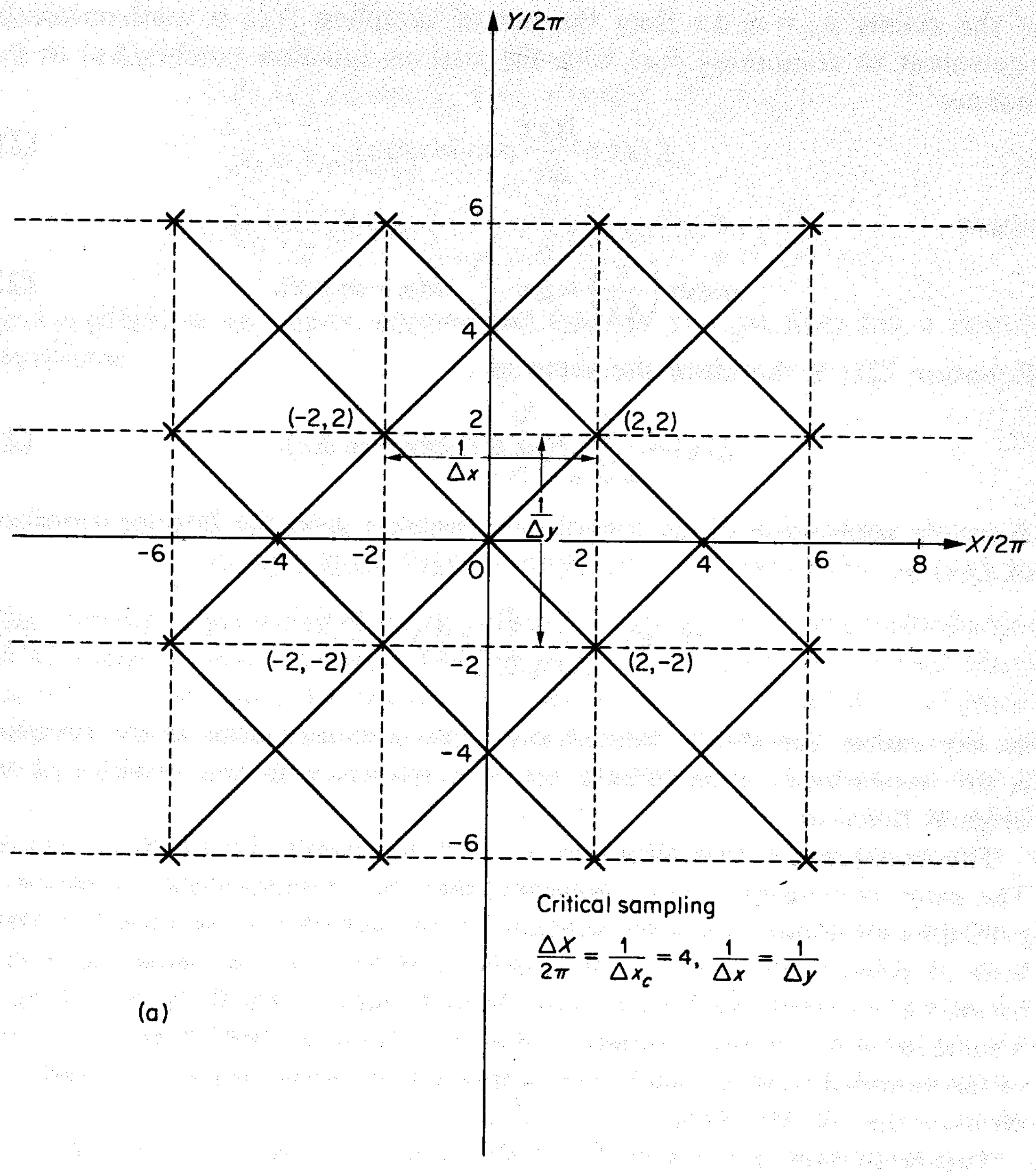

Figure 5. Illustration, using the letter X, of the consequences of using a sampled Fourier transform to reconstruct the original object with an inverse transform. (a) Critical sampling, (b) oversampling, (c) undersampling. Note that $\Delta X \, \Delta x = 2\pi$

The critical sampling rate Δx_c is 2π divided by the reciprocal of the maximum size of the object in the X-direction. This is more formally understandable by considering $F(x, y)$ that has, in general, for $y = 0$, a term in it of the form $\sin[(L/2)x \cos\theta)]\exp(iX_0x)$, where L is the length of a line in the object and X_0 is its mid-point. Such a function involves the products $\sin[(L/2)x \cos\theta]\cos(xX_0)$ and $\sin[(L/2)x \cos\theta]\sin(xX_0)$. This means

