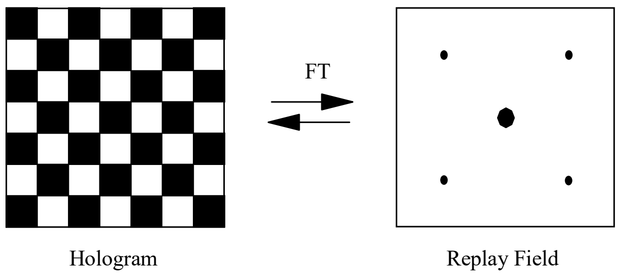
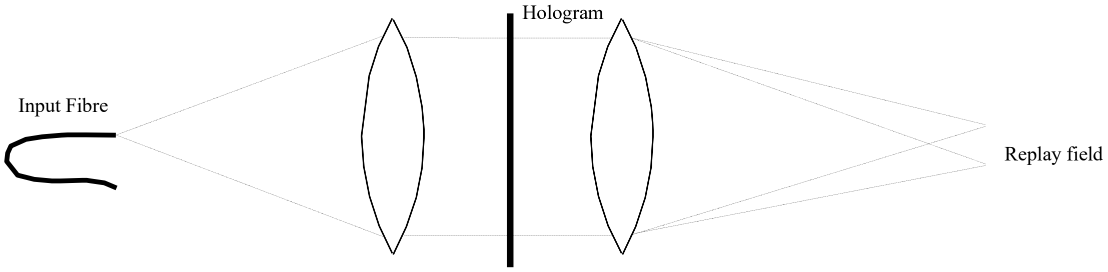
4B11 Crib 2021

Q1 a) [25%] By combining an array of these pixels at various positions on a regular grid, it is possible to generate a complex amplitude function in the far field. Such a 2-D combination of these pixels in various positions is defined as a Hologram and the pattern generated by the hologram if the far field is the Replay Field. The translation between the two is the Fourier transform.





b) [30%] The far field of a single square pixel is its Fourier transform: $F(u,v) = A\Delta^2 \mathrm{sinc}(\pi\Delta u)\mathrm{sinc}(\pi\Delta v)$

The original rectangular aperture is defined as a single pixel. By combining an array of these pixels at various positions on a regular grid, it is possible to generate a complex amplitude function in the far field. By altering the value of the amplitude *A* of each pixel, centred on a grid of interval *b* (in the example of two pixels above Δ was equal to *b*, but may not always be so), it is possible to add up the 2-D sinc functions and create an arbitrary 2-D distribution in the far field region. By superimposing all the exponential phase terms due to the shift and varying the amplitude A, it is possible to create useful patterns in the far field. In general terms, the broader the feature or combination of pixels, the smaller or more delta function-like the replay object. The exact structure of this distribution depends on the shape of the 'fundamental' pixel and the number and distribution of these pixels in the hologram. The pattern we generate with this distribution of pixels is repeated in each lobe of the sinc function from the fundamental pixel. The lobes can be considered as spatial harmonics of the central lobe which contains the desired 2-D pattern. The spatial coordinates (*u,v*) are related to the original absolute coordinates used earlier in the diffracted aperture (α,β) by the relation.

$$u = \frac{k\alpha}{2\pi f} \qquad v = \frac{k\beta}{2\pi f}$$

Note that as shown before, there are other orders appearing in the replay field due to the orders of the first suppressed zero in the sinc envelope. This effectively limits the useable area in the replay field to half if overlapping hologram replay patterns are to be avoided.

