposition on it of an extraneous pattern not present in the original, but which is a function of both the original and the scanning system.

2. Roughly half the frequency range occupied by the transmitted signal is idle. Its frequency spectrum consists of alternating strong bands and regions of weak energy. In the latter the signal energy reproducing the original is at its weakest, and gives rise to the strongest part of the extraneous pattern. In a television system these idle regions are several hundred to several thousand cycles wide and have actually been used experimentally as the transmission path for independent signaling channels, without any visible effect on the received picture.

3. With respect to the blurring of the original all reasonable shapes of aperture give about the same result when of equivalent size. The sizes (along a given dimension) are determined as equivalent when the apertures have the same radius of gyration (about a perpendicular axis in the plane of the aperture).

4. With respect to extraneous patterns certain shapes of aperture are better than others, but all apertures can be made to suppress them at the expense of blurring. An aperture arrangement is presented which almost completely eliminates extraneous pattern while about doubling the blurring across the direction of scanning as compared with the usual square aperture. From this and other examples the degradation caused by the extraneous patterns is estimated.

IN the usual telephotographic or television systems the image field is scanned by moving a spot or elementary area along some recurring geometrical path over this field. In the more common arrangement this path consists simply of a series of successive parallel strips. Imagining the path developed or straightened out (or in the more common case, the strips joined end to end), this method of scanning is equivalent to transmitting the image in the form of a long narrow strip.

The theoretical treatment of such transmission has usually been developed by completely ignoring variations in brightness across the image strip, assuming the brightness to have a uniform distribution across this strip. This permits the image to be analyzed as an ordinary one-dimensional or single Fourier series (or integral) along the length of the strip; and the theory is then developed in terms of the one-dimensional steady state Fourier components. Such a method of treatment naturally gives no information in regard to the reproduction or distortion of the detail in the original image across the direction of scanning, nor, as will appear below, does it give any detailed information in regard to the fine-structure distribution of energy over the frequency range occupied by the signal.

The need of a more detailed theoretical treatment originally arose in connection with studies of the reproduction of detail in telephotographic systems, especially in comparisons of distortion occurring along the direction of scanning with that across this direction. Later, this same need was strikingly shown by the discovery that a television signal leaves certain parts of the frequency range relatively empty of current components. Certain considerations indicated that a large part of the energy of a signal might be located in bands at multiples of the frequency of line scanning. Actual frequency analyses more than confirmed this suspicion. The energy was found to be so closely confined to such bands as to leave the regions between relatively empty of signal energy.

Such bands and intervening empty regions are illustrated by the examples of current-frequency curves in Fig. 1. These curves were taken with the various subjects as indicated, and the television current was generated by an apparatus scanning a field of view in 50 lines at a rate of about 940 lines per second. The energy is grouped in bands at multiples of 940 cycles and the regions between are substantially devoid of current components. In addition to the bands shown by the curves, it is known that similar bands occur up to about 18,000 cycles and that there is also a band of energy extending up from about 20 cycles.

Certain of the relatively empty frequency regions were also investigated by including a narrow band elimination filter in a television circuit. The filter eliminated a band about 250 cycles wide and was variable so that the band of elimination could be shifted along the frequency scale at will. By shifting the region of elimination along in this manner it was found that a band about 500 or 600 cycles wide could be removed from a television channel between any two of the current components without producing any detectable effect on the reproduced image.

At a later date a 1500-cycle current suitable for synchronization was introduced into a relatively empty frequency region, transmitted over the same channel with a television current, and filtered out—all without visibly affecting the image.

These results indicated quite clearly the need of a more complete

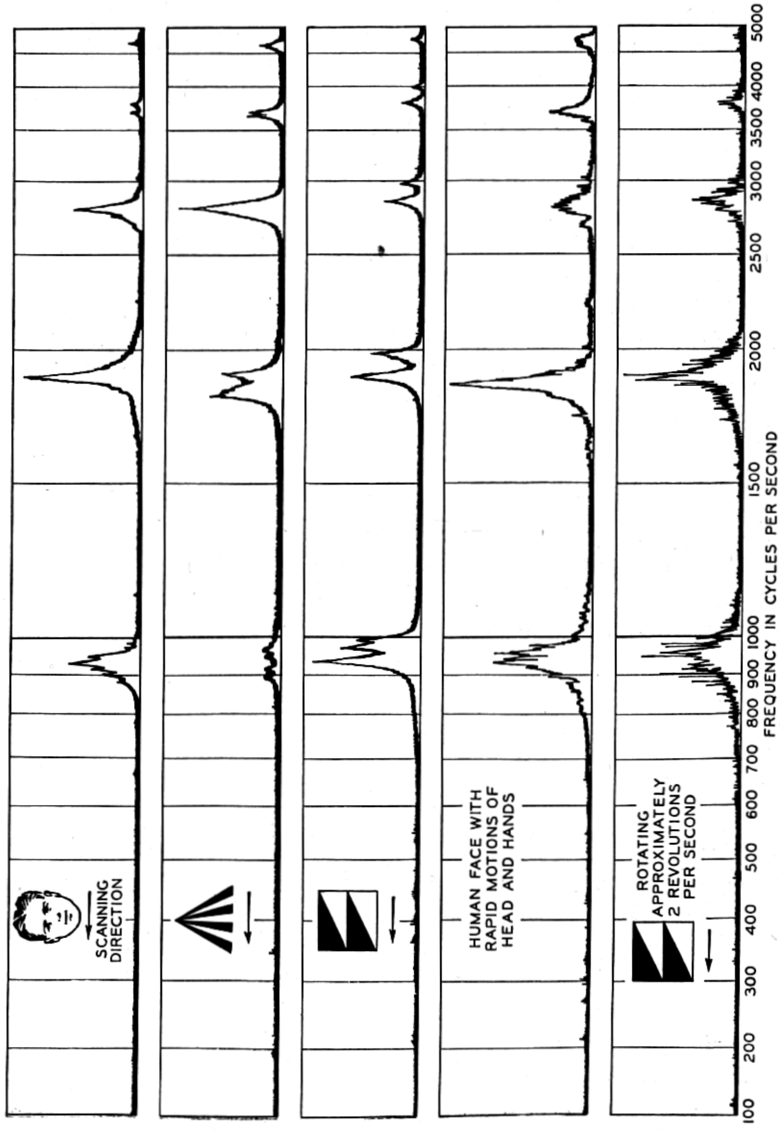


Fig. 1—Frequency analyses of television currents.

theory of the scanning processes used in telephotography and television and led to the study outlined in the following pages. Since this study will be confined to characteristics of the scanning processes all other processes in the system, wherever used, will be assumed to be perfect and cause no distortion.

The general trend of this more complete theory can be foreseen when it is considered that to obtain an adequate reproduction of the original it is necessary to scan with a large number of lines as compared with the general pictorial complexity of this original. This means that for any original presenting a large scale pattern (as distinguished from a random granular background) the signal pattern along successive scanning lines will, in general, differ by only small amounts. Thus, the signal wave throughout a considerable number of scanning lines may be represented to within a small error by a function periodic in the scanning frequency. Since such a function, developed in a Fourier series, is equal to the sum of sine waves having frequencies which are harmonics of the scanning line frequency, it will be natural to expect the total signal wave to have a large portion of its energy concentrated in the regions of these harmonics.

Furthermore, the existence of signal energy at odd multiples of half the scanning frequency will indicate the existence of a characteristic in the picture which repeats itself in alternate scanning lines. It is to be expected that such detail in a picture cannot be transmitted without accurate registry between it and the scanning lines and that when the detail spacing or direction or both differ somewhat from the scanning line spacing and direction, beat patterns between the two will be produced in the received picture which may be strong enough to alter considerably the reproduction of the original.

These phenomena are exactly what is observed, and will be treated in more quantitative fashion in the discussion below.[1]

An Image Field as a Double Fourier Series

Let us first consider the usual expression of the image field as a single Fourier series. The picture will be considered as a "still" so that entire successive scannings are identical. Then if the long strip corresponding to one scanning extends from $-L$ to $+L$, the illumina-

[1] In the following treatment an effort has been made to confine the necessary mathematical demonstrations almost exclusively to two sections entitled, respectively, "Effect of a Finite Aperture at the Transmitting Station," and "Reconstruction of the Image at the Receiving Station." Even in these sections a number of conclusions are explained in text which do not require reading the mathematics if the demonstrations are taken for granted. The occasional mathematical expressions occurring in the earlier sections are very largely for the purpose of introducing notation.

tion E as a function of the distance x along the strip may be expressed as the sum of an infinite number of Fourier components, thus:

$$E(x) = \sum_{n=0}^{\infty} a_n \cos\left(\frac{n\pi x}{L} + \varphi_n\right). \tag{1}$$

In this summation a_n represents the intensity of the nth component and φ_n its phase angle. The complete array of these for all components will vary if the picture is changed.

The cosine series above is very convenient for physical interpretation. It will be simple, however, for some of the later mathematical work to use the corresponding exponential series. The cosine series can be returned to, each time, as physical interpretation is required. That is, since

$$2a \cos\left(\frac{\pi x}{L} + \varphi\right) = (ae^{i\varphi})e^{(i\pi x/L)} + (ae^{-i\varphi})e^{(-i\pi x/L)} \tag{2}$$

the series in equation (1) can be written

$$E(x) = \sum_{n=-\infty}^{+\infty} A_n \exp i\pi(nx/L) \tag{3}$$

if we make

$$A_n = (1/2)a_n \exp(i\varphi_n)$$

and

$$A_{-n} = (1/2)a_n \exp(-i\varphi_n) \tag{4}$$

and if we use the notation $\exp \theta = e^{\theta}$.

In this new summation the complex amplitude A_n represents both the absolute intensity and the phase angle of the nth component. The complex amplitude of the corresponding component with a negative subscript is merely the conjugate of this.

As has already been noted, however, and as might readily be expected, the single Fourier series in equations (1) or (3) above do not always represent a two-dimensional picture with sufficient completeness. In order to consider the two-dimensional field more in detail, let us assume that Fig. 2 represents such an image field of dimensions $2a$ and $2b$, and take axes of reference x and y as indicated. The brightness or illumination of the field is a function $E(x, y)$ of both x and y. Along any horizontal line (i.e., in the x direction, constantly keeping $y = y_1$) the illumination may be expressed as a single Fourier series

$$E(x, y_1) = \sum_{m=-\infty}^{+\infty} A_m \exp i(mx/a). \tag{5}$$

Along any other line in the x direction a similar series holds with different coefficients, that is, the A's are functions of y. They may

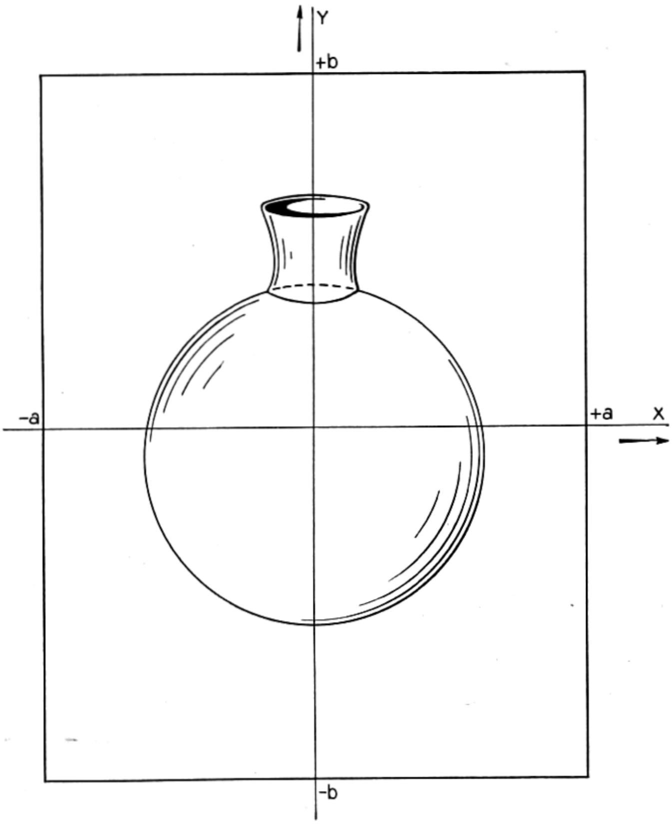


Fig. 2—Scanned field and image.

therefore each be written as a Fourier series along y

$$A_m = \sum_{n=-\infty}^{+\infty} A_{mn} \exp i\pi(ny/b). \tag{6}$$

Substitution in equation (5) gives the double Fourier series,

$$E(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} A_{mn} \exp i\pi \left(\frac{mx}{a} + \frac{ny}{b} \right). \tag{7}$$

For purposes of physical interpretation, as in the case of the simple Fourier series, it is desirable to combine the $+m$, $+n$ term with the $-m$, $-n$ term (giving the single $(m, +n)$th component) and similarly the $+m$, $-n$ with the $-m$, $+n$ terms (giving the single $(m, -n)$th component). This brings equation (7) back to a cosine series,

$$E(x, y) = \sum_{m=0}^{\infty} \sum_{n=-\infty}^{+\infty} a_{mn} \cos \left[\pi \left(\frac{mx}{a} + \frac{ny}{b} \right) + \varphi_{mn} \right] \tag{8}$$

when

$$A_{mn} = (1/2) a_{mn} \exp (i\varphi_{mn})$$

and

$$A_{-m-n} = (1/2) a_{mn} \exp (-i\varphi_{mn})$$

and where a_{mn} is always a real quantity. Each term of this series represents a real, two-dimensional, sinusoidal variation in brightness extending across the image field. The image is built up of a superposition of a series of such waves extending across the field in various directions and having various wave lengths.

Imagining brightness as a third dimension, we may, as an aid in visualizing the components of an image field, draw separate examples of various components as shown in Fig. 3. It will be noted that any given component (m, n) passes through m periods along any horizontal line in the image field, and through n periods along any vertical line. The slope of the striations with respect to the x-axis is therefore $-mb/na$ (the negative reciprocal of the slope of the line of fastest variation in brightness). For the same values of m and of n, the m, $+n$ component and the m, $-n$ component have equal wave lengths but are sloped in opposite directions to the x-axis. If m is zero the crests are parallel to the x-axis; if n is zero they are parallel to the y-axis. The component with both m and n zero is a uniform distribution of brightness covering the entire image field. The wave length of a component is

$$\lambda_{mn} = 1 \Big/ \sqrt{\left(\frac{m}{2a}\right)^2 + \left(\frac{n}{2b}\right)^2}.$$

A complete array of the components, up to m and n equal to 4, is illustrated in Fig. 4.

As of course is characteristic of the harmonic analysis, the wave lengths and orientations of the components are seen to vary only with the shape and size of the rectangular field, and to be independent of the particular subject in the field. A change of subject, or motion of the subject, merely alters the amplitudes of the components and shifts their phase; but their wave length and inclination with respect to the x-axis remain unchanged. Consequently, for the same rectangular field all subjects appearing in it may be considered as built up from the same set of components. For a "still" subject, the amplitudes and phase angles of the cosine components, or the complex amplitudes of the exponential components, remain constant with time. For a moving subject these complex amplitudes may be considered modulated as functions of time.