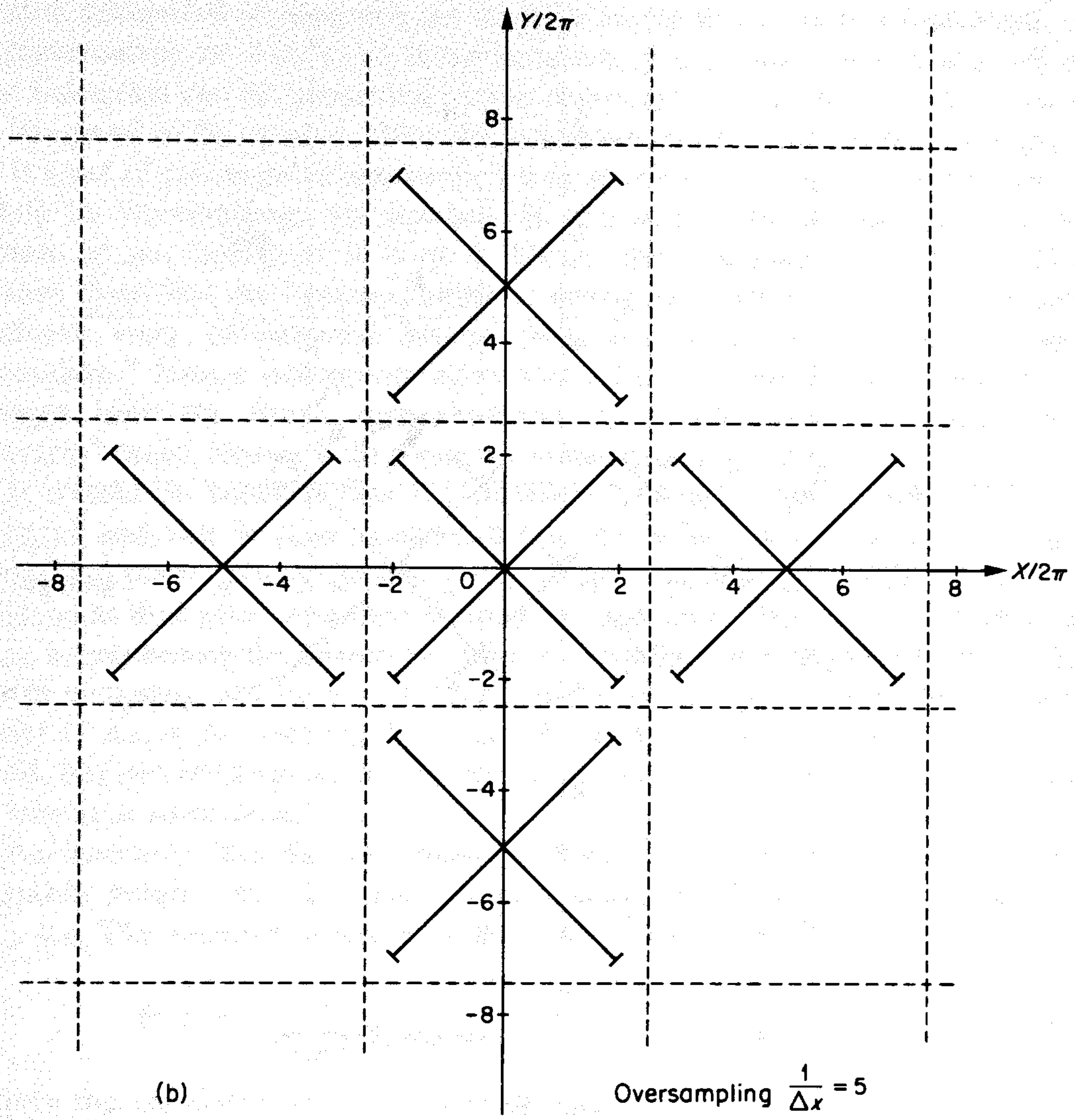

Figure 5b

that, in the x-direction, there are spatial frequencies

$$f_x = \left(X_0 \pm \frac{L}{2} \cos\theta\right) \Big/ 2\pi. \tag{25}$$

Similarly, for the y-direction, there are spatial frequencies

$$f_y = \left(Y_0 \pm \frac{L}{2} \sin\theta\right) \Big/ 2\pi. \tag{26}$$

The spatial frequency bandwidths are, therefore,

$$\Delta f_x = f_x^{(\max)} - f_x^{(\min)} = (L \cos\theta)/2\pi, \tag{27}$$

$$\Delta f_y = f_y^{(\max)} - f_y^{(\min)} = (L \sin\theta)/2\pi. \tag{28}$$

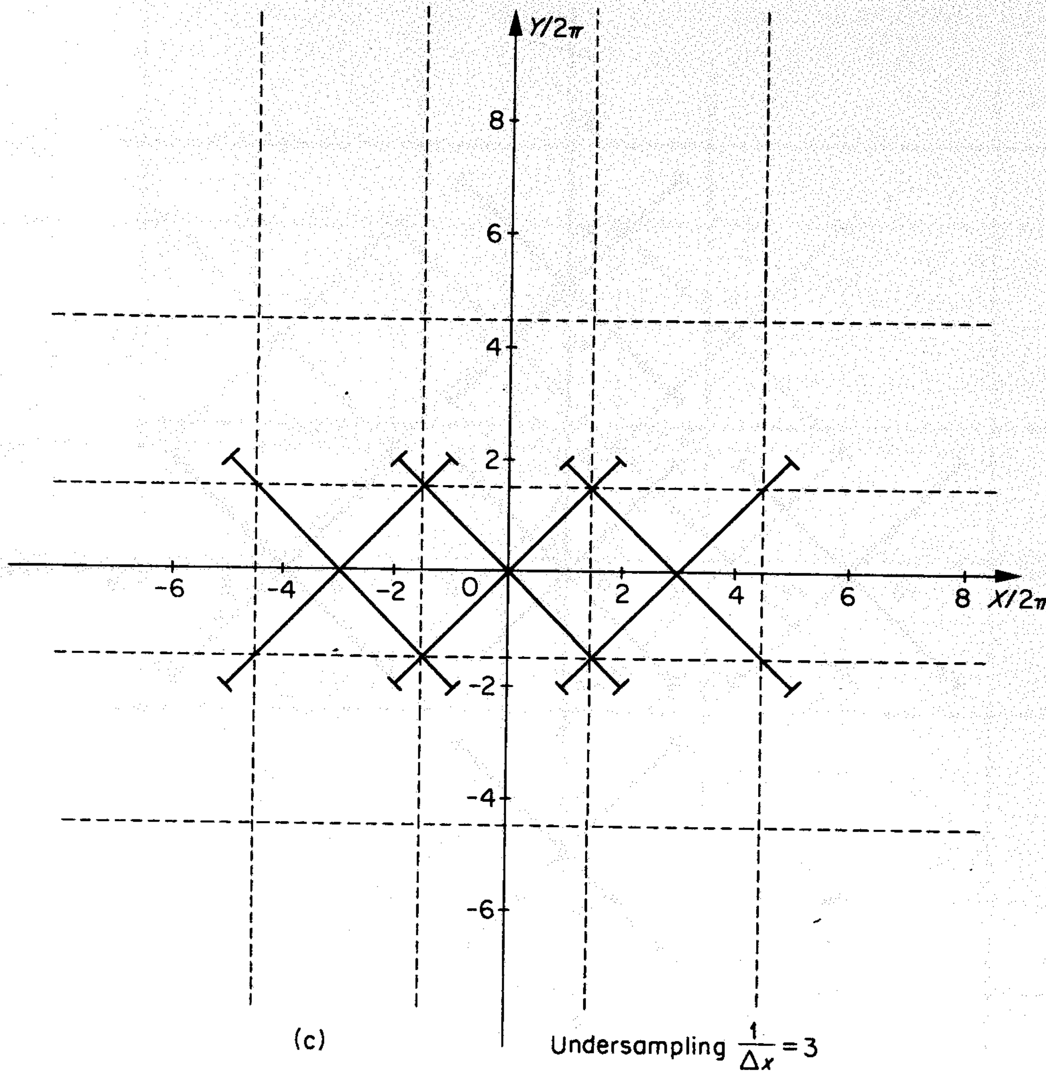

Figure 5c

$2\pi\Delta f_x$ and $2\pi\Delta f_y$ are the projections of the line on the X- or Y-axis and $2\pi f_x^{(\max)}$, $2\pi f_y^{(\max)}$ are the coordinates of the ends of the line. Δf_x and Δf_y are found for a complex object, i.e. a sequence of letters, by taking the difference between the maximum and minimum X-values and the maximum and minimum Y-values. Hence Δx_c is equal to Δf_x^{-1} or Δf_y^{-1}, whichever is the largest. It is evident that if the periodic array is spaced at intervals of $1/\Delta x_c$ then, along X say, $X_{\max} - X_{\min} = 2\pi/\Delta x_c$.

5. BINARY FOURIER TRANSFORM HOLOGRAMS

5.1 Representation of $F(x, y)$ at sampling points

The Fourier transform $F(x, y)$ is a complex quantity of the form

$$F(x, y) = A(x, y)e^{-i\alpha(x, y)}, \qquad (29)$$

that is computed at intervals Δx and Δy along the x- and y-directions. A representation of $F(x, y)$ at these sampling points must be devised before the hologram can be generated by the computer. If, initially, the (x, y) plane is imagined to be opaque, then the hologram can be constructed by making in it a set of rectangular apertures. Each aperture is completely transparent, while its surroundings are opaque. In this sense, the transmittance at all points of the hologram is 0 or 1. Hence the holograms are binary. This makes them like the familiar, halftone newspaper pictures. Continuous-tone pictures were investigated briefly[8] but, as pointed out by Brown and Lohmann,[9] binary holograms allow more light to reach the reconstructed images than the usual, thin emulsion, grey holograms. This and other features[9] make binary holograms an attractive proposition.

It should be recalled that the problem in hand is the representation of $F_s(x, y)$ and this is done, essentially, by constructing an aperture at each sampling point, whose height W_{nm} is proportional to the amplitude of $F_s(x, y)$ at that point and then shifting the aperture off-centre, by an amount P_{nm}, to represent the phase of $F_s(x, y)$. This aperture is located within the mnth sampling cell of the xy plane and centred upon the mnth sampling point $(n\,\Delta x, m\,\Delta x)$ (setting $\Delta x = \Delta y$). A variety of aperture shapes could be used, but the rectangular one is extensively discussed in the literature which is why it is used here.