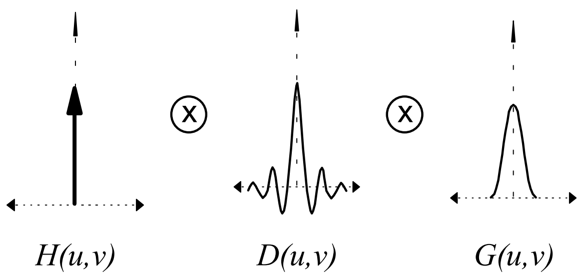
The ideal hologram replay field *H(u,v)* is designed as an array of delta functions in desired positions. The lens aperture *p(x,y)* is a large circular hole, so the FT *P(u,v)* will be a first order Bessel function (like a circular sinc function). The hologram aperture is a large square of size *NΔ* and its FT, *H(u,v)* will be a sharp sinc function. The effect of the FT of the illumination *G(u,v)* is to add a Gaussian profile. Hence, the profile of the spots in the hologram replay field will not be delta functions, they will be delta functions convolved with a Bessel function convolved with a sinc function convolved with a Gaussian function. This means that the replay field will not look exactly as expected, spots which are placed next to one another will interfere due to the tails of the Gaussian, sinc and Bessel functions and the individual desired sharp 'spots' become ringed blobs with finite width. In most cases, the hologram aperture will be smaller than the lens, so the effect of *p(x,y)* can be ignored.

In the examples given the illumination of the hologram is uniform and that the hologram and lens extend to infinity. This is not the case in the real world, as there are a finite number of hologram pixels creating a aperture over the hologram and the light used to illuminate it will not be uniformly distributed. In all the examples we are assuming that the illumination source is a collimated monochromatic laser which generates high quality parallel wavefronts with a wider diameter than the hologram or the lenses. Such a source will usually have an intensity distribution which can be expressed as a Gaussian beam profile or function.

$$g(x,y) = A_G e^{l\left(x^2+y^2\right)}$$

The entire illumination system (apodisation) can be modeled as a sequence of multiplied functions. The input illumination distribution *g(x,y)* times the hologram aperture *d(x,y)* times the total aperture of the FT lens (if it has a smaller diameter than the hologram) *p(x,y)*. Hence effect of the FT on these functions results in a convolution of their transforms.

$$F(u,v) = G(u,v) \otimes D(u,v) \otimes P(u,v)$$



c) [25%] The size of the replay field scales with the pixel pitch, the focal length of the FT lens and the wavelength of the laser light source. The maximum width of the replay field is defined by the central lobe of the sinc envelope.

$$\alpha_M = \frac{f\lambda}{\Delta} \quad \beta_M = \frac{f\lambda}{\Delta}$$

If we assume the green wavelength of 532nm to be the central scaled size of the replay field, then we can calculate scaling factors to use to scale the target replay field of the blue and red holograms in order to make them all the same size.

Blue scale = $\alpha_{MG}/\alpha_{MB} = \lambda_B/\lambda_G$. = 532/450 = 1.18 Red scale = 532/670 = 0.794

This is independent of the focal length and pixel pitch, assuming an achromatic transform lens.

The phase modulation of the SLM also depends on the wavelength of the light so there will be a residual zero order due to imperfect pi phase which will require filtering or blocking.

d) [20%] The difference between 2D and 3D is that need for depth information on the 3D hologram. There are several ways that depth can be included in the calculation or algorithm. The commonest is to include a Fresnel component to the diffraction calculation usually through a modified Fourier transform. The calculation process depends on the way in which the data is presented to the system. There are two main ways:

Layers or Fresnel slices. The 3D target is reduced into a series of 2D slices and a hologram is calculated for each one. They are then synthesised into a single 3D hologram.

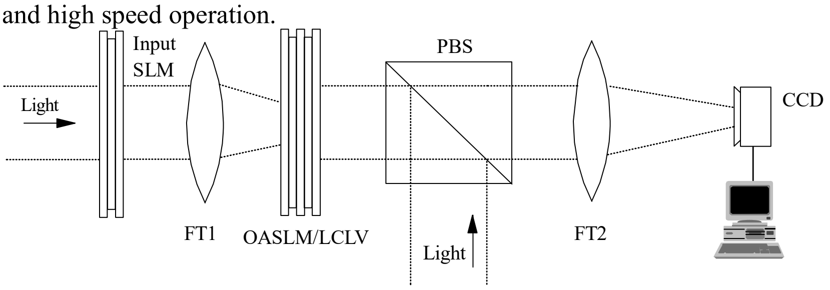
Point clouds/polygons. The 3D object is defined as a series of points or polygons in 3D space. Each point is then diffracted back to the hologram place and then summed up to form a complex hologram that is then quantised to suite the modulation.