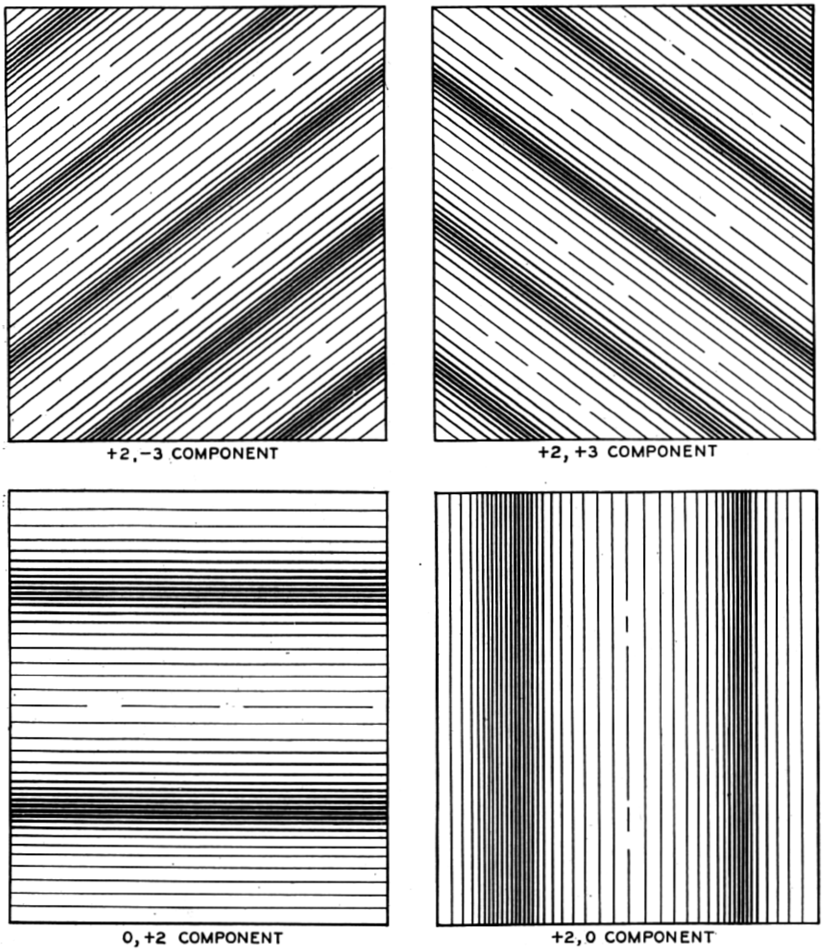


Fig. 3—Examples of field components.

The real amplitudes a_{mn} for a circular area of uniform brightness on a black background are relatively easily calculated, and this subject is also a good one to study as a picture from some points of view because it has a simple sharp border sloping in various directions. The amplitudes for a circle of unit illumination of radius R are

$$a_{mn} = \frac{\pi R^2}{2ab}\left(\frac{\lambda_{mn}}{2\pi R}\right) J_1\left(\frac{2\pi R}{\lambda_{mn}}\right), \tag{9}$$

where J_1 is the first order Bessel function. In this particular subject all components of a given wave length have equal amplitudes; and the

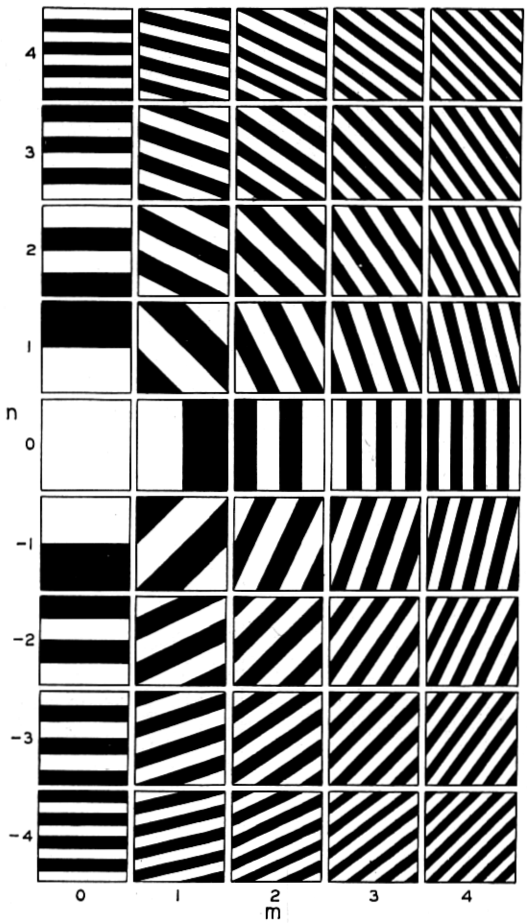


Fig. 4—Array of field components.

amplitudes may therefore be plotted as a function of wave length alone, as in Fig. 5. The curve illustrates the rapidity with which the amplitudes fall off for the higher order components in a subject of this nature.

The Frequency Spectrum of the Signal

When an image field is scanned by a point aperture tracing across it, each portion of the picture traversed causes variations in the light reaching the light sensitive cell and is thus translated into a corre-

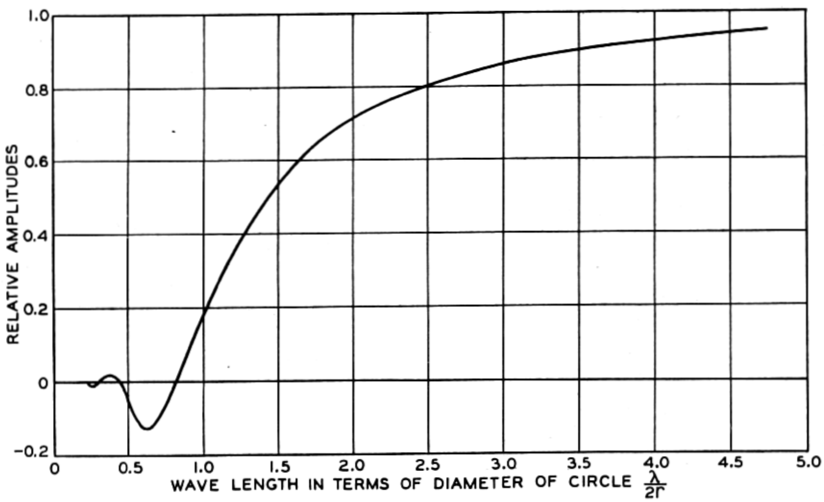


Fig. 5—Amplitudes of field components for a circular area of brightness.

sponding signal. Further, each Fourier component in the field is translated into a corresponding Fourier component in the signal. An equivalent translation occurs when a pencil of light traces over a photographic film in telephotography, or when a subject is scanned by a beam of light in television, whether or not a simple flat two-dimensional image is ever physically formed at the transmitting station. For clarity and simplicity, the discussion will be confined to the case in which a point aperture traces across a plane image field.

In most systems the aperture traces a line across the field and then there is a sudden jump back to the beginning of the next succeeding line. This discontinuous motion is naturally not easily subjected to mathematical treatment. It is much simpler to deal with the equivalent result that would be obtained if the scanning point, instead of tracing successive parallel paths across the same field, moved continuously across a series of identical fields. Such an equivalent scanning motion can fortunately easily be used because a double Fourier series represents not only a single field, but a whole succession of identical image fields covering the entire xy plane, and repeated periodically in both the x and y direction as illustrated in Fig. 6.

The equivalent of scanning a single field in parallel lines is obtained by assuming that the scanning point moves across the repeated fields along a sloping path as indicated. Let u be the velocity parallel to

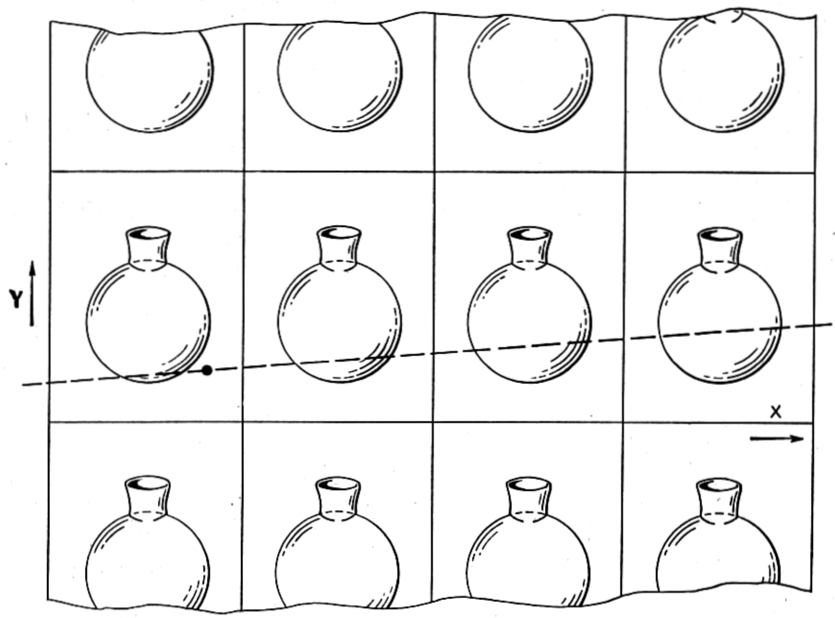


Fig. 6—Array of periodically recurring scanned fields.

the x axis and v the velocity parallel to the y axis. Then the picture illumination at the scanning point at any instant, and consequently the signal current, may be obtained by substituting

$$x = ut, \qquad y = vt$$

in the double Fourier series representing the image field, equation (8). Of course the entire expression must be multiplied by a factor K which is the constant ratio between the signal current and the picture illumination. This gives for the real signal as a function of time [2]

[2] It will be noted that this process does not explore the picture completely, inasmuch as, no matter how fine the scanning, there will always be unexplored regions between scanning lines. In this respect the process is quite analogous to that followed in analyzing a function of a single variable into a simple Fourier series when the values of the function are given only at discrete (even though closely spaced) values of the variable. The complete exact theory, which necessarily depends upon the size and shape of the finite scanning spot or aperture, will be given further below.

$$I(t) = K \sum_{m=0}^{\infty} \sum_{n=-\infty}^{+\infty} a_{mn} \cos \left[\pi \left(\frac{mu}{a} + \frac{nv}{b} \right) t + \varphi_{mn} \right]. \tag{10}$$

Thus if u and v are constants, each wave of the image field gives rise to a corresponding Fourier component of the signal. The frequencies of the signal components are

$$f = \frac{mu}{2a} + \frac{nv}{2b}. \tag{11}$$

The frequency spectrum of the signal is thus made up of a series of possible discrete lines, the position of which in that spectrum is determined by u and v, that is, by the particular scanning motion employed. We shall designate these lines by the indices $m, +n$ and $m, -n$, as they are correlated with the particular components of the image field that generated them.

A different choice of values for u and v (so long as these, once having been chosen, remain constant) changes the location of the lines in the frequency spectrum, but their amplitudes, depending only on the corresponding components of the image field, remain unchanged. In other words, the lines in the frequency spectrum of the signal are characteristic of the image field, and the scanning motion merely determines where they will appear in the frequency spectrum. Thus, if for a given subject the distribution of energy over the frequency scale is known for one method of scanning, it can be predicted for a great many other methods.

To scan a field in lines approximately parallel to the x axis, the velocity v must be made small compared to u. Under such conditions, $u/(2a)$ of equation (11) is the line scanning frequency and $v/(2b)$ is the frequency of image repetitions (or "frame frequency"). The frequency spectrum of the signal for a "still" picture thus consists of certain fundamental components at multiples of the line scanning frequency $u/(2a)$, each of which is accompanied by a series of lines spaced at equal successive intervals to either side of it. The spacing between these satellites is the image repetition or frame frequency $v/(2b)$.

If the picture changes with time the amplitudes of these fundamental lines and their satellites are modulated, also with respect to time. In other words they each develop sidebands or become diffuse. The diffuseness will not overlap from satellite to satellite unless the frequency of modulation becomes as great as half the frame frequency.

Thus for motions in the picture which are not too fast to be expected to be reproduced with reasonable fidelity, this diffuseness of the fundamental lines and their satellites will not obliterate their identity.

A diagrammatic arrangement of some of the possible lines in a frequency spectrum, with their corresponding m and n indices, is shown in Fig. 7.

It is important to note that the correlation between the wave lengths of the field components and the frequencies of the current components is not the one that is naturally assumed on first consideration. We

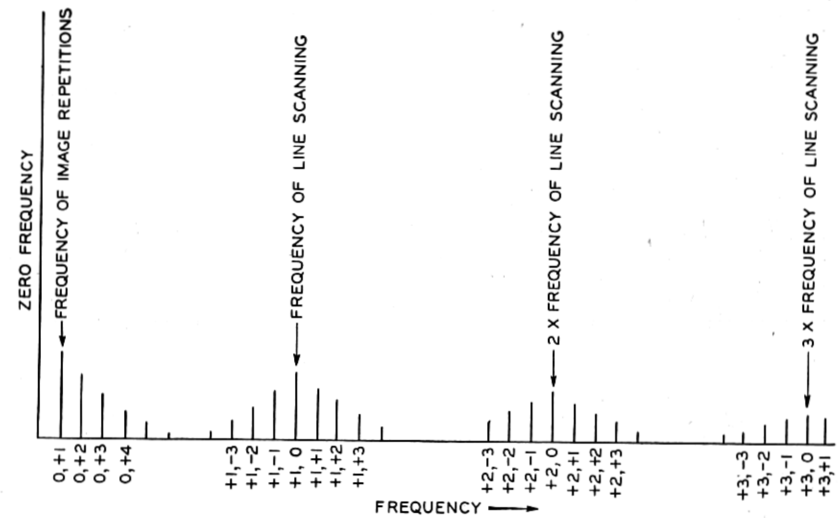


Fig. 7—Diagram of signal frequency spectrum.

are quite likely to make the erroneous assumption that high frequencies correspond to all sharp changes in brightness and that low frequencies correspond only to slow changes. The error in this assumption is readily realized by noting that sharp changes in brightness may generate very low frequencies if the scanning point passes over them in a sloping direction. An actual correlation is shown schematically in Fig. 8. It is seen that the same general type of correlation is repeated periodically over the frequency scale at multiples of the line scanning frequency. There are evidently numerous regions of the spectrum in which short image waves, or fine grained details of the image field, may appear in the signal. They are not confined to the high-frequency region alone.

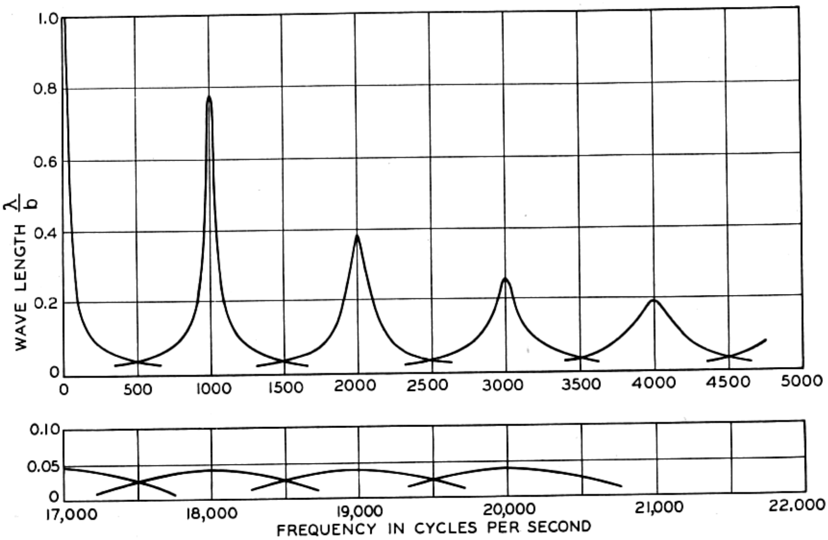


Fig. 8—Correlation between wavelength and frequency of signal components.

In telephotography the frequency of line scanning is usually low and the groups of lines in the frequency spectrum are so closely spaced that such fine grained details of the signal are of little practical importance as far as the electrical parts of the system are concerned. In television, however, these bands are widely spaced, of the order of 1000 cycles or so apart, and such details of the signal are quite important.