An aperture, like the one shown in Figure 6, has a constant width $c\,\Delta x$, variable height $W_{nm}\,\Delta x$, and variable location along x, within the cell, $P_{nm}\,\Delta x$. The transfer function of the whole hologram is[10]

$$H(x, y) = \sum_n \sum_m \operatorname{rect}\left(\frac{x - [n + P_{nm}]\,\Delta x}{c\,\Delta x}\right) \operatorname{rect}\left(\frac{y - m\,\Delta x}{W_{nm}\,\Delta x}\right), \tag{30}$$

where the standard rectangle function

$$\begin{aligned} \operatorname{rect}(x) &= 1 \quad 0 \leqslant |x| \leqslant \tfrac{1}{2} \\ &\ 0 \quad \tfrac{1}{2} < |x| < \infty, \end{aligned} \tag{31}$$

implies that

$$\left| \frac{x - (n + P_{nm})\,\Delta x}{c\,\Delta x} \right| \leqslant \tfrac{1}{2}, \tag{32}$$

$$\left| \frac{y - m\,\Delta x}{W_{nm}\,\Delta x} \right| \leqslant \tfrac{1}{2}, \tag{33}$$

and hence, explicitly,

$$-c\frac{\Delta x}{2} + (n + P_{nm})\,\Delta x \leqslant x \leqslant c\frac{\Delta x}{2} + (n + P_{nm})\,\Delta x, \tag{34}$$

$$-\frac{W_{nm}}{2}\,\Delta x + m\,\Delta x \leqslant y \leqslant \frac{W_{nm}}{2}\,\Delta x + m\,\Delta x, \tag{35}$$

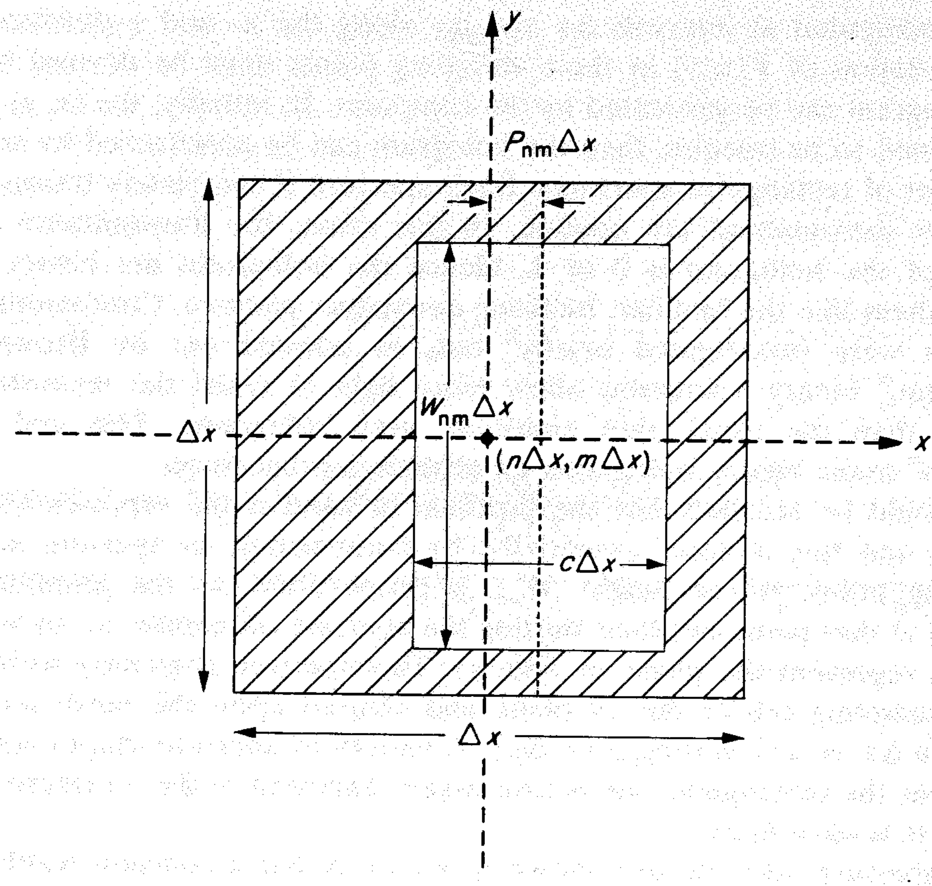

Figure 6. An aperture in a cell in (x, y) space. The area of the cell is $(\Delta x)^2$ and it is centred on the point $(n\,\Delta x, m\,\Delta x)$. The aperture is offset from this point by an amount $P_{nm}\,\Delta x$ but is symmetric about the x-axis

showing that measurements may just as well be made, with respect to the cell centre $m\,\Delta x$, $n\,\Delta x$, in the coordinate system x', y' where

$$\left(P_{nm}-\frac{c}{2}\right)\Delta x\leqslant x'\leqslant\left(P_{nm}+\frac{c}{2}\right)\Delta x, \tag{36}$$

$$-\frac{W_{nm}}{2}\Delta x\leqslant y'\leqslant\frac{W_{nm}}{2}\Delta x. \tag{37}$$

Suppose that a basic reconstruction system like the one shown in Figure 7 is used.[1,10] An off-axis plane wave $\exp(-\mathrm{i}xS)$ from a point source S illuminates the hologram. This choice, incidentally, allows a phase variation along x, and this is consistent with encoding the phase in the hologram by means of shifts P_{mn} in the x-direction only. The amplitude at the hologram plane is $\exp(-\mathrm{i}xS)\,H(x, y)$ so that $G(X, Y)$, the inverse Fourier transform, appears

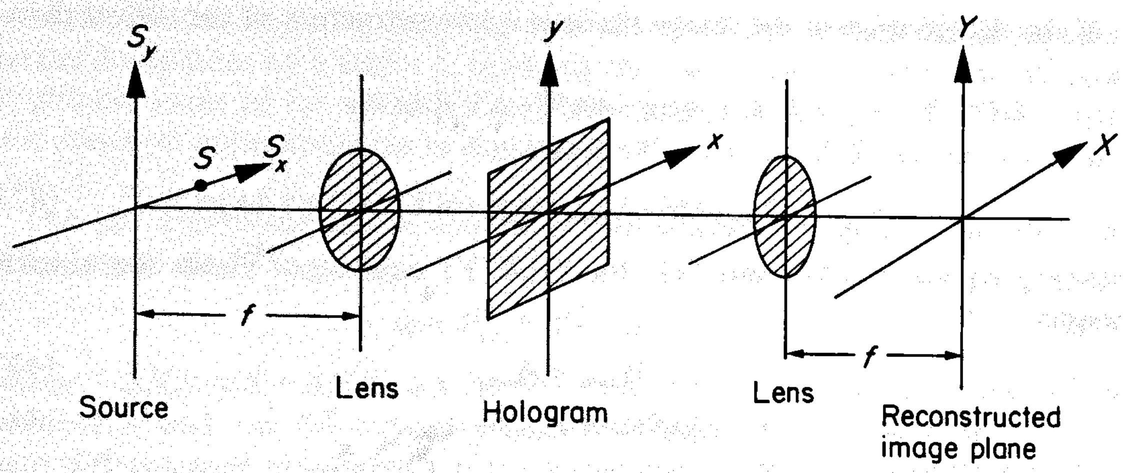

Figure 7. Optical reconstruction system used in the theory of section 5

at the image plane, i.e.

$$G(X, Y) = \sum_n \sum_m \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \mathrm{e}^{-\mathrm{i}(xX+yY)} \mathrm{e}^{-\mathrm{i}xS} \, \mathrm{d}x \, \mathrm{d}y, \tag{38}$$

where use is made of equations (30) to (35). Here, $x_{\min}$, $x_{\max}$ and $y_{\min}$, $y_{\max}$ are, in fact, given by equations (34) and (35). Since $G(X, Y)$ is also required to be a reconstruction of the object, a prescription for identifying W_{nm} and P_{nm} therefore suggests itself.