Any discussion of depth information in the holograms is acceptable here

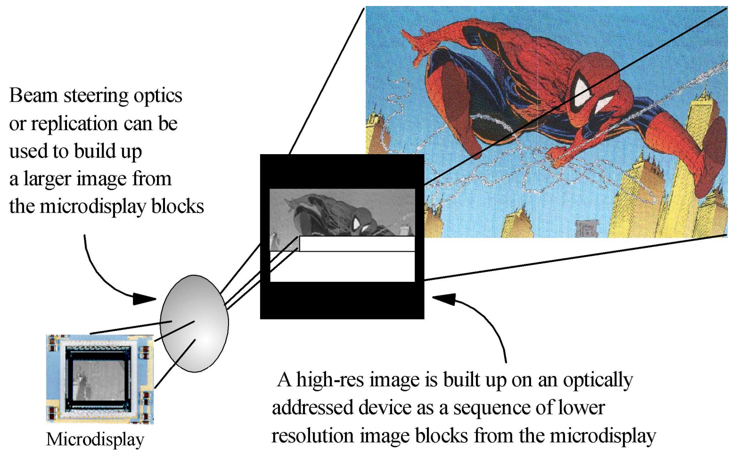
[a) Rather poor definitions of a CGH and Fourier transform, good assumptions made. b) Fairly standard derivation, but many answers followed a poorly laid out proof without much logic. Quite a few missed the relevance of N. c) Quite a few missed the fact that the holograms needed scaling with the wavelength. d) A few picked up the 3D Fresnel]

Q2 a) [30%] An OASLM takes the intensity of an incoming optical signal and converts it into an image which is displayed on a SLM. It is a way of capturing, processing and storing optical intensity information within an optical system. Two main applications:

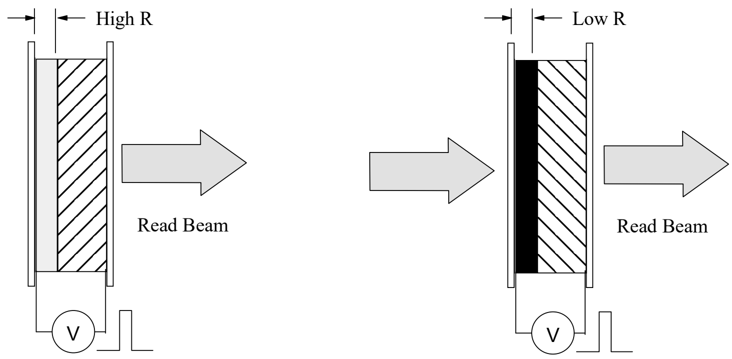
The OASLM can be used as the non-linearity in the Fourier plane of an optical correlator such as a JTC. The OASLM effectively thresholds the intensity of the spectrum from the input plane, which gives good sharp correlation peaks and high speed operation.



A second application is in projection displays, where the OASLM acts as a storage device. The almost continuous resolution can be exploited by storing low res tiles spatially onto the OASLM to make a higher res image or hologram. The OASLM can be used to convert from incoherent to coherent light.



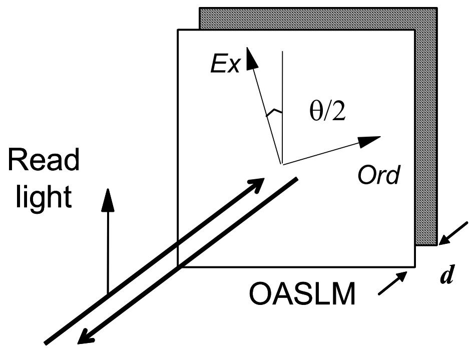
A LC OASLM can be made from a thin layer of LC sandwiched between a conductive glass electrode and a photoconductive layer. There is an external electrical field applied, common to the whole OASLM and is used to time the read and write cycles of the OASLM. The photoconductive layer is usually a thin film such as CdS or amorphous silicon (aH:Si) which has the required photo-active properties. When the external electrical field is positive (write cycle) a voltage appears across the aH:Si and LC layers. If the aH:Si is not illuminated (in a dark state) the its resistivity is very high, and most of the voltage appears across the aH:Si. This means that the voltage across the LC is insufficient to switch the molecules. If the aH:Si is illuminated (bright state), then the resistivity of the aH:Si is low and the voltage appears across the LC, causing it to switch.