As a specific example, it is interesting to plot the frequency spectrum of the television signal that results from scanning a circular area of uniform brightness on a black background. So far as the present theory extends, this may be done by converting the field components of equation (9) into current components with the aid of equations (10) and (11). Taking $b/a = 1.28$, the radius of the circle as $b/3$, and assuming that the field is scanned in 50 lines 20 times per second, we obtain the amplitude-frequency spectrum shown in Fig. 9. Since it is not convenient to show the individual current components—only 20 cycles apart—the curve shows simply the envelope of the peaks of these components. At low frequencies, the energy is largely confined to bands at multiples of the line scanning frequency, 1000 cycles, and to an additional band extending up from zero frequency. In the retions between the bands, the signal components are so small that they do not show when plotted to the same scale. At higher frequencies the signal energy as thus far computed is not confined to such bands. It

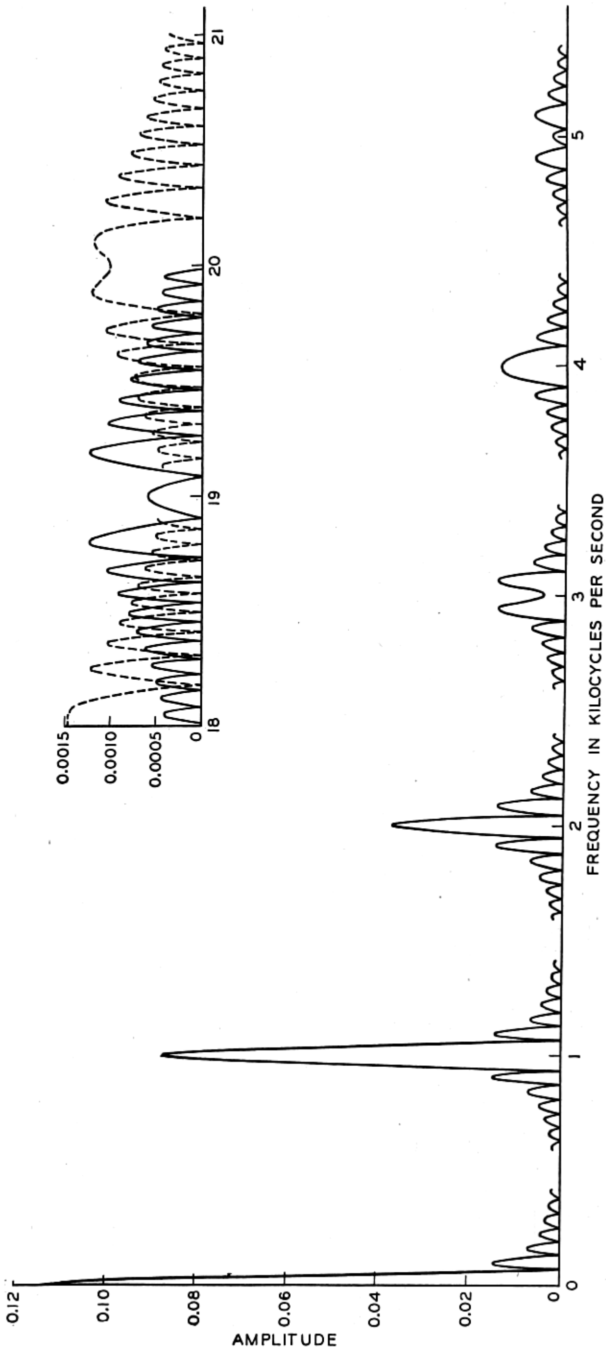


Fig. 9—Amplitudes of signal components for a circular area of brightness (diameter, 1/3 of field length).

will be shown farther on, however, that the effect of the use of a finite aperture for scanning is to confine the signal energy more rigorously to such bands throughout the frequency range.

The theoretical energy distribution for the circular area is in excellent agreement with actual frequency analyses of television currents, which show the energy confined to bands at multiples of the line scanning frequency with apparently empty regions between. It is evident from the theory so far, however, that these regions are not really empty but are filled with weak signal components representing fine details of the subjects; and subjects of greater pictorial complexity than a simple circular area may be devised to give large signal components in such regions. We must therefore look for other factors to explain why these frequency regions do not transmit any appreciable details of an image.

Confusion in the Signal

With the usual method of scanning, one such factor is the confusion of components in the signal. This confusion arises from the fact that two or more image components sloping across the field in different directions may intercept the line of scanning with their crests spaced exactly the same distance apart along this line of scanning. As the scanning point passes over them they thus give rise to signal current components of exactly the same frequency. Consequently the two image components are represented by a single, confused, signal current component that can transmit no information whatever in regard to their relative amplitudes and phases. This confusion evidently depends on the scanning path.

If the image field is scanned in N lines, the velocity v of the scanning point parallel to the y axis is

$$v = \frac{b}{Na} u \tag{12}$$

and the signal frequencies from equation (11) are

$$f = \frac{u}{2a}\left(m + \frac{n}{N} \right). \tag{13}$$

Field components with indices m, n and m', n' such that

$$m + \frac{n}{N} = m' + \frac{n'}{N} \tag{14}$$

give rise to current components of the same frequency.

In other words, the bands of components in the frequency spectrum really overlap. Consequently the components of one band may coincide in frequency with the components of adjacent bands. Such coinciding components are illustrated schematically in Fig. 10.

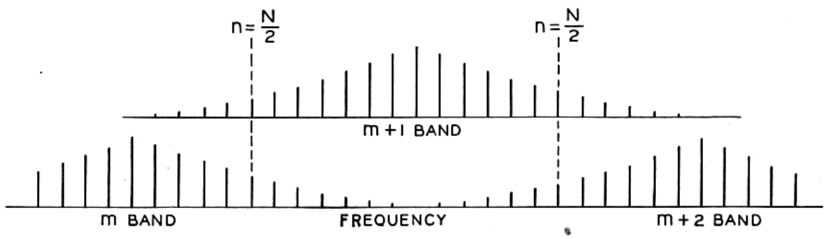


Fig. 10—Coinciding lines of confused bands.

It is obvious that a single a-c. component cannot transmit the separate amplitudes and phases of two or more image components. Consequently the receiving apparatus has no information to judge how the components in the original image are supposed to be distributed in the reproduction.

The situation is most serious where the intensities of coinciding components have the same order of magnitude, that is, at the centers of the frequency regions intermediate to the strong bands. The confusion in these regions is the most important factor that renders them incapable of transmitting any appreciably useful image detail.

On first consideration it would appear that the overlapping of bands in the signal might result in a hopeless confusion. The situation is saved, however, by the fact that components with large n numbers will tend to be weak due to the convergence of the Fourier series, and are further reduced, as will be shown later, by the effects of a scanning aperture of finite size. They therefore do not usually seriously interfere with the stronger components. The interference usually manifests itself in the form of serrations on diagonal lines and occasional moiré effects in the received picture.

Confusion in the signal may be practically eliminated by using an aperture of such a nature that it cuts off all components with n numbers greater than $N/2$, that is, cuts off each band before it reaches the center of the intervening frequency regions so that adjacent bands do not overlap. The practical possibilities of this arrangement will be discussed further below.

The mere elimination of confusion in the signal itself does not necessarily prevent the appearance of extraneous components in the reproduced image. The receiving apparatus itself must be so designed that when it reproduces all the image components represented by a given signal component, it suitably suppresses all those but the dominant one desired.

Effect of a Finite Aperture at the Transmitting Station

In the preceding pages the scanning aperture has been assumed as infinitesimal in size, or merely a point. In any actual scanning system the necessary finite size of the aperture introduces effects which will now be considered for the transmitting end.

Let us first review briefly the usual theory of this effect when the picture is analyzed simply as a one-dimensional Fourier series. According to equation (3) above, this series is

$$E_1(x) = \sum_{n=-\infty}^{+\infty} A_n \exp i\pi(nx/L).$$

Let ξ be a coordinate fixed with respect to the scanning aperture as shown in Fig. 11 and let the optical transmission of the aperture for

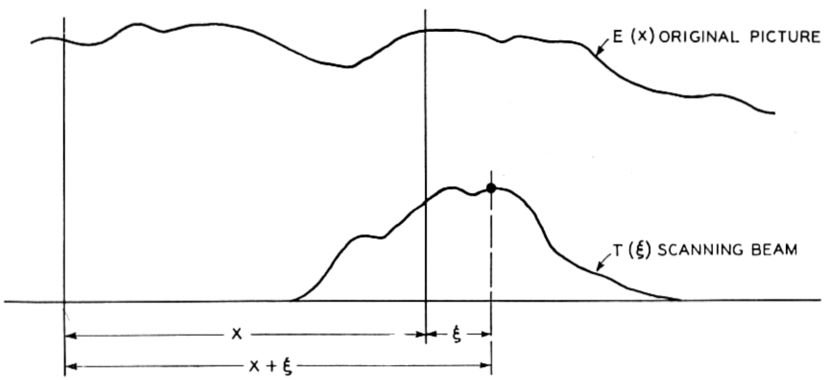


Fig. 11—Analysis of one-dimensional scanning operation.

any value of ξ be $T(\xi)$.[3] Then if x is taken as a coordinate of the origin of ξ the illumination at any point ξ of the aperture is

$$E_1(x + \xi) = \sum_{n=-\infty}^{+\infty} A_n \exp i\pi(n[x + \xi]/L), \qquad (15)$$

[3] This optical transmission may represent either the transparency of an aperture of constant width or the width of an aperture which is a shaped hole in an opaque screen.

so that the total flow of light through the aperture at any position x is

$$F_1(x) = \int_{\text{aperture}^4} T(\xi)E_1(x + \xi)d\xi. \tag{16}$$

Since x is a constant with respect to the integration the exponential term may be factored and the part involving x only may be brought outside the integral sign. This gives

$$F_1(x) = \sum_{n=-\infty}^{+\infty} Y(n)A_n \exp i\pi(nx/L), \tag{17}$$

where

$$Y(n) = \int_{\text{aperture}} T(\xi) \exp i\pi(n\xi/L)d\xi. \tag{17'}$$

For a symmetrical aperture (that is, about the origin of ξ)

$$Y(n) = \int_{\text{aperture}} T(\xi) \cos (\pi n\xi/L)d\xi \tag{17''}$$

and $Y(n)$ in this case is, therefore, a pure real quantity.

The important conclusion to be drawn from equation (17) as to the effect of a finite transmitting aperture is that it multiplies the complex amplitude A_n of each original image component by a quantity $Y(n)$ which is independent of the picture being scanned. This is entirely similar to the effect of a linear electrical network in a circuit, and the quantity $Y(n)$ is quite analogous to the transfer admittance of that network.

The quantity $Y(n)$ has been plotted for variously shaped apertures in Fig. 12. For convenience in comparison, the ordinates of each curve have been multiplied by a numerical factor to make $Y(0) = 1$. The curves show the characteristics that are by this time familiar, which are that the effect of the finite size of the scanning aperture in the transmitter is similar to that of introducing a low-pass filter in the circuit, namely, cutting down the amplitudes of the signal components for which n is numerically high, i.e., the high-frequency components.

The curves are remarkable, however, in that in the useful frequency band (i.e. from $n = 0$ to something like half of the first root of $Y(n) = 0$) all the distributions considered give practically the same transfer admittance if the dimensions of the beam along the direction of scanning are suitably chosen, as has been done in the figure. This results from

[4] The integral is mathematically taken from $-\infty$ to $+\infty$ but the regions outside the aperture give no contribution since the integrand is there equal to zero.

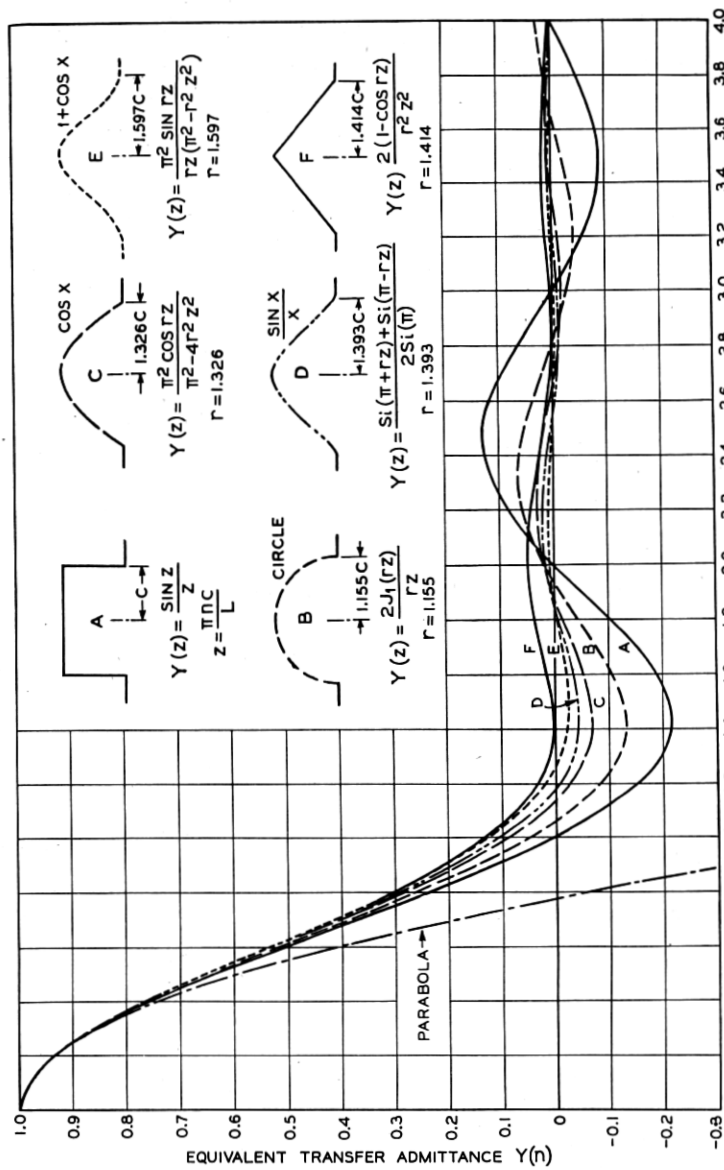


Fig. 12—Equivalent transfer admittances for various distributions of light in the scanning beam.

the physical limitation that the illumination in any part of the beam must be positive, that is, the illumination from one part of the beam must always add to that from another part and cannot subtract from it.[5] This observation enables one to define the resolution of two apertures of different shapes as being equal along a certain direction when their transfer admittances in the useful frequency range show the same filtering effect, if that direction is used as the direction of scanning. This will occur when the radii of gyration (about a normal axis in the plane of each aperture) are equal. According to this definition all the apertures illustrated in Fig. 12 have the same resolution along a horizontal direction.