The integrations in equation (38) are quite easy to perform and lead to

$$G(X, Y) = c(\Delta x)^2 \operatorname{sinc}\left(c\frac{\Delta x}{2}[X+S]\right) \sum_n \sum_m W_{nm} \operatorname{sinc}\left(\frac{W_{nm}}{2}\Delta x Y\right) \times \exp\{-[\mathrm{i}\Delta x(X+S)(n+P_{nm}) + \mathrm{i}mY]\}. \tag{39}$$

This is not a regular series and in order to identify it term by term with a regular Fourier series representation of the object the following approximations[10] ought to be made:

$$\begin{aligned} R = \operatorname{sinc}\left(c\frac{\Delta x}{2}[X+S]\right) &\simeq 1, \\ \operatorname{sinc}\left(\frac{W_{nm}}{2}\Delta x Y\right) &\simeq 1, \\ \exp(-\mathrm{i}XP_{nm}\Delta x) &\simeq 1, \end{aligned} \tag{40}$$

so that

$$G(X, Y) \simeq c(\Delta x)^2 \sum_n \sum_m W_{nm} \exp[-\mathrm{i}\,\Delta x S(n+P_{nm})] \exp[-\mathrm{i}\Delta x(Xn+Ym)]. \tag{41}$$

If the distribution in the image plane is a reconstruction of the object then

$$G(X, Y)=\int\int F(x, y)\exp[-i(xX+yY)]\,dx\,dy$$

$$\approx\sum_n\sum_m A_{nm}\exp(-i\alpha_{nm})\exp[-i(Xn+Ym)\,\Delta x]c(\Delta x)^2. \quad (42)$$

Making equations (41) and (42) for $G(X, Y)$ equivalent yields the simple results

$$W_{nm}=A_{nm} \quad (43)$$

$$(\Delta xS)(n+P_{nm})=\alpha_{nm} \quad (44)$$

to within a multiple $2\pi L_{nm}$, where L_{nm} is an integer.[9] The neglect of L_{nm} can cause cell overlap,[9] but we neglect it here without serious consequences. Hence, in the construction of the hologram, the height W_{nm}, and position P_{nm}, of the aperture in the cell directly record the amplitude and the phase of $F_s(x, y)$ for the sample point $(n\,\Delta x, m\,\Delta x)$. Obviously the aperture should be fairly close to either side of the centre of the cell so that, on average, $\langle P_{nm}\rangle=0$ and $\langle\alpha_{nm}\rangle=0$. Equation (44) can be simplified by setting $\Delta xS=2\pi M$, where M is an integer. This step ensures that

$$e^{i\Delta xSn}=e^{i2\pi Mn}=1$$

so that it can, in effect, be forgotten. Hence the tidier version of equation (44) is

$$P_{nm}=\alpha_{nm}/2\pi M. \quad (45)$$

The brightness[10] of the image at $X=0$, the centre of the object, is proportional to $[\sin(\pi cM)/\pi M]^2$. The latter expression has maxima when

$$|cM|=\tfrac{1}{2}, \tfrac{3}{2}, \ldots, \quad (46)$$

thus making the brightness at $X=0$ proportional to $1/(\pi M)^2$. The brightness varies as M^{-2} so the lowest value $M=1$ must be selected. This choice of M, from (46), makes $c=\tfrac{1}{2}$. The next choice is $c=\tfrac{3}{2}$ which is rejected on the rather obvious grounds that it defines an aperture larger than the cell size. Even $c=\tfrac{1}{2}$ can cause cell overlap but it is not a serious problem.

Extreme values of R, in equation (40), correspond to $X=\pm(\Delta X/2)$ where $\pm(\Delta X/2)$ refers to a reconstructed object lying between $-(\Delta X/2)$ and $\Delta X/2$. Now $\Delta xS=2\pi M$, $cM=\tfrac{1}{2}$, and $\Delta x\,\Delta X=2\pi$ so that R becomes $c\,\mathrm{sinc}[(\pi/2)(1\pm c)]$. Hence, if this factor is not suppressed there is a brightness ratio of $[(1+c)/(1-c)]^2$ from one side of the object to the other. The neglect of it is not very important, because the eye adjusts quite well to its absence, and it is quite easy, if necessary, to incorporate into any computer program. Furthermore, if oversampling is used the ratio is, in any case,

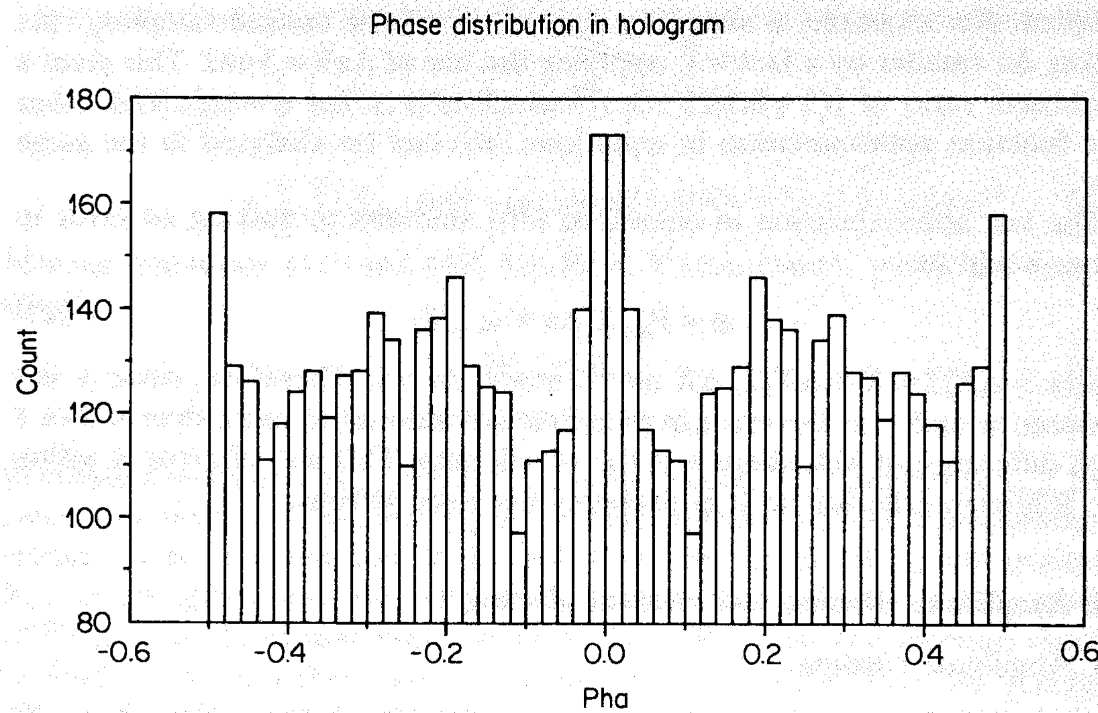

Figure 9. Phase histogram: no random phasing. Note that the phase distribution is symmetrical about the origin

into the low-amplitude range. This feature is a consequence of using letters as an object. It would seem prudent to introduce some form of clipping to chop off the top end of the amplitude range, where so few counts are recorded and then renormalize. This has the effect of making the apertures wider, for the other amplitudes, than they would otherwise be and hence the image field ought to be brighter.

The nature of the clipping used in the computer program is as follows. The original amplitudes are in the range $0 \leqslant A \leqslant 1$ and the number of histogram levels, or channels, is 50 (incidentally this number is easily located in the program and is easily changed). Suppose that by the time some channel number n_L is reached 99 per cent of the amplitudes are counted. A clipping factor $50/n_L$ can then be formed and all the original amplitudes converted to $A' = (50/n_L)A$. This scales up all the amplitudes and effectively makes the original amplitude $A = A_L$, at n_L, equal to unity but permits a very small number of amplitudes to become greater than unity. This is only an apparent difficulty because for any $A' > 1$ the aperture is completely open and so few cells are involved that it does not matter. An examination of Table 1, the contents of which come out on output channel 3, confirms the points discussed above. The critical sampling rate Δx_c and the modified sampling rate actually used is also given in Table 1 as is some information used later concerning the plotting of the hologram. Note from Table 1 and Figure 9 that the phase distribution is symmetrical about the origin.