Dark State (Not illuminated) Bright State (Illuminated)

If a nematic LC were to be used, then it would offer the potential for grayscale modulation of either intensity or phase. It also offers the potential for a nonlinear optical response if the modulation characteristics are used in conjunction with polarisors or in a twisted structure.

b) [25%] A reflective OASLM uses a layer of FLC twice as it is reflected off the photoconductive layer. In order to obtain binary phase, the SLM must act as a half-wave plate, which in reflection will be a quarter-wave plate. The FLC must have a switching angle of 90 degrees for optimal performance. For binary phase the light should be linearly polarised and bisect the two switched states of the FLC.



For the layout above, the analyser should be horizontally aligned, however if the switching angle of the FLC is 90 degrees and the retardation is pi, then it is not needed as the OASLM will give pure binary phase.

c) [20%] For binary phase the device must act as a half-waveplate with a retardance of pi.

$$\Gamma = \frac{2\pi d \Delta n}{\lambda} = \pi \quad \text{Hence d} = 2.95\text{x}10^{-6}\text{m (2.95um)}$$

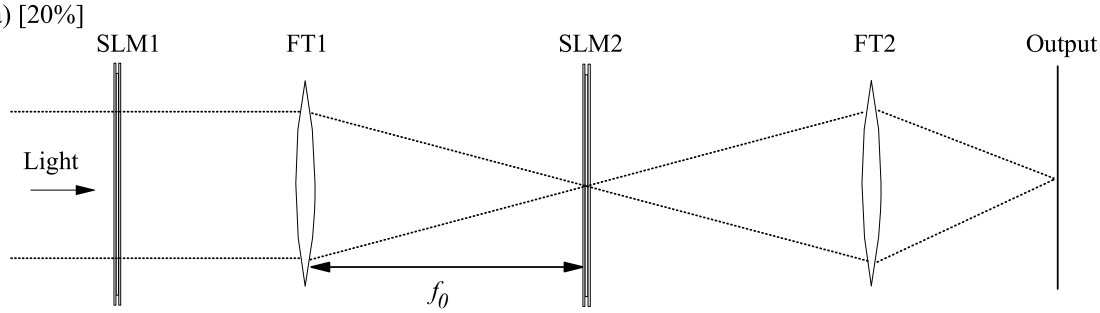
However, for a reflective device we need half this thickness hence d = 1.48um.

d) [25%] In order to make a reflective OASLM there must be a reflective layer next to the photoconductive layer sandwiched between the photoconductor and the LC. Most photoconductive layers are not ideal reflectors, although aH:Si is about 60% reflective at 650nm, hence is can be used as a reflective layer. The problem with this is that the aH:Si is also sensitive to 650nm so if care is not taken, the read light will act as a write beam for the OASLM.

Another option is to insert a mirror between the LC and the a-Si:H, however this is not ideal either. A metal mirror will be conductive which is Ok across the layers, but not ideal laterally as the charges from the aH:Si will leak away from the image. Could use a dielectric mirror or Bragg reflector as these are insulating.

[*a) Overall a well answered section with most defining an OASLM quite well. b) Not many got this and did not see the relevance of the reflective structure. c) Most got this with a few being out by a factor of 2. d) Some managed to get the fact that the mirror had to be pixelated.*]

Q3 a) [20%]



The basic optical layout for a BPOMF follows directly from the theoretical expectations. The input light illuminates SLM1 which is used to display the input image. SLM1 is also a FLC SLM, but it is used in intensity mode (black and white). The SLM is an N_IxN_I array of square pixels, with a pitch of Δ_I, we are assuming that there is no pixel deadspace. The modulated light then passes through lens f_0 which performs the FT of the input image. The FT is formed in the focal plane of the lens and will have a finite resolution. Once the FT of the input (matched in size to SLM2) has passed through SLM2, the product of the input FT and the BPOMF has been formed. This is then FT'ed again by the final lens and the output is imaged by a CCD camera.

- Light source is collimated coherent laser illumination
- FT1 and FT2 are positive focal length lenses
- SLM1 and SLM2 are FLC transmissive SLMs. SLM1 is amplitude, SLM2 is phase.