When the picture is analyzed as a two-dimensional Fourier series the equations which have been given above become

$$E_1(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} A_{mn} \exp i\pi \left(\frac{mx}{a} + \frac{ny}{b} \right),$$

$$E_1(x + \xi, y + \eta) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} A_{mn} \exp i\pi \left(\frac{m[x + \xi]}{a} + \frac{n[y + \eta]}{b} \right), \tag{18}$$

$$F_1(x, y) = \iint_{\text{aperture}} T_1(\xi, \eta) E_1(x + \xi, y + \eta) d\xi d\eta, \tag{19}$$

$$F_1(x, y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} Y_1(m, n) A_{mn} \exp i\pi \left(\frac{mx}{a} + \frac{ny}{b} \right), \tag{20}$$

where

$$Y_1(m, n) = \iint_{\text{aperture}} T_1(\xi, \eta) \exp i\pi \left(\frac{m\xi}{a} + \frac{n\eta}{b} \right) d\xi d\eta. \tag{20'}$$

For an aperture symmetrical about both ξ and η axes

$$Y_1(m, n) = \iint_{\text{aperture}} T_1(\xi, \eta) \cos \pi \left(\frac{m\xi}{a} + \frac{n\eta}{b} \right) d\xi d\eta. \tag{20''}$$

[5] The shape of the transfer admittance curve near $n = 0$ depends upon the power of n in the first variable term of the Taylor expansion for $Y(n)$ about $n = 0$, and upon the sign of this term. Assuming a symmetrical aperture, the expansion from equation (17'') is

$$Y(n) = \int T d\xi - \frac{\pi^2 n^2}{2!\, L^2} \int \xi^2 T d\xi + \frac{\pi^4 n^4}{4!\, L^4} \int \xi^4 T d\xi - \cdots.$$

Since T is everywhere positive the first variable term is always in n^2 and negative. The shape of the curve near $n = 0$ is, therefore, always a parabola (indicated in Fig. 12), which can be made the same parabola by suitably choosing the two disposable constants in the aperture. Even after departing from this common parabola, the curves maintain the same general shape over a substantial range; for the next variable term is in n^4 and positive, and has the same order of magnitude for all usual types of apertures. Consequently, the curves for these apertures have approximately the same shape over a wide range extending up from $n = 0$. The results are the same for an unsymmetrical aperture, but the reasoning is more involved.

In the two-dimensional case $T(\xi, \eta)$ is defined, for a hole in an opaque screen, as unity throughout the area of the hole, and zero for the screen. Where the aperture is covered with a non-uniform screen T may take on intermediate values.

The transfer admittances have been calculated for a variety of shapes of aperture in Appendix I. It will be noted that for those types of aperture for which T can be separated into two factors, one a function of ξ only and the other a function of η only, namely, for which

$$T_1(\xi, \eta) = T_\xi(\xi) \cdot T_\eta(\eta), \tag{21}$$

then equation (20′) becomes

$$Y_1(m, n) = \int_{\text{aperture}} T_\xi(\xi) \exp(i\pi m\xi/a) d\xi \int_{\text{aperture}} T_\eta(\eta) \exp(i\pi n\eta/a) d\eta$$
$$= Y_\xi(m) \cdot Y_\eta(n) \tag{22}$$

and Y_ξ and Y_η are each one-dimensional integrals of the type illustrated in Fig. 12.

The rectangular aperture is a simple case of this type. Assume the field to be scanned in N lines and take the dimensions of the aperture, $2c$ and $2d$ parallel to the x and y axes, respectively, as

$$\frac{c}{a} = \frac{d}{b} = \frac{1}{N}. \tag{23}$$

Then

$$Y_\xi(m) = \frac{\sin \pi mc/a}{\pi mc/a}$$

and

$$Y_\eta(n) = \frac{\sin \pi nd/b}{\pi nd/b}$$

and the frequency corresponding to a given signal component mn is, from equation (11)

$$f = \frac{u}{2a}\left(m + \frac{n}{N}\right).$$

Thus, $Y_1(m, n)$ considered as a function of the signaling frequency corresponding to each component of indices mn, consists of a succession of similar curves $u/2a$ cycles apart, corresponding to the successive integral values of m (these curves are themselves really not continuous but consist of a succession of points $u/(2aN)$ cycles apart. For convenience, however, the drawings will always show the curves as con-

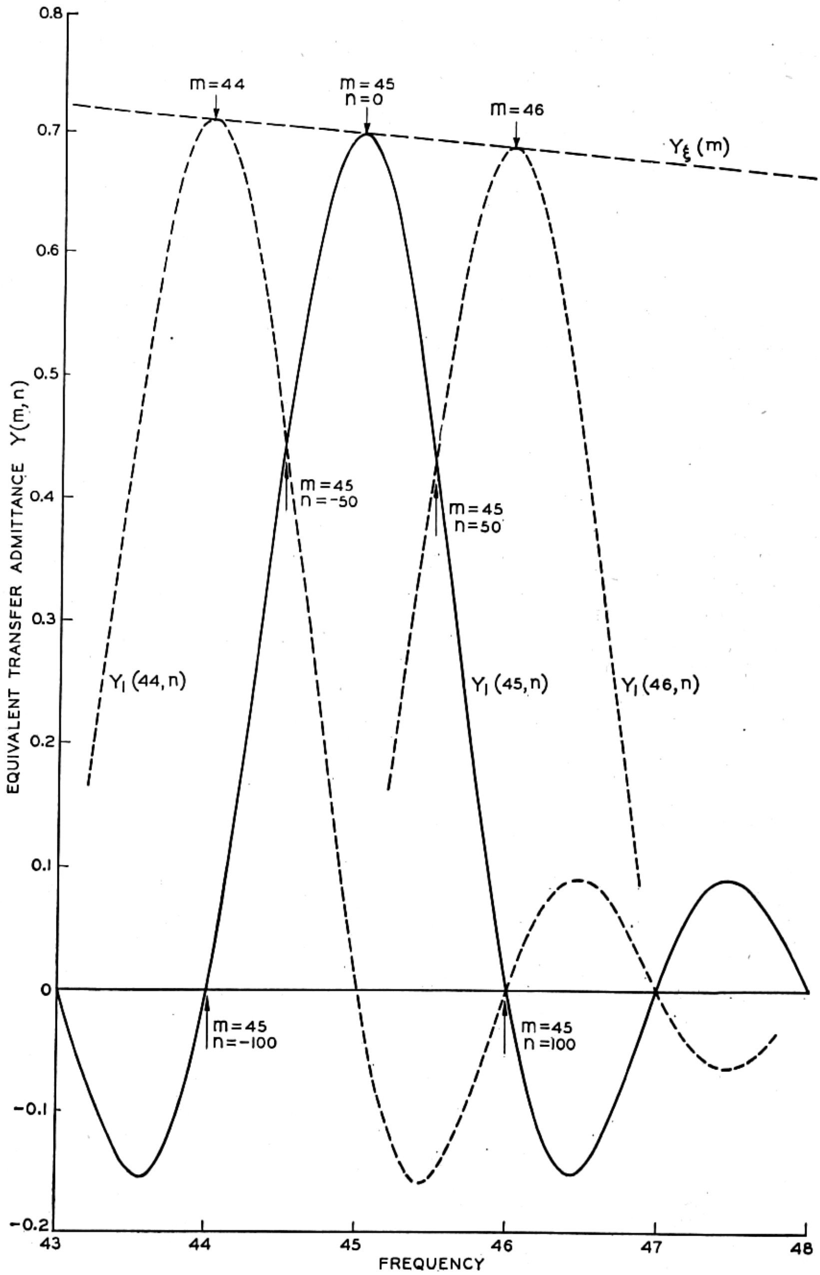


Fig. 13a—Detail of equivalent transfer admittance of aperture for two-dimensional scanning.

tinuous). Each of the curves is of the equation

$$Y_1(m, n) = Y_\xi(m) \frac{\sin \frac{2\pi Nc}{u}\left(f - \frac{mu}{2a}\right)}{\frac{2\pi Nc}{u}\left(f - \frac{mu}{2a}\right)}$$

and therefore has a peak of the value $Y_\xi(m)$ at the point where $n = 0$ or $f = mu/2a$, and trails off from the peak in each direction according to a curve of the same shape as curve "A" in Fig. 12. The successive curves are all of identical shape, but each one is to a reduced scale of ordinates as compared with the preceding (in the useful frequency range) as imposed by the factor $Y_\xi(m)$.

The peaks, it will be noted, occur at the frequencies occupied by what have been called the fundamental components (as distinguished from the satellite lines) in the discussion above on the frequency spectrum of the signal.

Assuming N to be 100 and for simplicity taking the factor $u/2a$ as equal to 1, a plot is shown in Fig. 13a of $Y_1(m, n)$ over a very limited region near the upper end of the useful frequency range. The curve shown in a solid line represents $Y_1(m, n)$ for $m = 45$, and the dotted curves on either side represent the function for $m = 44$ and 46, respectively.

The function has been redrawn for the complete useful range of frequencies and a little beyond, in Fig. 13b, with the frequencies to a logarithmic scale. This logarithmic plot opens out the scale at the low frequencies and enables the fine structure of the function to be indicated there, and still enables the complete range of useful frequencies to be shown without requiring a prohibitive size of drawing (it has, however, the disadvantages that the distortion in the frequency scale then masks the symmetry of the individual curves around the fundamental lines, the similarity of shape of these individual curves, and also the constant frequency separation between the successive fundamental lines).

The function $Y_1(m, n)$, as is clear from equation (22) and Figs. 13a and b, consists of a sort of envelope function $Y_\xi(m)$, "modulated" by a fine structure function $Y_\eta(n)$. The latter function has the value unity at the positions of the fundamental lines in the frequency spectrum of Fig. 7 and diminishes for the satellite lines in the same way that the envelope function diminishes for the fundamental lines away from zero frequency. It will be seen that the envelope function is the only one obtained by the simple one-dimensional analysis. The

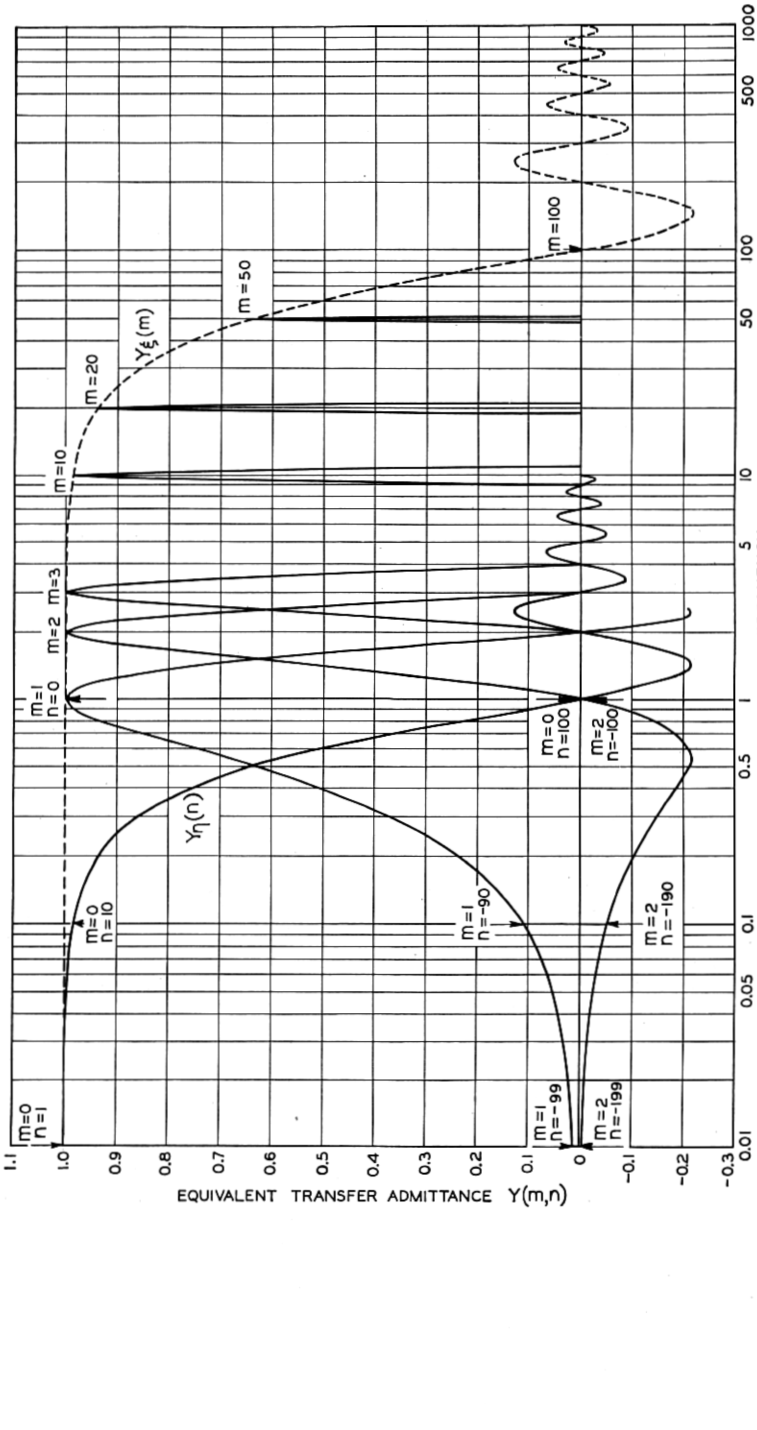


Fig. 13b—Equivalent transfer admittance of aperture for two-dimensional scanning, to a logarithmic frequency scale.

complete function shows, by the very small transfer admittance in the regions half-way between the fundamental lines, an additional reason why the signal currents in these regions will be weak and relatively incapable of transmitting appreciable image detail.

Examination of the other apertures for which computations are given in Appendix I will show that, in general, for all ordinary apertures the same broad phenomena are observed as for the rectangular aperture, although it is not always possible to express the complete function in the simple product form above, in which case the curves for the successive values of m will vary gradually in shape.

The final signal current is proportional to the light flux through the aperture, given in equation (20). Neglecting constant factors it may, therefore, be written as

$$I(t) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} Y_1(m, n) A_{mn} \exp i\pi \left(\frac{mu}{a} + \frac{nv}{b} \right) t. \qquad (24)$$

Reconstruction of the Image at the Receiving Station

At the receiving station the signal current is translated back into light to illuminate an aperture moving in synchronism with the one at the sending end. Neglecting constant factors the flow of light $F_2(t)$ to the receiving aperture is

$$F_2(t) = I(t). \qquad (25)$$

Let $E_2(x, y)$ be the resulting apparent illumination (integrated with respect to time) at a point x, y of the reproduced image, or, in telephotography, the integrated exposure of the recording film at this point. This illumination may be expressed as a double Fourier series, similar to equation (7) (but primed subscripts will be used to distinguish them from those of that equation).

$$E_2(x, y) = \sum_{m'=-\infty}^{+\infty} \sum_{n'=-\infty}^{+\infty} B_{m'n'} \exp i\pi \left(\frac{m'x}{a} + \frac{n'y}{b} \right), \qquad (26)$$

where

$$B_{m'n'} = \frac{1}{4ab} \int_{-b}^{+b} \int_{-a}^{+a} E_2(x, y) \exp -i\pi \left(\frac{m'x}{a} + \frac{n'y}{b} \right) dxdy. \qquad (27)$$

Reproduction of detail in the image may be studied by comparing these components with the corresponding ones of the original image.