Table 1. Typical output on channel 3 for the pen plotter with linewidth = 0.5 mm. 13 lines fill the cell. Units = 3.375

SAMPLE RATE = 0.967
MODIFIED SAMPLE RATE = 0.630
DISTRIBUTION OF AMPLITUDE AND PHASE IN HOLOGRAM

AMP	NUMBER	PHASE	NUMBER
0.010	1700.0	-0.49	150.0
0.030	1566.0	-0.47	129.0
0.050	792.0	-0.45	126.0
0.070	604.0	-0.43	111.0
0.090	390.0	-0.41	118.0
0.110	374.0	-0.39	124.0
0.130	260.0	-0.37	128.0
0.150	220.0	-0.35	119.0
0.170	124.0	-0.33	127.0
0.190	78.0	-0.31	128.0
0.210	36.0	-0.29	139.0
0.230	56.0	-0.27	134.0
0.250	36.0	-0.25	110.0
0.270	34.0	-0.23	136.0
0.290	22.0	-0.21	138.0
0.310	30.0	-0.19	146.0
0.330	18.0	-0.17	129.0
0.350	20.0	-0.15	125.0
0.370	8.0	-0.13	124.0
0.390	4.0	-0.11	97.0
0.410	0.0	-0.09	111.0
0.430	0.0	-0.07	113.0
0.450	0.0	-0.05	117.0
0.470	4.0	-0.03	140.0
0.490	8.0	-0.01	173.0
0.510	0.0	0.01	173.0
0.530	4.0	0.03	140.0
0.550	4.0	0.05	117.0
0.570	0.0	0.07	113.0
0.590	0.0	0.09	111.0
0.610	0.0	0.11	97.0
0.630	0.0	0.13	124.0
0.650	0.0	0.15	125.0
0.670	0.0	0.17	129.0
0.690	0.0	0.19	146.0
0.710	0.0	0.21	138.0
0.730	0.0	0.23	136.0
0.750	0.0	0.25	110.0
0.770	0.0	0.27	134.0
0.780	0.0	0.29	139.0
0.810	4.0	0.31	128.0
0.830	0.0	0.33	127.0
0.850	0.0	0.35	119.0
0.870	0.0	0.37	128.0
0.890	0.0	0.39	124.0
0.910	0.0	0.41	118.0
0.930	0.0	0.43	111.0
0.950	0.0	0.45	126.0
0.970	0.0	0.47	129.0
0.990	4.0	0.49	158.0

SUGGESTED CLIP = 2.941TO CHANGE TYPE 1 ELSE 0

(b) *Random phasing*

It will be seen later that the hologram that can be plotted using the theoretical work dealt with up to now has a star-like structure. Furthermore, each line of the object is associated with a high-density region of the hologram. In a typical letter-group object there is a considerable number of horizontal lines and vertical lines so that a significant overlap occurs. If the hologram is produced automatically as a film output from the computer or is graph-plotted and then photographically reduced, very poor image quality can be expected. This troublesome feature would be overcome if it were possible to spread out the hologram (Fourier) amplitudes, over the hologram plane, in some way so as to reduce the area of overlaps. A method of doing this that is very easy to include in a computer program is to multiply the object amplitudes by a random phase factor. After doing this the Fourier amplitude distribution becomes more uniform.

Suppose the object function $G(X, Y)$, is arbitrarily multiplied by a phase factor $\exp(-iaX-ibY)$ then, as we discussed earlier, its Fourier transform shifts to $F(x-a, x-b)$. The introduction of such a phase factor into the function $G(X, Y)$ does not make any physical difference because it is eliminated from the observable $|G(X, Y)|^2$. Since this is the case, then a and b can be quite arbitrarily chosen and an excellent choice would be to make them random numbers. This, in fact, is what is done in the program and they are selected by a random number routine that produces uniformly distributed numbers $0 \leqslant r \leqslant 1$. The random number r is then converted to a set $r' = 2r - 1$ that lies on the interval $-1 \leqslant r' \leqslant 1$ and a fraction of the maximum size of x and y is used to form

$$a = \frac{r'}{4} x_{\max}, \tag{48}$$

$$b = \frac{r'}{4} y_{\max}. \tag{49}$$

The effect of using a and b is to spread out or equalize the amplitudes. This means that, before clipping, there is a greater occupancy of the higher channel numbers as is seen in Table 2 and Figure 10. Note also that now the phase distribution is no longer symmetrical.

5.3 Drawing the hologram

The computer graphics package is used to generate the hologram in the form of a large-scale pen plotter plot that is later reduced photographically or in the form of a directly usable 35 mm microfilm slide.

In order to find W_{nm} and P_{nm} the computer program determines the complex Fourier transform as an amplitude FXY(i, j) and a phase ANGLE

Table 2. Typical output on channel 3 for the microfilm unit with linewidth = 0.2 mm. 33 lines fill the cell. Units = 3.375

SAMPLE RATE = 0.967
MODIFIED SAMPLE RATE = 0.638
DISTRIBUTION OF AMPLITUDE AND PHASE IN HOLOGRAM

AMP	NUMBER	PHASE	NUMBER
0.010	280.0	-0.49	145.0
0.030	636.0	-0.47	140.0
0.050	734.0	-0.45	112.0
0.070	554.0	-0.43	121.0
0.090	455.0	-0.41	131.0
0.110	399.0	-0.39	124.0
0.130	311.0	-0.37	129.0
0.150	278.0	-0.35	107.0
0.170	310.0	-0.33	116.0
0.190	260.0	-0.31	124.0
0.210	223.0	-0.29	108.0
0.230	244.0	-0.27	120.0
0.250	213.0	-0.25	113.0
0.270	211.0	-0.23	116.0
0.290	188.0	-0.21	131.0
0.310	135.0	-0.19	135.0
0.330	164.0	-0.17	131.0
0.350	124.0	-0.15	120.0
0.370	95.0	-0.13	116.0
0.390	78.0	-0.11	120.0
0.410	70.0	-0.09	120.0
0.430	52.0	-0.07	144.0
0.450	49.0	-0.05	141.0
0.470	59.0	-0.03	158.0
0.490	43.0	-0.01	117.0
0.510	30.0	0.01	162.0
0.530	30.0	0.03	136.0
0.550	28.0	0.05	148.0
0.570	22.0	0.07	119.0
0.590	16.0	0.09	117.0
0.610	14.0	0.11	139.0
0.630	11.0	0.13	106.0
0.650	11.0	0.15	120.0
0.670	14.0	0.17	114.0
0.690	14.0	0.19	107.0
0.710	4.0	0.21	131.0
0.730	10.0	0.23	134.0
0.750	3.0	0.25	147.0
0.770	4.0	0.27	125.0
0.790	4.0	0.29	140.0
0.810	2.0	0.31	119.0
0.830	3.0	0.33	110.0
0.850	2.0	0.35	120.0
0.870	1.0	0.37	116.0
0.890	3.0	0.39	122.0
0.910	4.0	0.41	164.0
0.930	2.0	0.43	157.0
0.950	0.0	0.45	148.0
0.970	0.0	0.47	120.0
0.990	3.0	0.49	140.0