The main limitation of this system is the alignment and scaling of the FT of the input image on SLM1 and the filter in SLM2. They must be matched and aligned to the nearest pixel and maintained by the opto-mechanics of the system. As the SLM shrinks in size, so does the pixel pitch which makes the alignment even more critical.

b) [20%] The FT is formed in the focal plane of the lens and will have a finite resolution (or 'pixel' pitch) given by. This comes from the sinc envelope for a pixellated object.

$$\Delta_0 = \frac{f_0 \lambda}{N_1 \Delta_1}$$

There are N_1 'pixels' in the FT of the input image on SLM1, hence the total size of the FT will be $N_1\Delta_0$. The BPOMF is displayed on SLM2 in binary phase mode. SLM2 is also a FLC device with N_2xN_2 pixels of pitch Δ_2. The FT of SLM1 must match pixel for pixel with the BPOMF on SLM2 in order for the correlation to occur. For this reason we must choose f_0 such that.

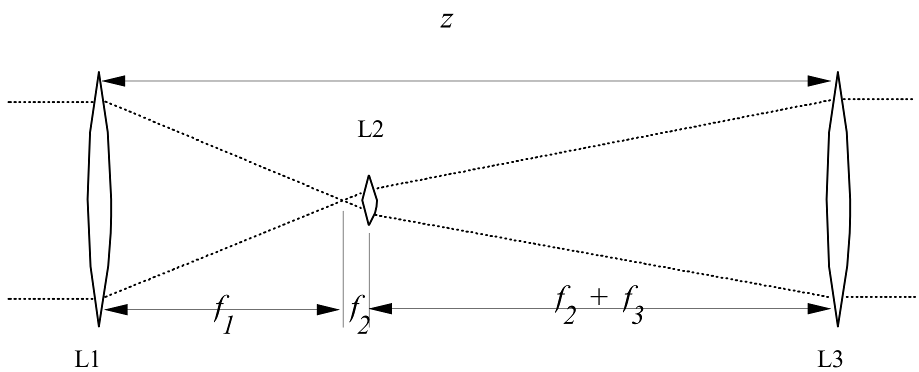
$$N_1\Delta_0 = N_2\Delta_2$$

Hence we can say.

$$f_0 = \frac{N_2\Delta_2\Delta_1}{\lambda}$$

Example SLM1 and SLM2 are both 128x128 pixel FLC devices, with a 220μm pixel pitch at a wavelength of 633nm. The required focal length to match the input FT to the BPOMF in this case is 9.787m which is clearly impractical as an experimental system. The overall length of the BPOMF would be in excess of 20m!

c) [30%] It is possible to shorten the actual length of the optical transform whilst still keeping the effective focal length that is desired by including further lenses in a combination lens. One technique is to combine a positive lens with a negative lens to make a two lens composite. This gives a length compression of around $f_0/5$ which in the example above is still 2m and impractical. Furthermore, the two lenses combine in aberrations which leads to poor correlations due to optical quality.



The three lens combination above provides a much higher length compression of around $f_0/15$. The only drawback is that the more elements you have, the more position errors there will be when the optical system is mounted. The three lens system is an FT lens followed by a two lens telescope to magnify the FT to the size of SLM2. The first lens f_1 generates the FT of the input in its focal plane. The size of this FT will be too small, as f_1 is chosen to be much shorter than f_0. The FT is then magnified by lenses f_2 and f_3 which are configured in a telescope. Lenses f_2 and f_3 are spaced $f_2 + f_3$ apart so that an object placed a distance f_2 in front of lens f_2 will be magnified after lens f_3 by a magnification of f_3/f_2. Hence, if the FT is placed a distance f_2 in front of lens f_2, then it will be magnified after f_3. We can now set the magnification factor so that the input FT matches the size of the BPOMF.

$$Magnification = \frac{f_3}{f_2} = \frac{N_2\Delta_2\Delta_1}{f_1\lambda}$$

d) [30%] If the chosen lenses and SLMs fit this equation, then the FT will be the same size as the BPOMF and the total length of the three lens system will be.