The apparent illumination is the same as if the aperture traced a single strip across repeated fields in the xy plane as illustrated in Fig. 14, and all of the repeated fields included between $y = -b$ and

$y = +b$ were cut out and superposed to form the image. Let $E_3(x, y)$ be the illumination of this strip. Then, since the exponential

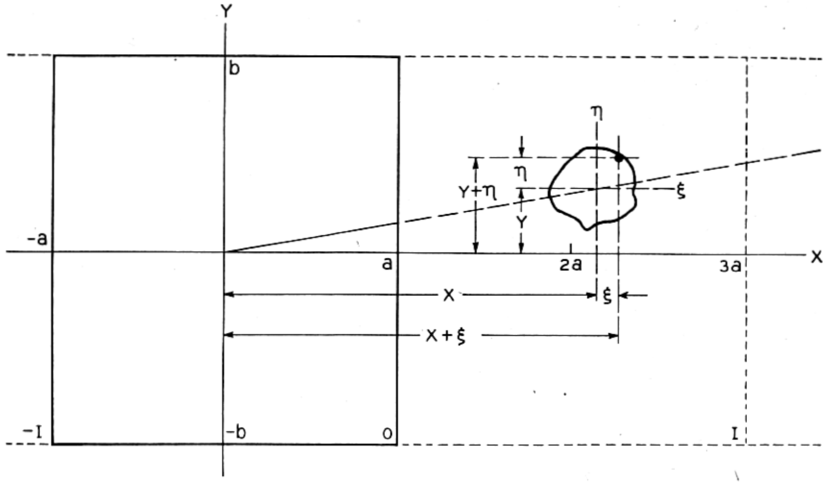


Fig. 14—Analysis of received picture.

factor of the integrand in equation (27) is periodic in x and identically reproduced in each of the fields, the integral is equal to

$$B_{m'n'} = \frac{1}{4ab}\int_{-b}^{+b}\int_{-\infty}^{+\infty} E_3(x, y) \exp -i\pi\left(\frac{m'x}{a} + \frac{n'y}{b}\right) dxdy. \quad (28)$$

The limits $-b$ to $+b$ in y and the infinite limits in x may be used because the illumination is zero everywhere outside of the strip.

Again taking a coordinate system $\xi\eta$ fixed with respect to the aperture, such that

$$x = \xi + ut, \qquad y = \eta + vt \quad (29)$$

the instantaneous illumination of any point covered by the aperture is, neglecting constant factors,

$$T_2(\xi, \eta)I(t) = T_2(x - ut, y - vt)I(t). \quad (30)$$

The total illumination of any point xy in the image strip is thus

$$E_3(x, y) = \int_{-\infty}^{+\infty} T_2(x - ut, y - vt)I(t)dt.$$

Substitution in integral (28) and a change in the order of integration gives

$$B_{m'n'} = \frac{1}{4ab}\int_{-\infty}^{+\infty}\int_{-b}^{+b}\int_{-\infty}^{+\infty} T_2(x - ut, y - vt)I(t) \cdot \exp -i\pi\left(\frac{m'x}{a} + \frac{n'y}{b}\right) dxdydt. \quad (31)$$

Changing to the $\xi\eta$ system

$$B_{m'n'} = \frac{1}{4ab}\int_{-\infty}^{+\infty}\int_{-b-vt}^{b-vt}\int_{-\infty}^{+\infty} T_2(\xi, \eta)I(t) \exp -i\pi\left(\frac{m'u}{a} + \frac{n'v}{b}\right)t \cdot \exp -i\pi\left(\frac{m'\xi}{a} + \frac{n'\eta}{b}\right) d\xi d\eta dt. \quad (32)$$

This integral may be considered as the surface integral of a function $\varphi(\eta, t)$ taken over a strip shaped area shown in Fig. 15, in elements of

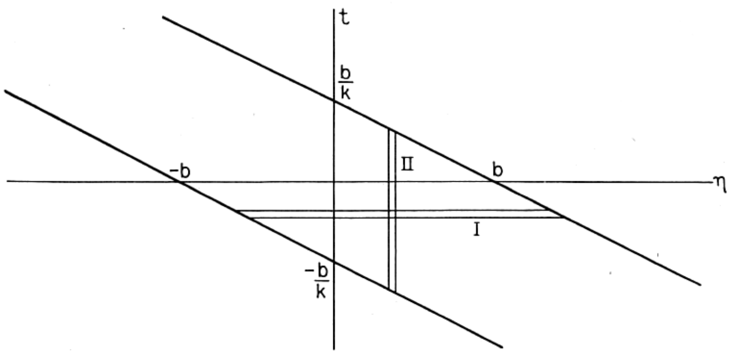


Fig. 15—Equivalent integration regions.

the type indicated as I. From this it may be seen that, when also integrated in elements of the type indicated as II,

$$\int_{-\infty}^{+\infty}\int_{-b-vt}^{b-vt} \varphi(\eta, t)d\eta dt = \int_{-\infty}^{+\infty}\int_{(-b-\eta)/v}^{(b-\eta)/v} \varphi(\eta, t)dtd\eta. \quad (33)$$

Consequently

$$B_{m'n'} = \frac{1}{4ab}\int_{-\infty}^{+\infty}\int_{(-b-\eta)/v}^{(b-\eta)/v}\int_{-\infty}^{+\infty} T_2(\xi, \eta)I(t) \exp -i\pi\left(\frac{m'u}{a} + \frac{n'v}{b}\right)t \cdot \exp -i\pi\left(\frac{m'\xi}{a} + \frac{n'\eta}{b}\right) d\xi dtd\eta. \quad (34)$$

Consider now the intensity $B_{m'n'}$ of a final reproduced picture component m', n' resulting from a single component m, n in the signal as expressed by equation (24). The integral becomes

$$B_{m'n'} = \frac{A_{mn}Y_1(m, n)}{4ab} \int_{-\infty}^{+\infty} \int_{(-b-\eta)/v}^{(b-\eta)/v} \int_{-\infty}^{+\infty} T_2(\xi, \eta) \cdot \exp i\pi \left(\frac{m - m'}{a} u + \frac{n - n'}{b} v \right) t \cdot \exp -i\pi \left(\frac{m'\xi}{a} + \frac{n'\eta}{b} \right) d\xi dt d\eta. \quad (35)$$

It will be noted that the exponential function of t is periodic in t, one of the periods being $t_0 = 2b/v$. Furthermore, this is just the difference between the upper and lower limits in t. Hence the integral in t may be written

$$I = \int_0^{t_0} \exp i\pi \left(\frac{m - m'}{a} u + \frac{n - n'}{b} v \right) t dt.$$

This integral is zero except when

$$\frac{m - m'}{a} u + \frac{n - n'}{b} v = 0, \quad (36)$$

in which case

$$I = t_0. \quad (37)$$

The meaning of these last few equations is clear. It is, as would be expected, that a signal component m, n does not give rise to all components m', n' in the final received picture, but that these latter components are in general zero unless m' and n' satisfy a definite relationship with m and n, expressed by equation (36). A somewhat unexpected result is, however, that equation (36) allows some other m', n' components besides the normal one for which $m' = m$ and $n' = n$. That is to say, a given signal component m, n in the line will reproduce in the final picture not only a corresponding m, n component, but as has been foreshadowed in the discussion on confusion in the signal, it will also reproduce certain other components with different indices.

Let us consider first, however, the reproduction of the normal component for which $m' = m$ and $n' = n$, which is obviously allowed by equation (36). The amplitude B_{mn} is then, neglecting constant factors,[6]

$$B_{mn} = A_{mn} Y_1(m, n) Y_2(m, n), \quad (38)$$

where

$$Y_2(m, n) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} T_2(\xi, \eta) \exp -i\pi \left(\frac{m\xi}{a} + \frac{n\eta}{b} \right) d\xi d\eta. \quad (38')$$

[6] The constant factor neglected as compared with equation (35) is $t_0/(4ab)$. The t_0 is the period of image repetitions (or "frame period"). It appears here because the brightness of a single image depends on how quickly it is reproduced.

The quantity Y_2 it will be noted is almost the same, for the receiving aperture, as the Y_1 is in equation (20′) for the sending aperture. Thus, on the normally reproduced component the receiving aperture merely adds whatever filtering action it has to that which has already been caused by the sending aperture.

As noted, in addition to this normal component, the integral (35) exists in general for other values of m' and n' and thus gives rise to extraneous components in the reproduced image. If equation (36) is applied particularly to the usual system of scanning in N lines in which as in equation (12), $v = ub/(Na)$, it becomes

$$m + \frac{n}{N} = m' + \frac{n'}{N}. \tag{39}$$

For values of m' and n' satisfying equation (39), the reproduced component has the complex amplitude (neglecting constant real factors)

$$B_{m'n'} = A_{mn} Y_1(m, n) Y_2(m', n'). \tag{40}$$

Looking back at equation (14) and comparing it with equation (39) it may be seen that these components correspond in indices to the original image components that are confused in the signal to give only one signal component. The result is, therefore, after all quite reasonable from a physical point of view. For when a signal of a certain frequency is transmitted over the line the receiving apparatus has no information by which to judge which component in the original picture it is supposed to represent. So, as shown by equation (40) it impartially reproduces every one of the components it could possibly represent, each component with the intensity and phase it would have if it were really the one intended to be represented by the signal. The components are then all superimposed in the picture.

From this development it is clear that the process of scanning an image field in strips and reproducing it in a similar manner not only reproduces the components of the original image but also introduces extraneous components. The reproduced field thus consists of two superposed fields: a normal image built up from the normally reproduced components, and an additional field of extraneous components. Although not really independent, it is convenient to consider these two fields as existing separately, and thus to think of the normal image field as having an extraneous field superposed on it.

Considering the normal field alone, we may term the reproduction of its detail as the *reproduction of* NORMAL *detail.* There is a loss in such reproduction, for both the transmitting and receiving apertures introduce a relative loss in the reproduction of the shorter wave components. Consequently there is a loss of definition in the finer grained details of the normal image. This type of distortion due to aperture loss may be termed *simple omission of detail.*

In addition to the simple omission of detail, the normal image is masked by the presence of the extraneous field. The more pronounced features of this field are the line structure and serrated edges that it superposes on the normal image. Its presence is not only displeasing, but it also masks the normal image components and thus results in a further loss of useful detail. This type of loss may be termed a *masking of detail* or a *masking loss.* It is true that the extraneous components may sometimes give rise to an illusory increase in resolution across the direction of scanning in special cases where they add on to the diminished normal components in just the right phase and magnitude to bring the latter back to their phases and intensities in the original image, giving no resultant distortion whatever. (In all such cases, however, to obtain this benefit it is necessary to effect a quite accurate register between the original image and the scanning lines or the distortion is very large. Such accurate registering is generally impractical and may be definitely impossible if the registry required for one portion of the image conflicts with that required in another portion. Such cases may, therefore, in general be disregarded.)

The Reproduction of Normal Detail

The preceding theory permits a numerical calculation of the reproduction of detail in the normal image. This is given directly by equation (38) above.

In order to make some of the discussion in the following pages more concrete and specific the sending and receiving apertures will be taken alike; this condition, therefore, gives $[Y(m, n)]^2$ as a measure of how well the various components are reproduced. If a picture be assumed in which all the original components have the same amplitude then $[Y(m, n)]^2$ is the amplitude of the reproduced normal components.

The relative admittance for any given pair of apertures may be calculated from equations (20′) or (38′). Such calculations have been made for various apertures and the results summarized in Appendix II.

The admittance of an aperture is not in general uniquely determined by the wave length of a component, but also depends on the orientation of the component with respect to the aperture. The admittances of reasonably shaped apertures do, however, decrease in general with increasing numerical values of the indices m and n; and the shorter wave components are, therefore, in general, less faithfully reproduced than the longer wave ones.

A circular aperture furnishes a simple example of such reproduction—because its admittance, from its symmetrical shape, is a unique function of the wave length of a component. In other words a circular aperture reproduces normal detail equally well in all directions. We may, therefore, simply plot $[Y(m, n)]^2$ as a function of the component wave length as in Fig. 16, and this single curve is a measure of how

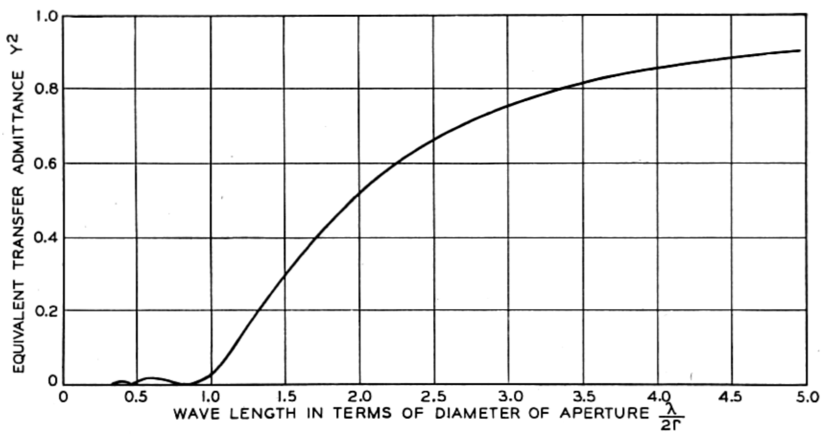


Fig. 16—Equivalent transfer admittance for circular apertures at both sending and receiving ends, *vs.* wavelength.

well the various normal components are reproduced. The shorter wave components are practically omitted in the reproduction of an image.

Other apertures do not reproduce normal image detail equally well in all directions because their admittances depend on the slope of a component. To simplify the consideration of such apertures we may resort to a practice commonly used in discussing telephotographic or television systems, and that is, we may take the resolution along the direction of scanning and across the direction of scanning separately as criteria of their performance.

Neglecting the small slope of scanning lines with respect to the x axis of the image field, the admittance of an aperture for components normal to the direction of scanning is $Y(m, 0)$. Consequently, we may take $[Y(m, 0)]^2$ as a measure of the reproduction of normal detail along the direction of scanning. In a similar manner we may take $[Y(0, n)]^2$ as a measure of the reproduction of normal detail across the direction of scanning.

It thus follows that an aperture gives the same resolution of normal detail along the direction of scanning and across the direction of scanning when the two admittances $Y(m, 0)$ and $Y(0, n)$ are substantially equal for components of the same wave length over the useful range. Circular apertures, square apertures and other apertures that are suitably symmetrical fulfill this condition exactly, and consequently give equal resolution of normal detail in the two directions.

The curve $[Y(0, n)]^2$ has been plotted, by way of illustration for a rectangular aperture, in the middle line of Fig. 17.

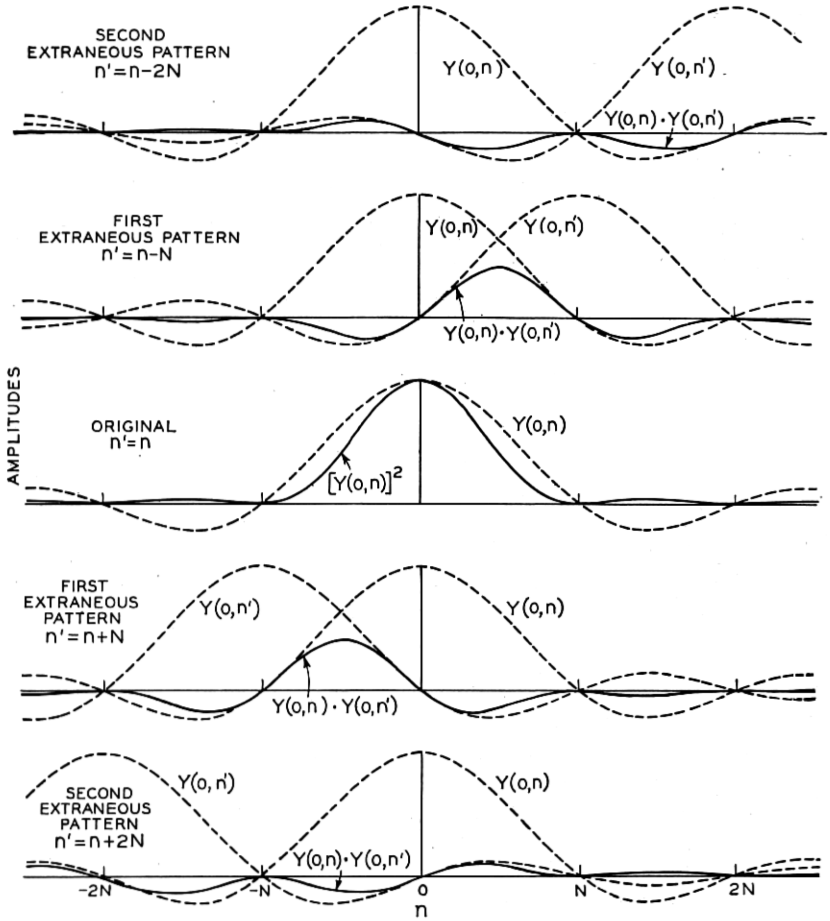


Fig. 17—Reproduction of original and extraneous patterns.