SUGGESTED CLIP = 1.471TO CHANGE TYPE 1 ELSE 0

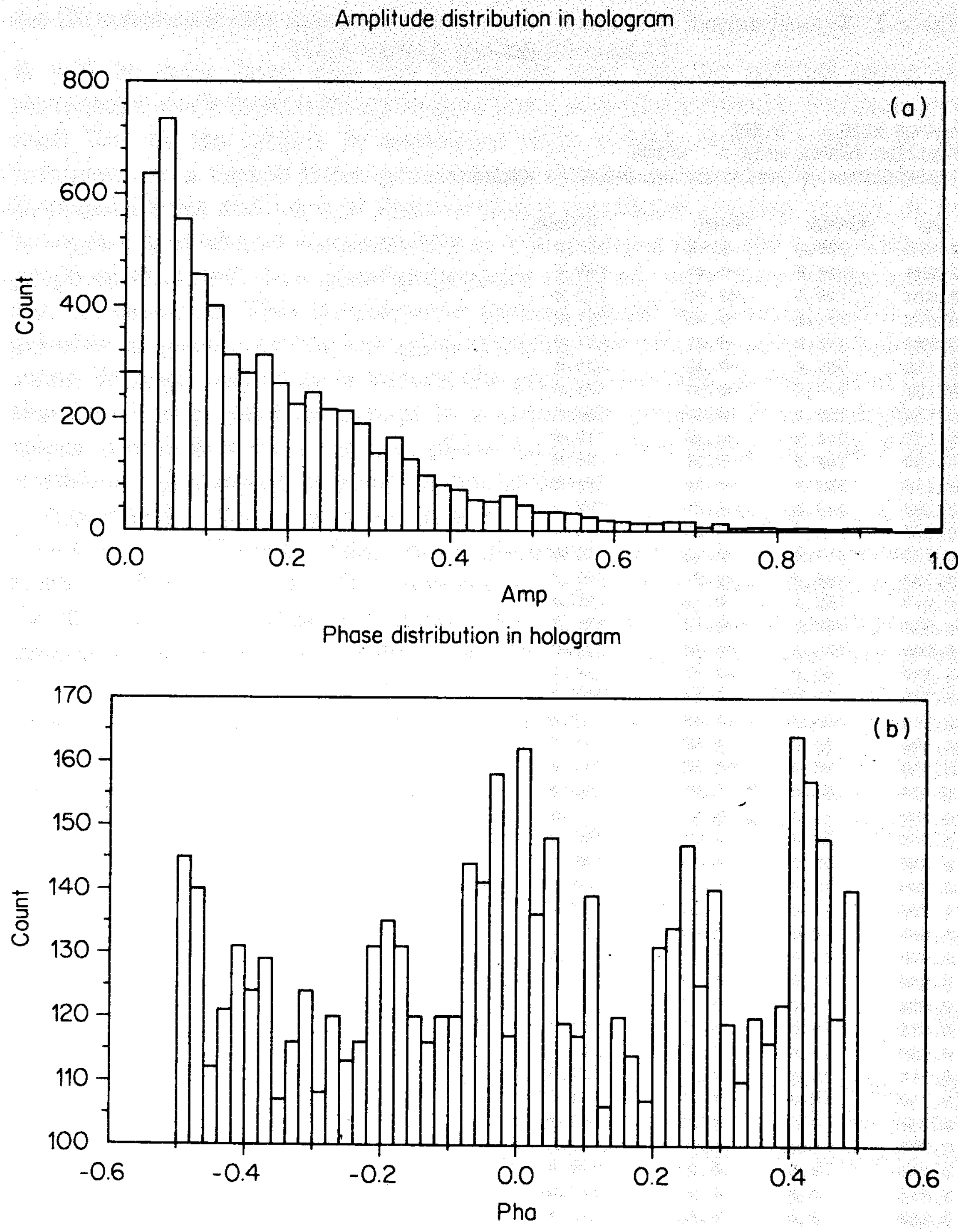

Figure 10. (a) Amplitude histogram: with random phasing. (b) Phase histogram: with random phasing

(i, j) that is normalized to 2π. The sampling points are the centres of cells covering the x–y plane whose dimensions are $(\Delta x)^2$ and are labelled (i, j). The origin of the coordinate system is in the centre of the array with a cell symmetrically disposed about the origin, as shown in Figure 11. Hence, if the number of sampling points in the x- and y-directions are i_m, j_m

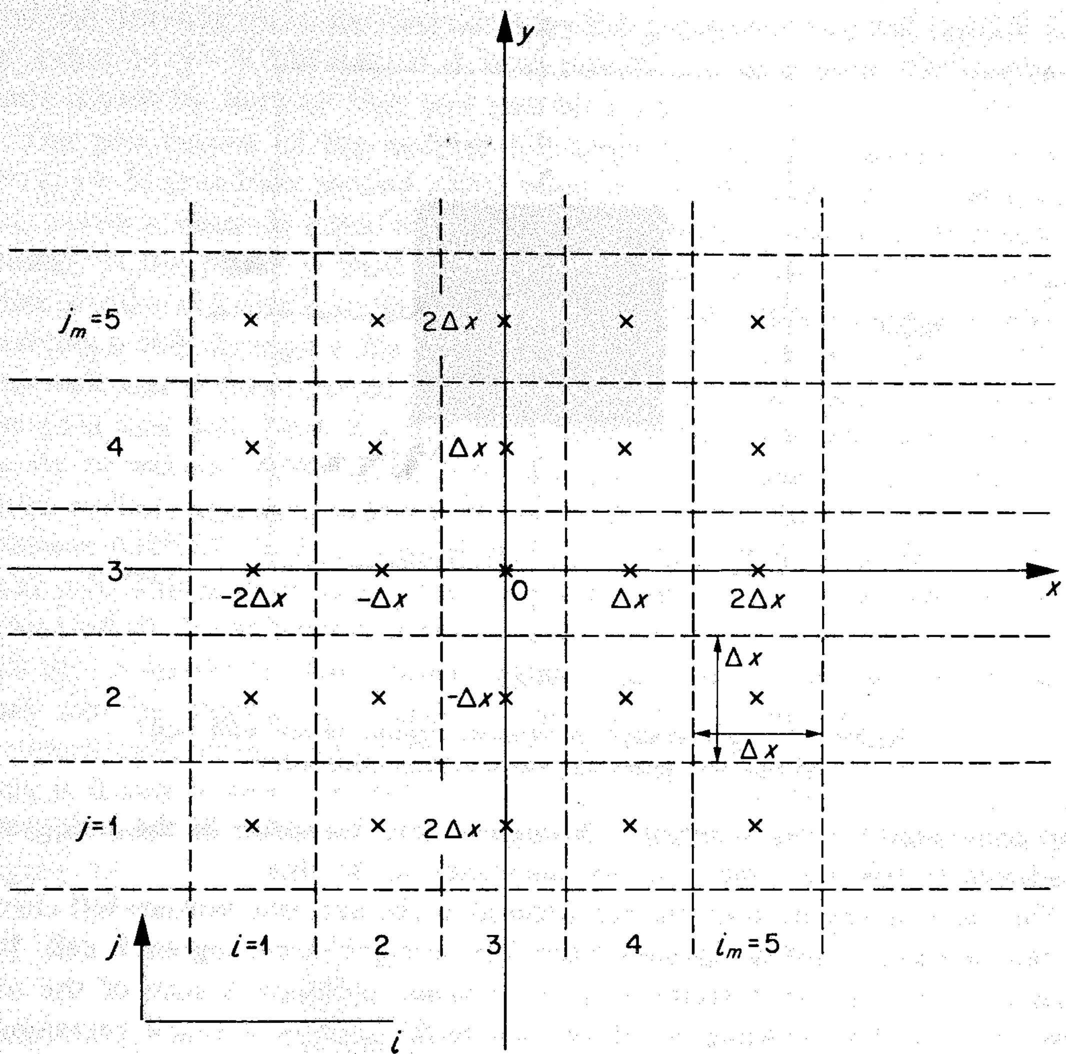

Figure 11. Distribution of sampling points in the (x, y) plane

(ISAMPS and JSAMPS in the program) then the coordinates of the sampling points are

$$x=[i-\tfrac{1}{2}(i_m+1)]\,\Delta x, \tag{50}$$

$$y=[j-\tfrac{1}{2}(j_m+1)]\,\Delta x. \tag{51}$$

An 80×80 cell array is used. The sampling point is $(n\,\Delta x,\ m\,\Delta x)$ and it is assumed that $F(x, y)$ is virtually constant over the cell area. Figure 6 shows the cell drawn, conventionally, as completely opaque, except for an aperture of height $W_{nm}\,\Delta x$ and width $c\,\Delta x$ that is displaced a distance $P_{nm}\,\Delta x$ from the centre. This is the aperture that is the basis of the preceding theoretical discussions. It is not, however, computer-generated in this way. It is, in fact, drawn, by the computer, in the manner of Figure 12 in which the aperture is made opaque by using a black ink pen plotter or by exposing a 35 mm film