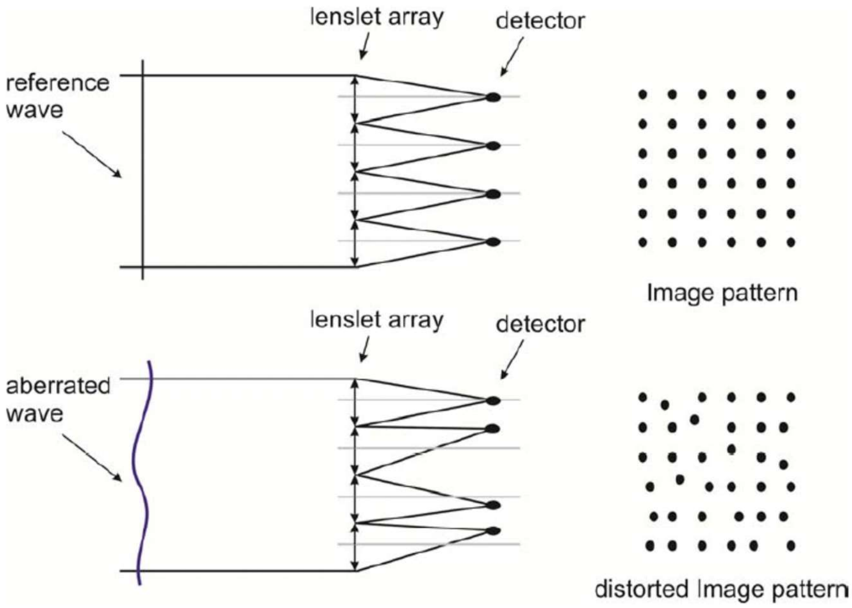
$$z = f_1 + 2f_2 + f_3$$

When choosing the lenses, care must be taken to find good quality low aberration doublets. Choose a suitable f_1 length. Aim to create an FT of the input image which needs scaling by an integer value of magnification to match SLM2. The choose f_2 and f_3 to make this magnification.

Example For the SLMs in the above example, a first lens f_1 = 250mm was chosen to perform the initial FT. This means a magnification of 39.1 is required for the telescope. From the available catalogue lenses, a combination of f_2 = 10mm and f_3 = 400mm was chosen to give a magnification of 40. A lens of f_3 = 365mm was also available, but was not chosen as the magnification was closer with 400mm. Although the other lens offers a shorter overall optical length, it is best to choose the lens combination which offers the closest magnification to that which is desired. The overall length of the system was z = 670mm, which means that the BPOMF can now be constructed on an optical table. Pros – low wavefront aberration, shorter length, Cons – more lenses, weight, alignment.

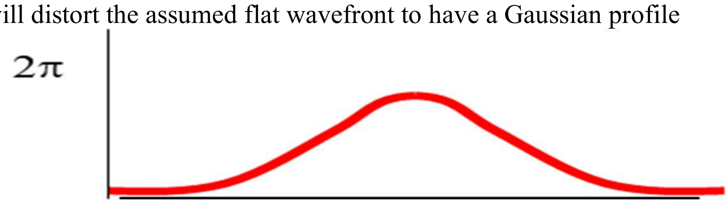
[*a) Well answered as a whole. Only a few candidates pointed out the simpler mechanical properties and several drew unnecessary full optical diagrams. b) Probably a bit simple derivation and not much of a challenge, a few pointed out the result from 1(b) l c) Similar derivation, but quite a few did not explain their magnification factor well. d) quite well answered despite being speculative*]

Q4 a)[20%] One such method for measuring the slope of the wavefront, is the Shack-Hartmann wavefront sensor. This is made by attaching a lens array to the front of a camera, spaced by the focal length of the lenslets. For a plane wave, a spot will be focussed on the optical axis of each corresponding lenslet in the array, as shown in the diagram below as the reference wave. For a distorted wave each focussed spot is displaced and this displacement is proportional to the slope of the wavefront. The incoming beam, whether it is the reference or the measurement beam, passes through the Hartmann screen which divides the wavefront into many subapertures. These are then focused on the Hartmann grid by the lenslet array. By comparing the difference in the coordinates between the expected and measured beams, we can obtain the slope of the wavefront.