It may be noted incidentally that the simple omission of detail which occurs in the reproduction of normal components is quite similar to the loss of resolution that an image suffers when it is reproduced through an imperfect optical system. Specifically the effect of a sending or receiving circular aperture alone, or $Y(m, n)$, is the same as that caused by an optical system which reproduces a mathematical point in the original as a circle of uniform illumination (circle of confusion) in the image of the same size (with respect to the image) as the scanning aperture. The effect of the two apertures in tandem, or $[Y(m, n)]^2$, may be very closely simulated by a circle of confusion of about twice the area of either aperture, as can be judged from the discussion which has been given above regarding the curves in Fig. 12.

The Extraneous Components

It will be clearly understood that the discussion immediately preceding has been confined entirely to the normal image components, that is, to the image that would be seen if no extraneous components were present. In particular, it should be clear that the reproduction of normal detail equally well in the direction of scanning and across the direction of scanning does not mean that the details of the total resultant image will be seen equally well in the two directions, for the extraneous components will to a certain extent mask the normal image.

In the same manner as for the normally reproduced components, the amplitudes of the extraneous components, according to the preceding theory, are given by equation (40) above, where

$$m' = m + \mu,$$
$$n' = n - \mu N,$$

where

$$\mu = \text{an integer} = m' - m = (1/N)(n - n'). \tag{41}$$

The composite transfer admittance $Y(m, n) \cdot Y(m', n')$ may therefore be taken as a measure of the extent to which the extraneous components are introduced. If a picture be assumed in which all the original components have the same amplitude then $Y(m, n) \cdot Y(m', n')$ is the amplitude of the extraneous components.

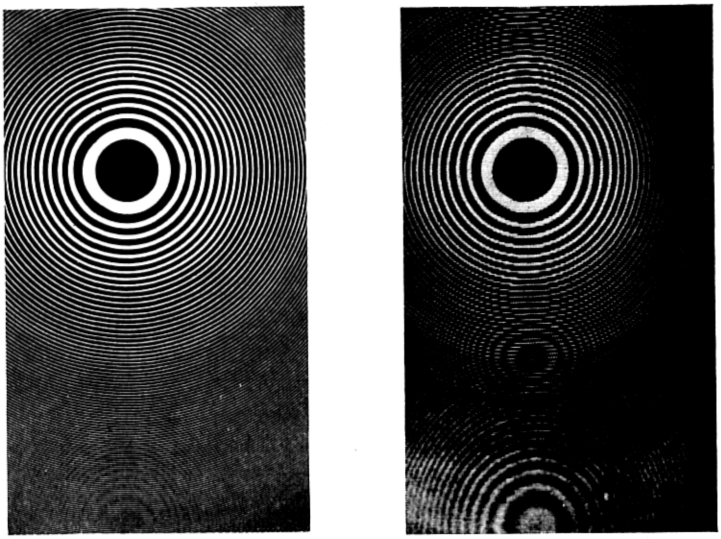
A given original component of indices m, n gives rise to a whole series of extraneous components, m', n', as μ ranges from 1 up through the positive integers and -1 down through the negative integers. As an illustration we have plotted the case of a rectangular aperture of a width just equal to the scanning pitch, in Fig. 17, which has just been referred to in considering the normal components. The two lines marked "first extraneous pattern" show the relative amplitudes for μ equal to 1 and -1, respectively, and those marked "second extraneous pattern," for μ equal to 2 and -2, respectively, for $m = 0$. (The shift from $m' = 0$ to $m' = \pm 1$ and ± 2 has been ignored since if N is at all large this has a negligible effect on $Y(m', n')$, as may be noted from Fig. 13.) An examination of Fig. 17 and a consideration of the nature of $Y(m, n)$ and $Y(m', n')$ shows that the principal interference effect will come from the pattern for which $|\mu| = 1$, and that the relative amplitudes become very small as μ increases in absolute magnitude. In general, therefore, only the first extraneous pattern may be considered as of really serious importance. Considering this pattern in Fig. 17 it will be seen that the amplitude $Y(0, n)$ $Y(0, n')$ increases as $|n'|$ increases from zero, the extraneous components becoming more and more comparable to the normal components. At $N/2$, both components are of the same amplitude, and the extraneous components are therefore masking the normal components. It will be noted that the index region at $N/2$ corresponds to the centers of the relatively empty regions in the frequency spectrum of the signal. The large masking effect caused by the extraneous components explains why such small signal energy as exists in these regions is almost completely incapable of transmitting any useful image detail.

It will be noted that the components with values of $|n'|$ in the neighborhood of $N/2$ and greater are in general almost parallel to the direction of scanning. The masking loss will therefore be greatest across the direction of scanning and practically negligible along the direction of scanning. This is quite reasonable because the extraneous components constitute the line structure of the reproduced image, and should therefore cause the greatest loss of detail across the direction of scanning.

For clarity in the explanation up to this point, masking loss has been discussed as if an extraneous component could only mask the normal component with which its indices happened to coincide. In reality the masking is of a more serious nature. An extraneous component undoubtedly obscures any normal component that has about the same wave length and the same slope across the field even though it does not exactly coincide in these characteristics.

More detailed curves than Fig. 17, showing the amplitudes of the extraneous components have been prepared in Appendix II. These also show the results for other index values of m than zero, and for other than the simple rectangular aperture. The results indicate that the extraneous patterns diminish in intensity progressively as more overlap is tolerated between adjacent scanning lines, at the expense, of course, of increased aperture loss for the normal components. This point will be taken up again below.

The reality of these extraneous components is strikingly demonstrated in Fig. 18, for which we are indebted to Mr. E. F. Kingsbury.



a. Original. *b*. Transmitted.

Fig. 18—Fresnel zone plate.

This shows at (*a*) the original of a Fresnel zone plate and at (*b*) the picture after transmission through a telephotographic system. The first extraneous pattern is very prominent in the lower corner of (*b*) and a detailed study of the slope and spacing of the extraneous striations shows them to be in exact accord with the theory which has been given.[7]

The special case of the extraneous components which are formed when the original consists of a flat field is of some interest due to the high visibility of these components under such a condition. This scanning line structure is quite familiar as an imperfection in many pictures

[7] The extraneous pattern, although it is (and should be according to the theory) very nearly a transposed reproduction of the original pattern, must not be confused with a long delayed echo of that original pattern. In other words, if only the lower half instead of the whole of (*a*) had been transmitted, the lower half of (*b*) would still have been exactly as it is, the extraneous components being generated entirely irrespective of whether components representing a similar configuration exist in other portions of the original or not.

In the region about half-way between the centers of the normal picture and the first extraneous picture the resulting pattern gives very much the appearance of another set of extraneous components. It is not such, however, that successive rings are not really bright and dark, as they would be in the case of a genuine extraneous component, but alternating uniform gray and striped black and white, so that the average intensity along the circumference of a ring is independent of the diameter of the ring, except for some photographic non-linearity.

transmitted by telephotography and television. It can be removed only by insuring that $Y(m', n')$ shall vanish whenever $n' = N$, so that

$$Y(0, 0) \cdot Y(\mu, \mu N) = 0.$$

The requirement can be met for the elementary shapes of apertures A, E and F of Fig. 12, but cannot be met in the others. In these other cases the overlap between adjacent scanning lines is usually adjusted so that the requirement is met for $\mu = \pm 1$, to remove the most serious pattern. Thus, for example, for the circular aperture B this requires an overlap of around 25 per cent.

The Reproduction of Detail

In optical instruments the reproduction of detail is usually measured by what is called the "resolving power" which in turn is defined from the smallest separation between two mathematical point (or parallel line) sources of light in the original which can be distinguished as double in the reproduced image.

For the present it is perhaps simpler to consider another criterion of the resolving power, namely, the shortest element length in an image resembling a telegraph signal, used as an original, which can be recognizably reproduced with certainty in the received picture. For reasons that have already been mentioned above it is necessary to insist that the received picture be recognizable with certainty without any registry requirement between the original image and the scanning lines.

Using this criterion for the resolution along the direction of scanning and assuming the apertures at the sending and receiving ends to be rectangular and of the same length with respect to the picture size, the minimum signal element required for a recognizable picture (as set by the apertures as distinguished from the electrical transmission circuits) will be of about the length of either aperture. For other shapes of aperture the minimum element length will be very nearly the length of the equivalent rectangular aperture using the term "equivalent" in the same sense that it was used in the discussion regarding Fig. 12.

According to the same criterion, for the resolution across the direction of scanning the minimum element length required for recognizable transmission, in the case of a rectangular aperture of width equal to the scanning pitch, will be twice the scanning pitch. It will be noted that this is twice the length which would be required if only the normal image components were reproduced, and this difference may be considered as a measure of the degradation caused by the masking effect of the extraneous components for this arrangement of apertures.

This figure for the degradation must be taken with a certain reserve, partly because the exact telegraph theory for the criterion of resolution considered has really been inferred rather than presented in complete logical form, and partly because the figure may be expected to vary according to the criterion of resolution chosen. Some rough studies have indicated the degradation to be materially less if the more conventional criterion of resolution (two parallel line sources of light) were used.

This degradation may be estimated in another manner. In Appendix II the extraneous components have been computed for a variety of apertures and degrees of overlap between adjacent scanning lines. In Fig. 19 there have been plotted the maximum amplitudes of these

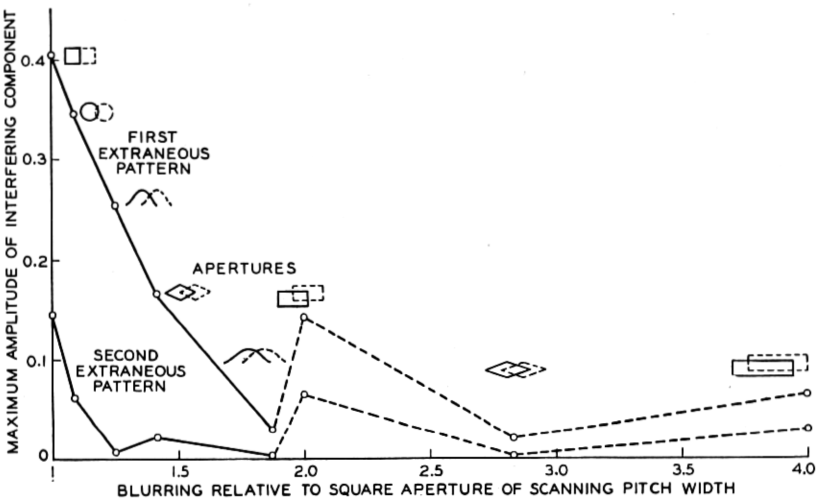


Fig. 19—Magnitude of extraneous components as a function of resolution.

extraneous components in each case (the first and second extraneous patterns being plotted separately) as a function of the relative coarseness of resolution for the normal image alone. This latter quantity is taken relative to a rectangular aperture of width equal to the scanning pitch, and, for example, for a rectangular aperture of width equal to twice the scanning pitch, is represented by the figure 2. For convenience, above the various points have been inserted small diagrammatic representations of the corresponding apertures. Also for convenience the points have been arbitrarily connected together.

From inspection of Fig. 19 several conclusions may be drawn, namely,

1. Considering apertures of a given shape, the more overlap allowed between adjacent scanning lines the weaker will be the extraneous patterns but the coarser will be the reproducible detail in the normal image.

2. Not all shapes of aperture are equally efficient in suppressing extraneous components, and at the same time retaining a given resolution of normal detail. Of the shapes considered, the rectangular aperture is least efficient in this respect, and the full-wave sinusoidal aperture (E in Fig. 12), is the most efficient.

3. Although not proved, it may be inferred from the figure that the finest resolution in the normal image that can be obtained (assuming a given scanning pitch) without showing a first order extraneous pattern on a flat field, is that obtained with the rectangular aperture of width equal to the scanning pitch.

4. With the most suitable aperture it is possible practically to suppress the extraneous components, at the expense of coarsening the normal reproducible detail to slightly under twice that given by the rectangular aperture just mentioned.

The last point in particular enables us to draw a conclusion in regard to the degradation contributed by the extraneous components. For a rectangular aperture of width equal to the scanning pitch it appears that the degradation amounts to a little less than doubling the coarseness of resolution to normal detail. This substantially checks the estimate which has already been made above. It may further be surmised for all the other shapes of aperture shown with a value of abscissa under 2 that as the degradation contributed by the extraneous components is reduced, the coarseness of resolution to normal detail is increased to just about make up for this, and that in the overall picture the minimum element length which can be recognizably reproduced remains substantially constant at about twice the scanning pitch.[8] For aperture arrangements with values of abscissa over 2, either the inefficiency in suppressing extraneous components, or the unnecessarily large overlap, tends to coarsen the overall resolution to a minimum elementary length greater than twice the scanning pitch. In this region the line connecting the points has been dotted.

[8] It may very well be that even if all these aperture arrangements transmit an about equal amount of information they do not give the same psychological satisfaction to the viewer at the receiving end. The general effect of a square aperture of scanning pitch width is to give a "snappy" appearance, disturbed, however, by the presence of the extraneous patterns. When these are removed, keeping the overall resolution about the same, the appearance becomes "woolly" or "fuzzy."

An Estimate of the Idle Frequency Regions

As mentioned at the beginning of this paper, the frequency regions between the strong bands appear to be empty when examined with a frequency analyzer of limited level range, or when a narrow band elimination filter is used in connection with visual observations of the reproduced image. These regions are not really completely empty, but do contain weak signal components as shown by the preceding theory, which are not, however, particularly useful inasmuch as, in the final result, they give rise about equally to components simulating the original picture and to masking extraneous components. The regions may, therefore, be considered as idle.

The factors determining the extent of these idle regions are too complicated to permit an exact theoretical evaluation of their width, but an estimate may be attempted from an inspection of Fig. 12 and of the curves given in Appendix II.

From Fig. 12 and the experience that along the direction of scanning the minimum recognizably transmitted elementary signal length is the length of a rectangular aperture it can be deduced that in the absence of extraneous components the useful band of an aperture extends up to the point where its relative admittance, for a single aperture, is in the neighborhood of 0.65. For two apertures in tandem the corresponding relative admittance is $0.65^2 = 0.42$.

Now in Fig. 27 of Appendix II the extraneous components are very small and may be considered negligible. According to the above criterion, therefore, the useful frequency band constitutes approximately 54 per cent of the total space. The idle frequency regions would, therefore, occupy the remainder, or 46 per cent of the total space.

Experimental examination of a television signal with a narrow band elimination filter gave the width of the idle regions as 50 to 60 per cent of the total space. This was for a field scanned with a circular aperture giving a one-quarter overlap of scanning strips. The discrepancy for a quantity so vaguely defined is not large but is probably due to incomplete utilization of even the theoretically active region by the television set because of inherent imperfections in parts of the complete system outside the scanning mechanism proper.

The width of the individual idle bands is then about half the frequency of repetition of scanning lines. For most systems of telephotography this runs in the order of magnitude of one cycle per second, making the waste regions very narrow and close together. For systems of television the waste bands come in much more significant "slices," although the same fraction of the frequency space is wasted. For example, in a 50-line system the waste bands are each about 500 cycles wide. In a system using a single sideband of one million cycles width the waste bands are each about 3300 cycles wide.