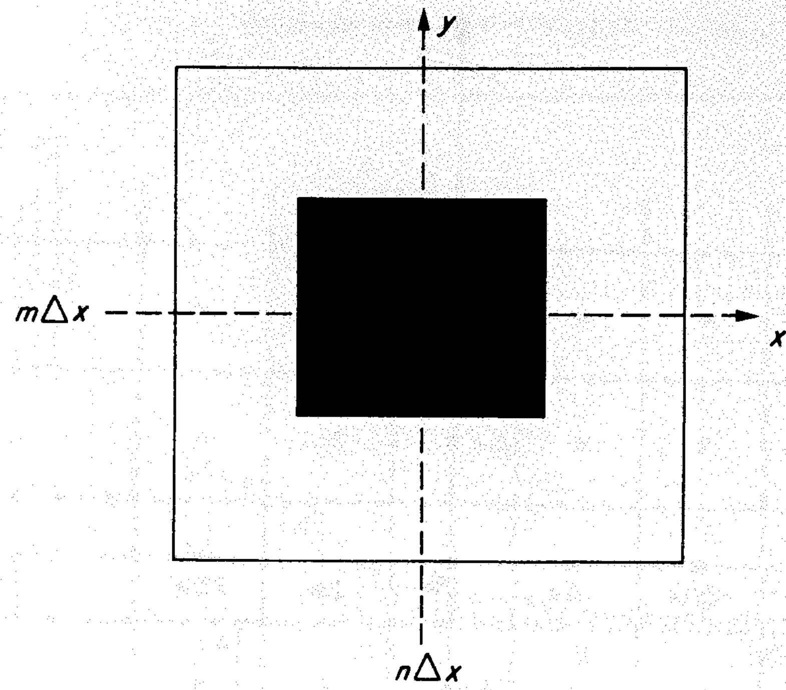

Figure 12. Appearance of opaque region of the unit cell of the hologram as drawn by the computer

unit controlled by the computer. A negative can be made of the holograms produced in this way, but it is not necessary to do this.

The plotting begins with the pen located at the extreme bottom left corner of the hologram and progresses from left to right, scanning each cell. It is then relocated at the extreme left, once again, to begin a scan of the next row of cells. The shading in of the apertures creates a black rectangular shape in each cell as shown in Figure 12.

The details of the way the program controls the behaviour of the plotter will now be discussed. The first important point is that the pen itself has a finite width, as also does the beam that is used in the film plotter device. A reasonable estimate of the pen width is 0.5 mm and for the beam 0.2 mm. This means that the minimum line thickness that can be made by the pen is ~1/50 in. Hence the minimum amplitude that can be represented graphically is limited by the linewidth of the plotter. The width of each aperture is a constant equal to half the width of the unit cell, and the unit cell size is found from a knowledge of the drawing area.

The area of the hologram is selected as either $3600\times270\ \text{mm}^2$, $3600\times 840\ \text{mm}^2$, or $430\times270\ \text{mm}^2$. Now if i_m and j_m (NXDIM and NYDIM in the subroutines) are the number of cells in the x- and y-directions then, for the 3600×270 area, say, the two quantities $x_s=3600/i_m$ and $y_s=840/j_m$ can be compared. The smallest of the two values x_s, y_s is the number of millimetres per unit cell that will just fill the drawing area, and for an 80×80 cell array $x_s=45$ mm and $y_s=3.375$ mm. The cell size is therefore $3.375\times3.375\ \text{mm}^2$

and the drawing units are upgraded, by the program, from the default value of 1 mm to $S = 3.375$ mm. It is now possible to determine the number of lines drawn by the pen that will just fill a unit cell.

The pen begins at the bottom left-hand cell of the hologram, labelled $i = 1$, $j = 1$. It is then moved along what in the plotting space is now called the y-axis a distance equal to $\frac{1}{4}$ (all distances refer to fractions of a unit cell width). If the phase is zero this is the edge of the aperture. The program then decides if a line is to be drawn. If it is not, the pen is moved on without drawing a line through $\frac{1}{2}$. By this time the pen is $\frac{1}{4}$ from the edge of the (1, 1) unit cell and is then moved through $\frac{1}{4}$ to test if a line needs to be drawn in the next unit cell. Now if a line does need to be drawn the pen first moves on by an amount ANGLE (I, J) to incorporate the phase (P_{nm} in the theory given earlier) and then draws a line of length $\frac{1}{2}$. It is then moved back by an amount ANGLE (I, J) to cancel the phase shift before moving on $\frac{1}{4}$ to the next cell. This is done across the complete row of cells. After this the pen is relocated at the left-hand side of the hologram and is incremented up the cell by $\frac{1}{2}$ a linewidth. The $\frac{1}{2}$ linewidth increment arises because the minimum that can be drawn in a cell is one linewidth, but the lines must be symmetrically drawn about the centre of the aperture, i.e. if a single line only is drawn it must symmetrically occupy the centre of the aperture. Using an increment of $\frac{1}{2}$ linewidth causes all the lines to overlap in the manner of Figure 13.

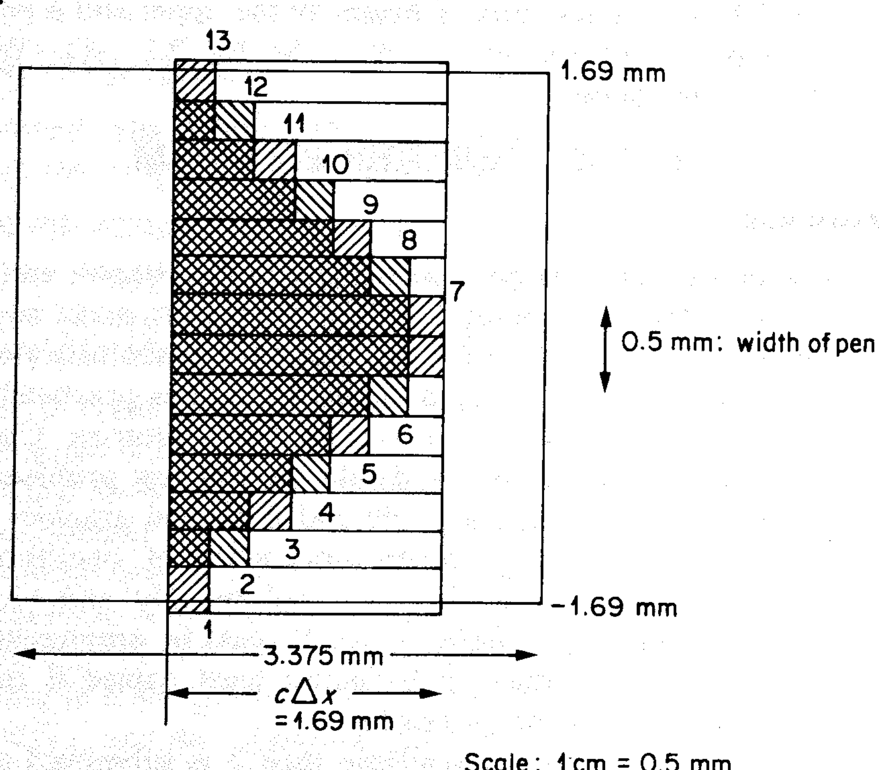

Figure 13. Scale cut-away magnification of the overlapping pen lines for the case $W_{nm} = 1$. 13 lines are used with a small overlap at the top and bottom of the cell

The number of lines that fill a unit cell of the hologram, using a $\frac{1}{2}$ linewidth increment, is

$$\text{NULINE} = 2\left[\!\left[\frac{\text{S}}{\text{ALINWD}} + 0.5\right]\!\right] - 1, \tag{52}$$

where the cell dimension is S, ALINWD is the linewidth of the plotter pen, or beam, and $[\![\ ]\!]$ denotes the truncated value of the number in integer form (equivalent to using IFIX in FORTRAN). For $\text{S} = 3.375$, and $\text{ALINWD} = 0.5$, NULINE is 13, as shown in Figure 13. Thinner lines are possible for the film output device so $\text{ALINWD} = 0.2$ is chosen in the program, giving $\text{NULINE} = 33$. The formula (52) ensures that, if the aperture is as wide as the cell, the cell is completely filled with lines. There is, with this method, a small overspill. This, however, is variable because the pen-width tends to vary. This possibility of variation in pen-width really makes overlapping advisable, even if there are other non-overlapping methods of filling the cell. The overspill does not prevent the hologram from behaving satisfactorily, since there are relatively few completely filled cells.