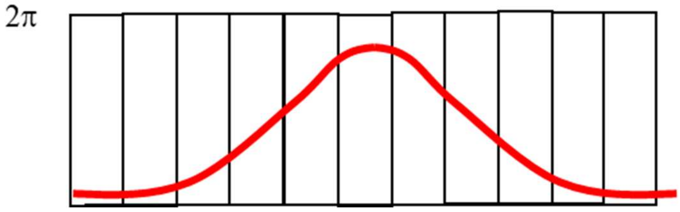


The main limitation comes from the zones defined by the lenslet array. Each lenslet samples a region of the incident wave. The spatial frequencies sensed by each zone will depend on its aperture. The more zones, the higher the frequencies of aberration detected, but also the higher the resolution of the camera needed and the more processing required. Each zone also limits the stroke of the sensor. If the slope of the wavefront in each zone is more than 2π, then it will push the centroid of the spot outside of the corresponding zone on the camera leading to a false detection for its neighbour. Also, if there is too much information in each centroid then it will distort the spot making it hard to track.

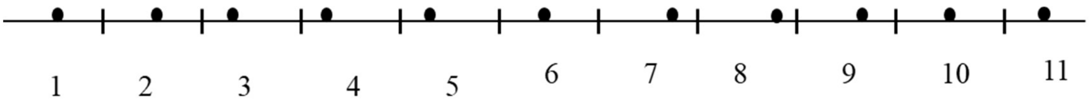
b) [35%] Aberration will distort the assumed flat wavefront to have a Gaussian profile



This is then sampled by the 11 zones of the central row of the SH sensor. Looking at the line scan (ignoring the other dimension), assume that the centroid of the focus spot shifts laterally with the slope of the phase across each lenslet zone. A slope of more than 2pi will lead to the cenroid crossing into the adjacent zone. From the above figure, we can estimate the approximate slope across each zone.



Zone	1	2	3	4	5	6	7	8	9	10	11
Slope(pi)	0	0	0.5	0.5	0.5	0	-0.5	-0.5	-0.5	0	0



c) [25%] With the DBS technique we take a hologram of random pixel values and then calculate its replay field. Then flip the binary value of a randomly positioned pixel and calculate the new replay field. Then subtract the two replay fields from the target replay field, sum up the differences to form a cost function for the hologram before and after the pixel change. If the cost function after the pixel has been flipped is less than the cost function before the pixel was flipped, then the pixel change is considered to be advantageous and is accepted. The new cost function is then used in comparison to another randomly chosen flipped pixel. The process repeats until no further pixels can be flipped to give an improvement in the cost function.

1) Define an ideal target replay field, T (desired pattern)
2) Start with a random array of binary phase pixels.
3) Calculate its replay field (FT), H0.
4) Take the difference between T and H0 and then sum up to make the first cost, C0.
4) Flip a pixel state in a random position.
6) Calculate the new replay field, H1.
7) Take the difference between T and H1 then sum up to make the second cost, C1.
8) If C0 < C1 then reject the pixel flip and flip it back.
9) If C0 > C1 then accept the pixel flip and update the cost C0 with the new cost C1.
10) Repeat steps 4 to 9 until |C0 - C1| reaches a minimum value.

The problem with DBS is that it is very slow to converge. There is also no guarantee that the algorithm will find a global minimum rather than a local minimum. This can be improved by using simulated annealing, however this makes the process even slower.

d) [20%] To correct for an aberration g(x,y) we have to include it in the algorithm steps above. At Step 2, multiply the initial hologram by g(x.y). This will improve the initial estimate, but has negligible impact on the overall optimisation.

The aberration can be included at Step 6, when calculating the new replay field after the pixel flip. We include the corresponding distortion at that pixel due to the aberration as well as the pixel flip. This works well as the calculation of the pixel flip does not involve a Fourier transform, If is done by purely calculating the change in the phase due to inverting the phase of the sinc function when flipping the pixel. This is then combined with the corresponding look up factor from g(x,y).

[*a) Mostly answered well, but a few omitted the slope relationship of the S/H sensor. b) Most got the idea, but their answers were very messy for a large number of marks. Almost nobody estimated the slopes. c) Well answered but no mention of possible error functions other than difference. d) A few good answers in this section showing an understanding of the principles of an adaptive CGH.*]