These idle frequency regions naturally lead to the questions whether (*a*) there is any way of segregating all of the relatively useless signal components in one region of the frequency spectrum so that the useful parts of the signal may be transmitted over a channel of about half the width, or (*b*) whether it would be worth while placing other communication channels in these waste regions. It must be realized, however, that even when the complete frequency space is utilized (by any one of a number of possible schemes), the required frequency band for transmitting a picture of given detail at a given rate is still only halved as compared with the simple system considered above, which is not a change in order of magnitude. The problems of transmitting the wide band of frequencies necessary, for example, in television, while lessened, therefore still remain.

APPENDIX I

The calculation of $Y_1(m, n)$ according to equation (20′) is, for the three simple apertures here considered, a straightforward mathematical process which will therefore not be reproduced. The results are plotted in the form of charts in the conventional manner for functions of two variables, namely as a series of contours, one of the two variables being kept constant for each contour. This constant value changes progressively for each successive contour.

The variables are taken as m and n, multiplied by parameters depending on the sizes of the scanning aperture and of the picture. Because of the obvious symmetry of the function, only half of each chart has been drawn. In order to avoid confusion the contours have been dotted when $|m|$ is greater than the first root of $Y_1(m, 0) = 0$. In one case the contour is shown in a dashed line when $|m|$ is equal to this root. Constant factors in the scale of ordinates have been neglected, to make $Y_1(0, 0) = 1$.

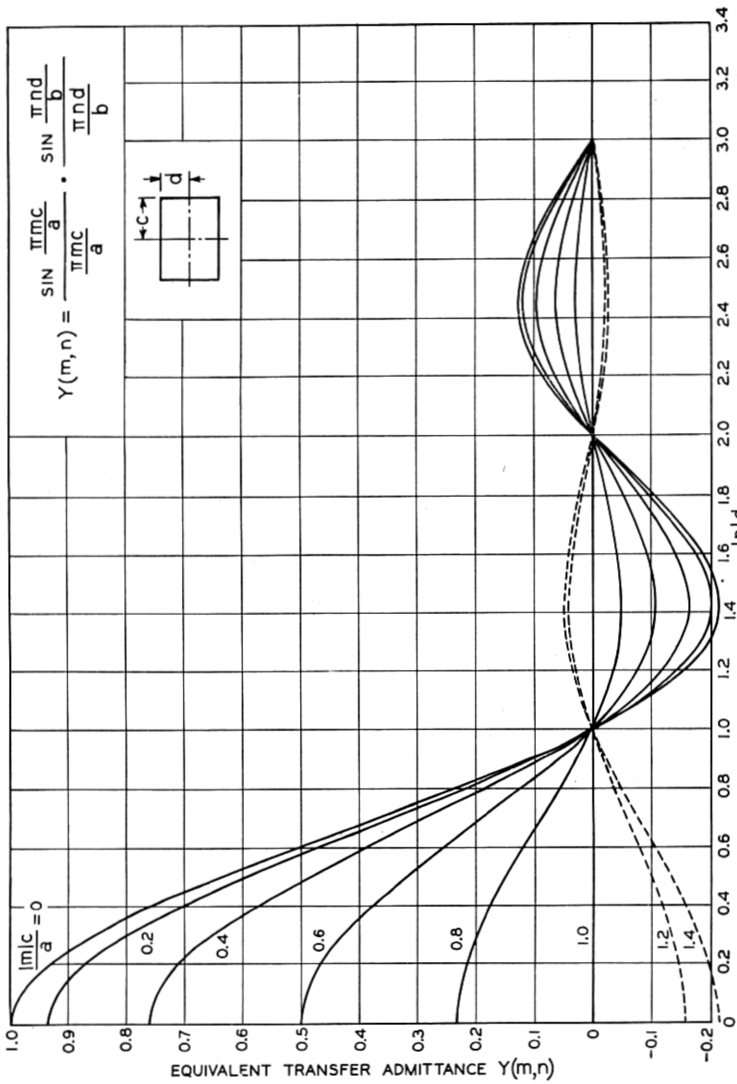


Fig. 20—Rectangular aperture.

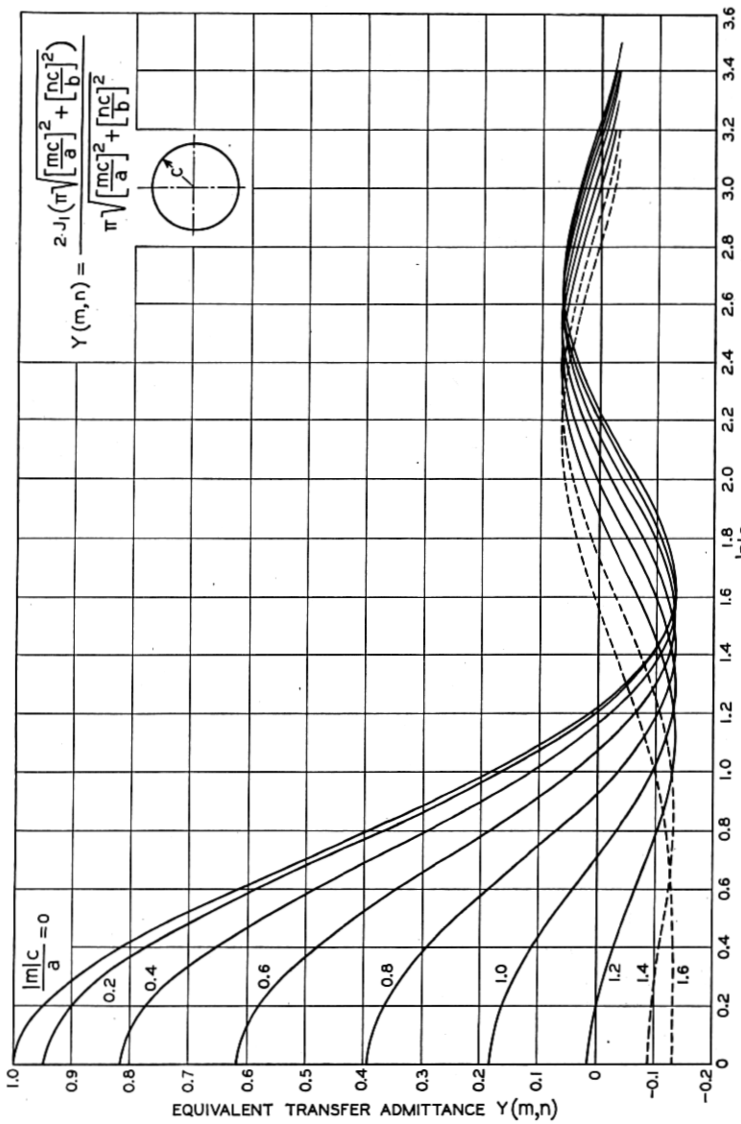


Fig. 21—Circular aperture.

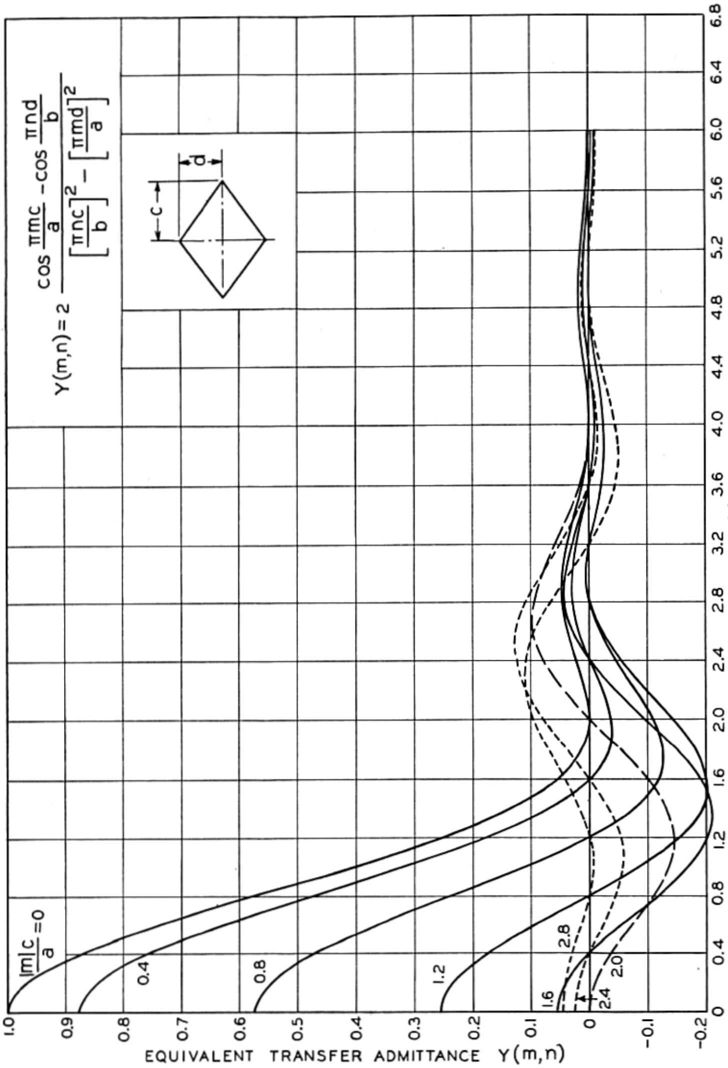


Fig. 22—Diamond shaped aperture.

APPENDIX II

As in the case of Appendix I, the calculation of $Y_1(m, n) \cdot Y_2(m', n')$ is a straightforward mathematical procedure which will not be reproduced. The results are again presented in the form of charts.

In these charts the intensities of the principal extraneous components are indicated by solid lines, while the higher order extraneous components are indicated by dotted lines. The normal components have been indicated by dashed lines.

It should be explained that what has really been plotted is $Y_1(m, n) \cdot Y_2(m, n')$ rather than $Y_1(m, n) \cdot Y_2(m', n')$. This is because the difference between m' and m, when multiplied by the parameters chosen for the charts, varies with the proportions of the scanning system used. As discussed in the text with regard to Fig. 17, however, this difference between m' and m has a negligible effect for any system employing a useful number of scanning lines.

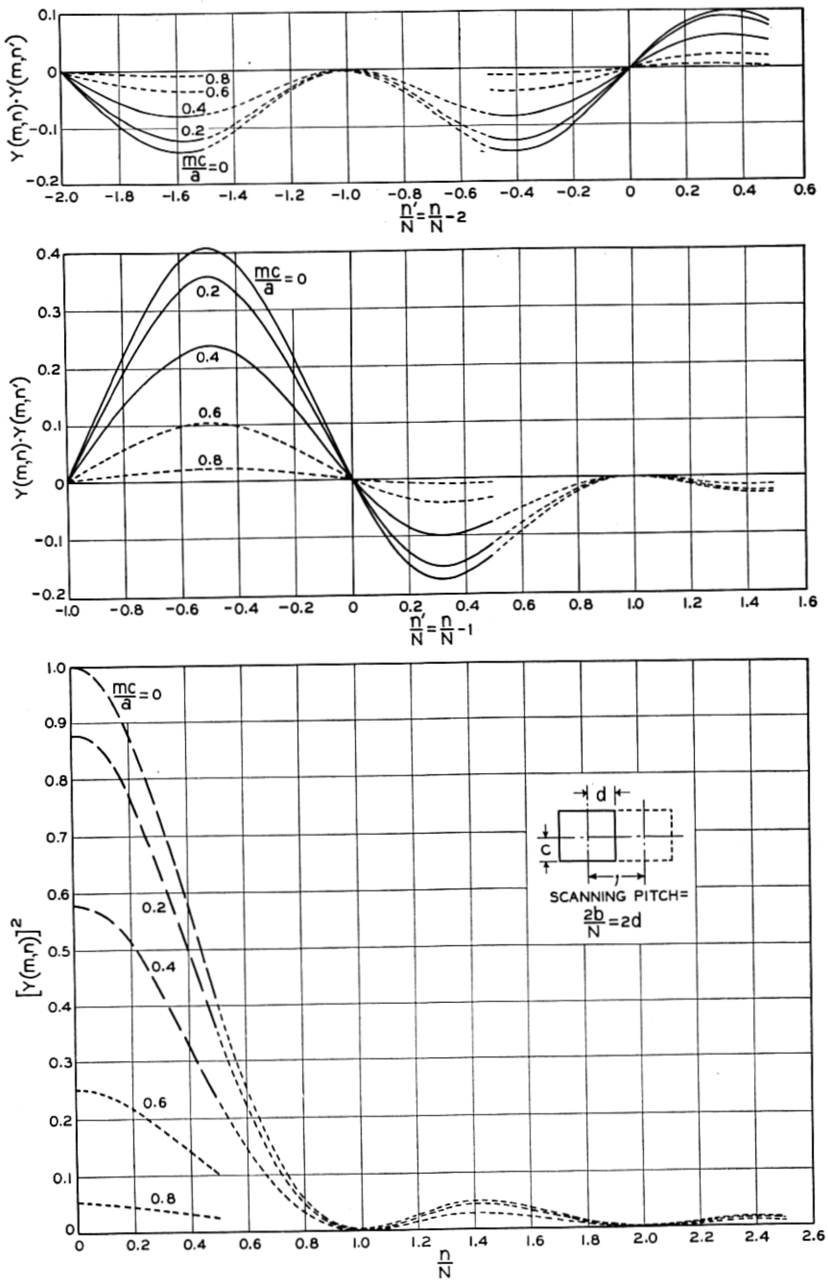


Fig. 23—Rectangular aperture with no overlap.

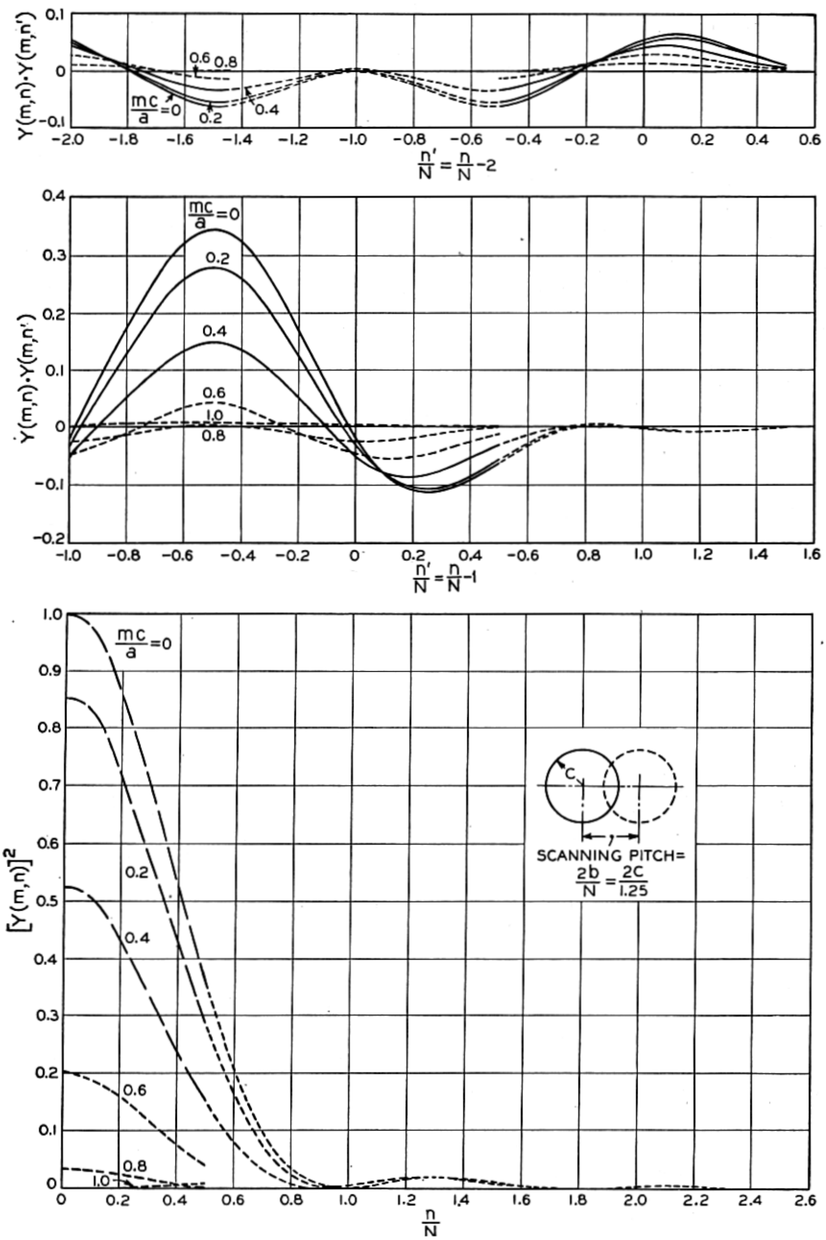


Fig. 24—Circular aperture with 25 per cent overlap.