The height of the aperture is equal to the amplitude $\text{W} = \text{FXY}(I, J) \times \text{CLIP}$ associated with the cell, where $0 \leqslant \text{W} \leqslant 1$. The program forms two fractions $\text{X} = \text{N/NULINE}$ and $\text{X1} = (\text{N}-1)/\text{NULINE}$ where N labels a line somewhere between the bottom and the top of the cell and $1 \leqslant \text{N} \leqslant \text{NULINE}$. For an aperture of height W the upper end is positioned at $(1+\text{W})/2$ and the lower end at $(1-\text{W})/2$. So for $\text{X1} < \frac{1}{2}(1-\text{W})$ and $\text{X} > \frac{1}{2}(1+\text{W})$ no lines are drawn.

6. THE COMPUTER PROGRAM

6.1 Description

The full computer program is listed at the end of the chapter and contains a list of comments sufficiently comprehensive for a user to make any necessary changes. Obviously a program like this must use a sophisticated graphics package. The one used here is called GINO-F that has practically universal use in the U.K. and is available in some other countries. Changing the program to fit an alternative system should present few problems.

The program can be run interactively and if a good graphics package is available this is an extremely satisfactory mode of operation. This is because many features can be demonstrated quickly and vividly in an interactive mode using video display units. It must be emphasized, though, that it is necessary to produce high-quality hard copies if the eventual reconstruction process is to be successful.

If interactive facilities are not available then it is submitted as a normal job to the computer, for batch processing, and the interactive mode requests, that are redundant in this mode, are simply filed via channel 2.

A control file, fed in through channel 1, contains the data that controls the whole program. The object can be any arrangement of 30 straight lines and the coordinates of these are read in through channel 4. The example chosen to illustrate this chapter is the letter group MESC that consists of 16 straight lines. The possibility of using 30 lines should allow most students to develop a hologram for their own initials or even their names. If 30 lines is not enough this can be easily increased through an elementary program modification.

The amplitude and phase distributions, together with the sample rate value, its modified value, the clipping factor, the drawing units of the hologram, and the number of lines needed to fill the aperture appear through channel 3.

Most of the rest of the program was discussed, in one form or another, in the earlier sections. The final stages, however, includes the possibility of producing a line printer output of the amplitude distribution. This is done by overprinting and, if required, appears through channel 3. We also thought that it may be interesting to place a window over the hologram and produce a set of holograms of varying size by simply changing the window. Finally, a facility for controlling the hologram development, whilst using the program in an interactive mode with a video display unit is included. However, the quality of the hard copies of such holograms needs careful monitoring.

6.2 Running the program

If it is assumed that a hologram is to be produced in a non-interactive manner then the following FORTRAN command is used:

```
FORTRAN PROGRAM=HOLO, NONFABS,     LIB=NAG, PLOT,
    DATA1=CONTROL,     DATA4=OBJECT,     WRITE2=OUT1,
    WRITE3=OUT2, JD(JT2000, MZ80K, URF).
```

The purpose in stating this here, and in this form, is that it literally shows what the program does.

The command shows that the program is in a file called HOLO and, since hologram production is likely to be a reasonably lengthy process, that the quick turn-round, limited batch service is not required (hence NONFABS). The comprehensive NAG library is requested and the desire to do plotting is expressed through PLOT. The control data are in a file called CONTROL, object data are in a file OBJECT, the interactive mode requests are dumped into a file OUT1, and the amplitude phase distribution, etc., is placed in file OUT2. The job description, JD(. . .), contains the job time JT that, for the 1904S computer, is $\simeq 2000$ s, the core requirement MZ of 80 K and an urgency code URF (fast).

The two data files are CONTROL and OBJECT. OBJECT contains:

	Format	Name in program
(1) Number of line segments in the object	I2	NULINE
(2) Coordinates of each end of a line (X1, Y1, X2, Y2)	4F(4.1, 1X)	APOINT

Note: Later in the program the definition of NULINE is changed to mean the number of lines that completely cover the aperture.

For the group of letters MESC, in the form  the file OBJECT contains 16:

−3.5	1.0	−3.5	−1.0
−3.5	1.0	−2.5	−1.0
−2.5	−1.0	−1.5	1.0
−1.5	1.0	−1.5	−1.0
−1.0	1.0	−1.0	−1.0
−1.0	1.0	0.0	1.0
−1.0	0.2	−0.5	0.2
−1.0	−1.0	0.0	−1.0
0.5	1.0	1.5	1.0
0.5	1.0	0.5	0.2
0.5	0.2	1.5	0.2
1.5	0.2	1.5	−1.0
0.5	−1.0	1.5	−1.0
2.0	1.0	3.0	1.0
2.0	1.0	2.0	−1.0
2.0	−1.0	3.0	−1.0

The units are set as centimetres for drawing the object as part of the output. The object is restricted to the area
$-10.0 \leqslant X \leqslant 10.0$
$-7.5 \leqslant Y \leqslant 7.5$

The file CONTROL contains:

	Format	Name in program
(1) Output device 1 = lineprinter: for overprinting display of amplitude distribution. 2 = Narrow plotter: normal size 3 = Wide plotter: uses very much wider paper. 4 = Microfilm output.	I1	NDEV
(2) Critical sampling rate multiplier, i.e. sampling rate actually used = critical rate × AMULT	F4.1	AMULT
(3) Number of sampling points in x-direction	I3	ISAMPS (or NXDIM)

(4) Number of sampling points in y-direction	I3	JSAMPS (or NYDIM)
(5) 1: With random phasing 0: Without random phasing	I1	NRAN
(6) 1: Modification of clipping factor 0: No modification of clipping factor	I1	NCLIP
(7) New clipping factor if 1 is used for (6)	F4.2	CLIP
(8) Number of extra (smaller holograms required by window variation	I1	NSM
(9) In the final section of subroutine DISPLAY, a further opportunity to modify the clipping factor occurs. It is appropriate only to the interactive mode and is suppressed by using 0.0 as data	F3.1	CLIP
(10) Related to the interactive opportunity, outlined in (9). Entering 0 as data suppresses this part of the program.	I1	NDEV

A typical example of this file is as follows:

2 : narrow plotter selected
0.66: AMULT
80 : number of cells in x-direction
80 : number of cells in y-direction
0 : no random phasing required
0 : no modification of clipping factor required
0 : no extra holograms requested
0.0 : suppression of interactive mode opportunity
0 : suppression of interactive mode opportunity

In the interactive mode all program input and output takes place at the terminal so that in the FORTRAN command = CONTROL, = OBJECT, = OUT1 and = OUT2 are omitted. The interactive mode is extremely valuable for instantly monitoring the influence of parameter variation on the hologram development, but the non-interactive mode is generally satisfactory for production runs.

6.3 Output from the program

The hologram, of course, appears on the graphical output device but, as pointed out earlier, in the non-interactive mode the interactive requests are dumped on to channel 2 and the amplitude and phase tables appear through channel 3.

Before the hologram is drawn, the object is plotted out and, as the comment in the program states, although it is not actually necessary, this action does provide a check on the object data. The amplitude and phase distributions are plotted next and, finally, the hologram (or holograms) are drawn.