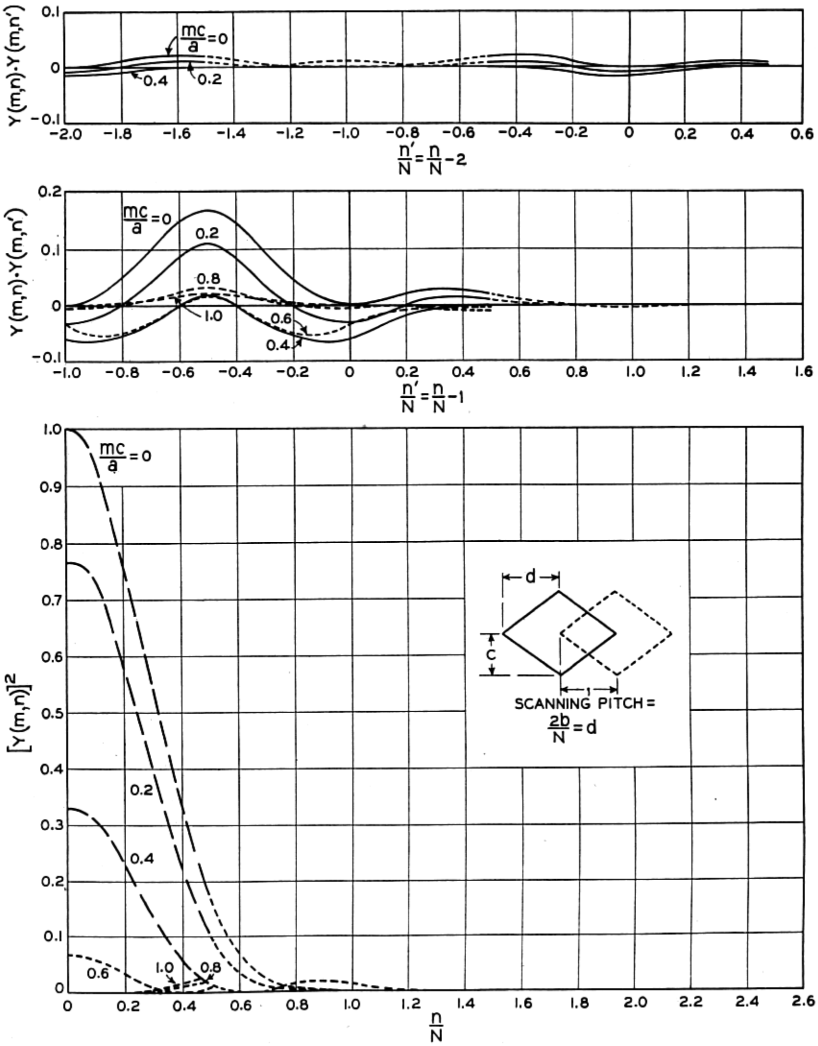


Fig. 25—Diamond shaped aperture with half diagonal overlap.

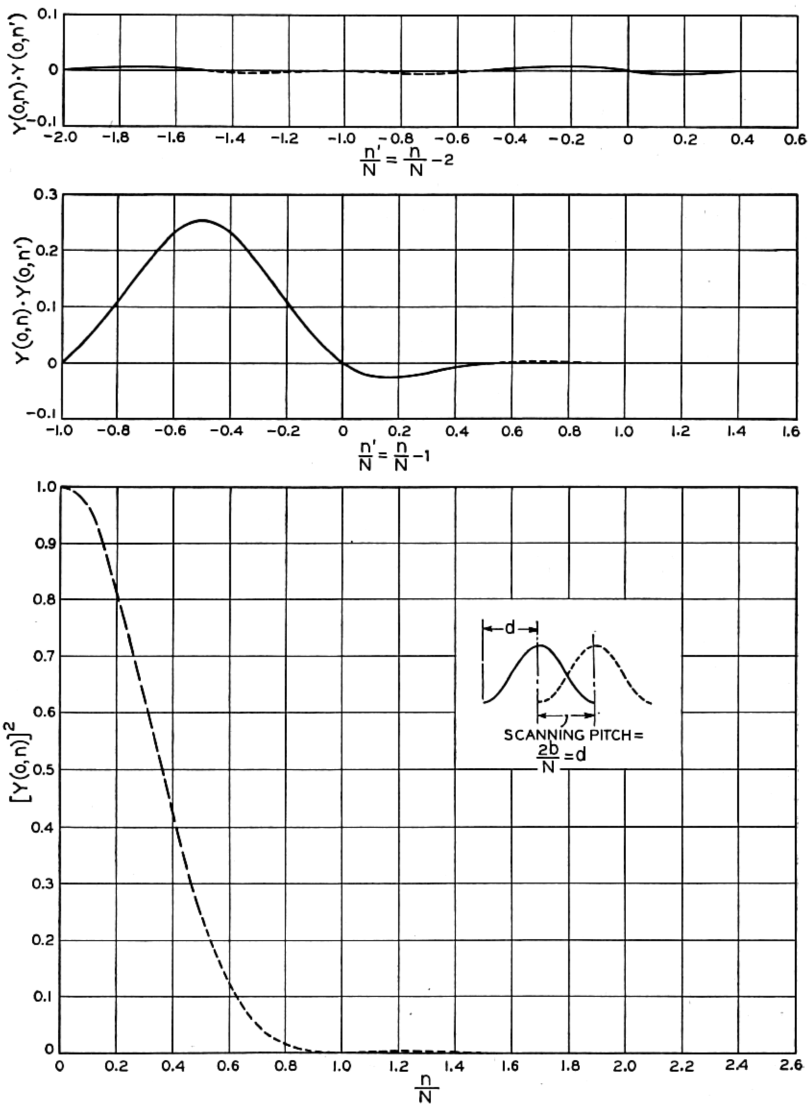


Fig. 26—Sinusoidal aperture with half wavelength overlap.

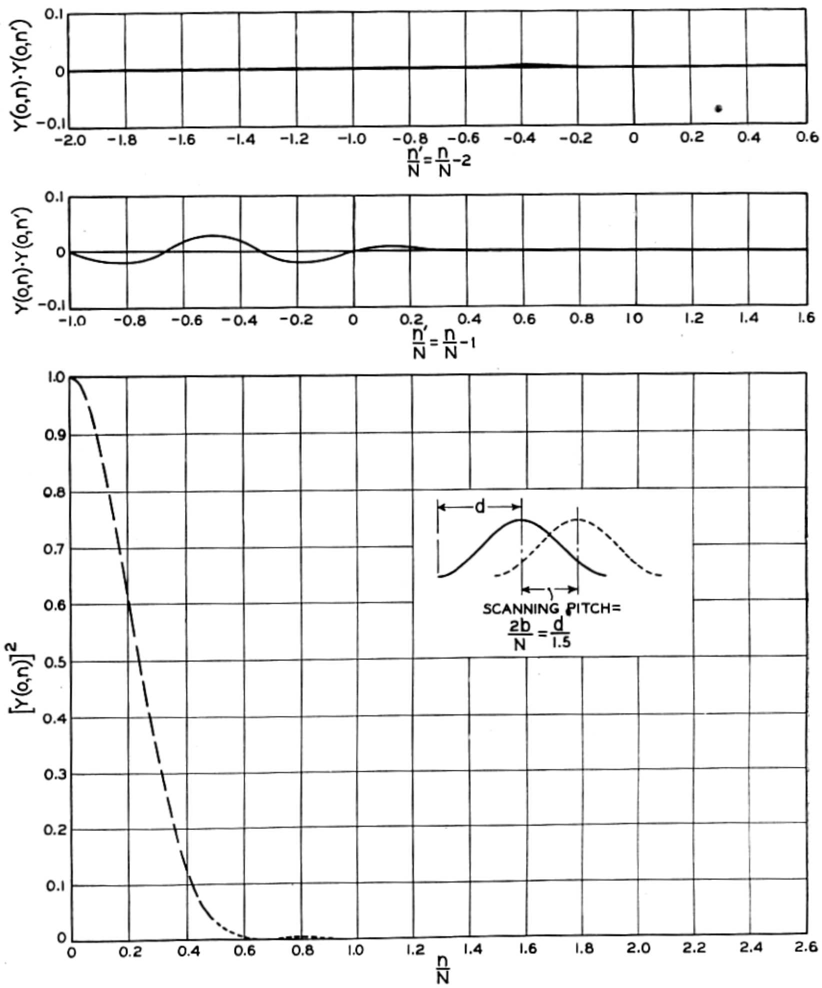


Fig. 27—Sinusoidal aperture with two-thirds wavelength overlap.

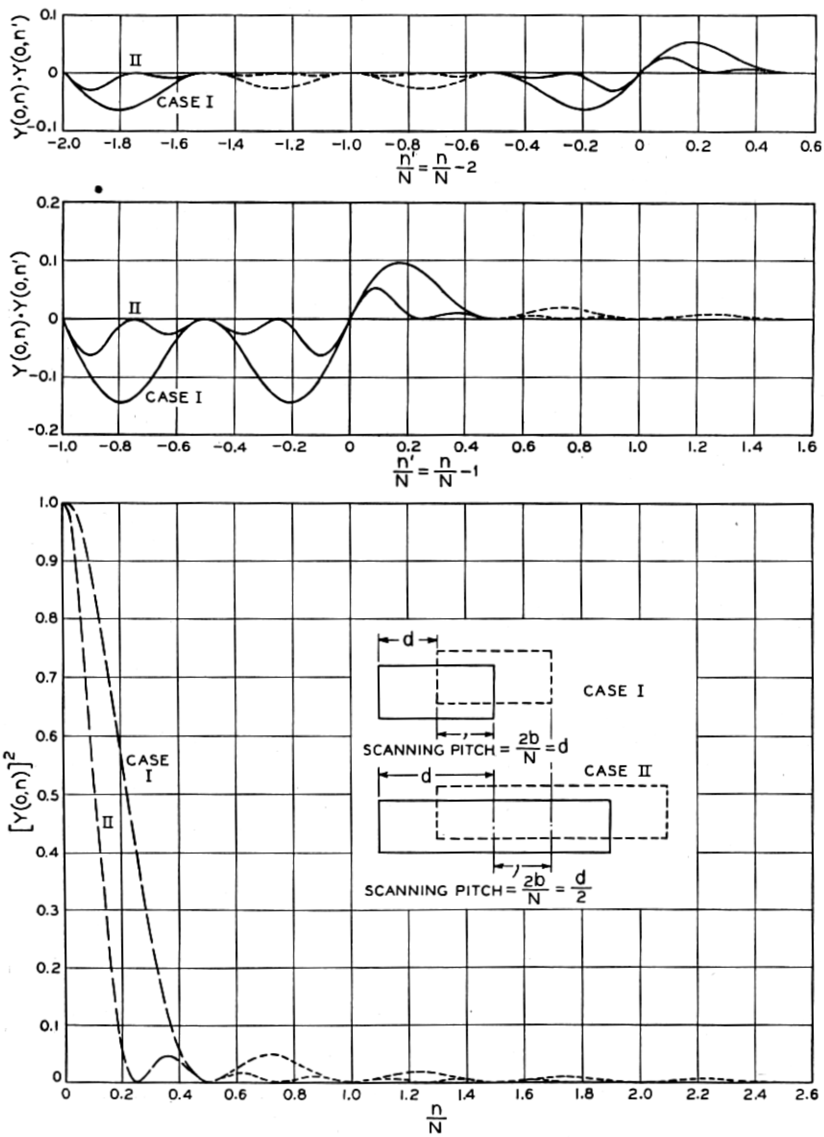


Fig. 28—Rectangular aperture with different degrees of overlap.

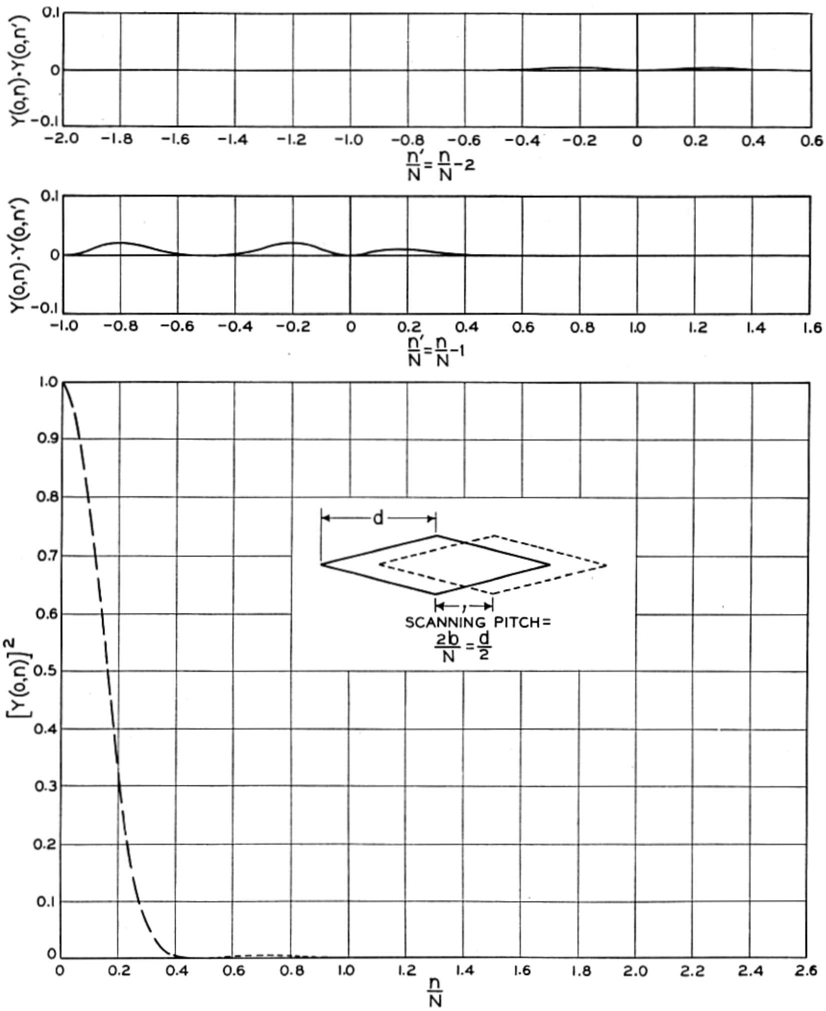


Fig. 29—Diamond shaped aperture with three-quarters diagonal overlap.