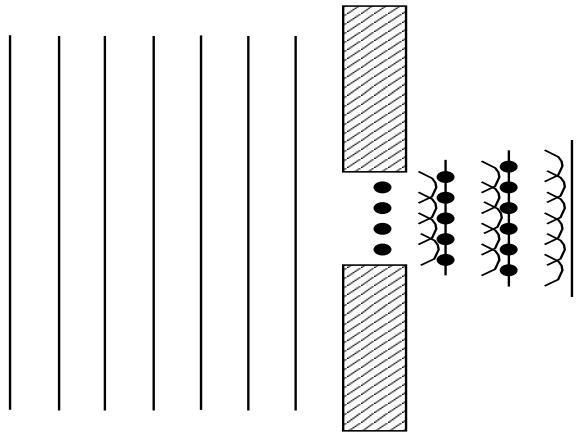
4B11 2015 Exam crib – answers are longer than expected from candidates

Q1 a) [30%] A plane wave can be considered as a infinite array of Huygen's wavelets all radiating spherical waves. If we take on single plane and split it into an infinite line of wavelets, then we can propagate each wavelet a distance and then sum it and all its neighbours some distance away. If we do this, then we gain another plane wave. Assumptions – planes are infinite in x and y. Light only radiates in the forward direction. Light is coherent.



[Good explanation of Huygens wavelets and how to gain plane waves and aperture diffraction. Most did not state assumptions such as forward propagation or coherence.]

b) [50%] If we consider an infinitely small differential of the aperture, *dS*, we can model this as a point source of light emitting spherical 'Huygens' wavelets with an amplitude of *A(x,y)dS*. If we consider only the part of the wavelets which are propagating in the forward (+*z*) direction and are contained in a cone of small angles away from the z axis, then we can evaluate the change in field *dE* at the point *P*, due to *dS*. As the wavelet *dS* acts as a point source, we can say that the power radiated is proportional to $1/r^2$ (spherical wavefront), hence the field *dE* will be proportional to *1/r*. We can see that for a real propagating wave of frequency ω and wave number *k*, ($k = 2\pi/\lambda$) we have a complex wave. The full complex field radiating from the aperture can be written in terms of exponentials as the cosine is just the real part of the complex exponential.

$$dE = \frac{A(x,y)}{r} e^{jwt} e^{-jkr} dS$$

Now, we need to change coordinates to the plane containing the point *P*, which are defined as [α,β].

$$R^2 = \alpha^2 + \beta^2 + z^2$$

$$r^2 = (\alpha - x)^2 + (\beta - y)^2 + z^2$$

Which can be rearranged to give.

$$r = R\sqrt{1 - \frac{2\alpha x + 2\beta y}{R^2} + \frac{x^2 + y^2}{R^2}}$$

The final full expression in terms of *x* and *y* (*dS* = *dxdy*) for *dE* will now be.

$$dE = \frac{A(x,y)e^{j\omega t}e^{-jkR\sqrt{1-\frac{2\alpha x+2\beta y}{R^2}+\frac{x^2+y^2}{R^2}}}}{R\sqrt{1-\frac{2\alpha x+2\beta y}{R^2}+\frac{x^2+y^2}{R^2}}}dxdy$$

If the point *P* is reasonably coaxial (close to the *z* axis, relative to the distance *R*) and the aperture *A(x,y)* is small compared to the distance *R*, then the lower section of the equation for *dE* can be assumed to be almost constant and that for all intents and purposes, *r = R.*

$$dE = \frac{A(x,y)}{R}e^{j\omega t}e^{-jkR\sqrt{1-\frac{2\alpha x+2\beta y}{R^2}+\frac{x^2+y^2}{R^2}}}dxdy$$

The similar expression in the exponential term in the top line of the original equation is not so simple. It can not be considered constant as small variations are amplified through the exponential. To simplify this section we must consider only the far field or Fraunhofer region where.

$$R^2 >> x^2 + y^2$$

In this case, the final term in the exponential ($(x^2 + y^2)/R^2$) can be considered negligible. To further simplify, we use the binomial expansion,

$$\sqrt{(1-d)} = 1 - \frac{d}{2} - \frac{d^2}{8} \ldots$$

Hence the simplified version of the field *dE*, can be expressed as:

$$dE = \frac{A(x,y)}{R}e^{j(\omega t-kr)}e^{jk\left(\frac{\alpha x+\beta y}{R}\right)}dxdy$$

The total effect of the dS wavelets can be integrated across *dE* to get an expression for the far field or Fraunhofer diffraction pattern.

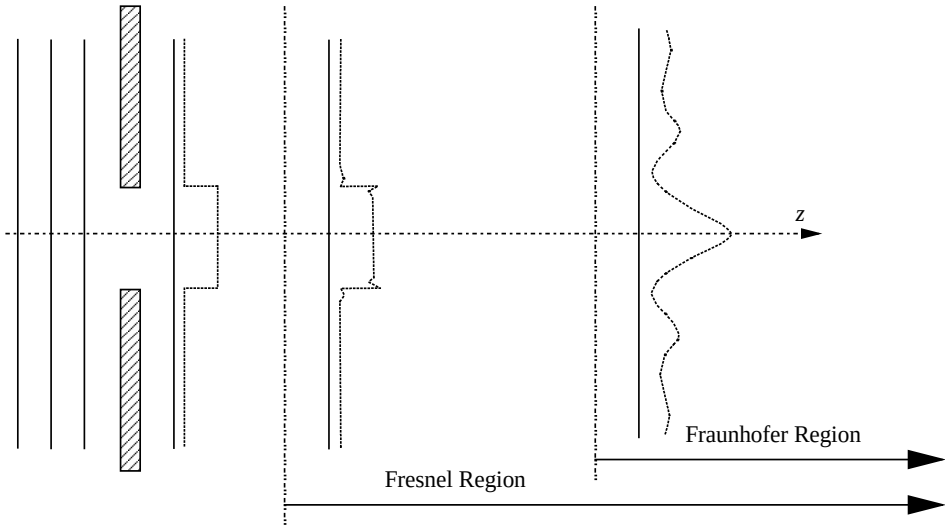
$$E(\alpha,\beta) = \frac{1}{R}e^{j(\omega t-kR)}\iint_{Aperture}A(x,y)e^{jk(\alpha x+\beta y)/R}dxdy$$

Hence the far field diffraction pattern at the point *P* is related to the aperture function *A(x,y),* by the Fourier transform.

[Fairly standard book work question, well answered in general, but many answers followed a poorly laid out proof without much logic to the assumptions made. Many omitted the origin of the 1/r relationship.]

c) [20%] The regions of the approximation are defined such that in the far field or Fraunhofer region, the approximations are accurate, hence the field distribution *E(x,y)* only changes in size with increasing *z*, rather than changes in structure. In the case where the approximation is bearably accurate, we are in the Fresnel region. Before the Fresnel region, the evaluation of *E*

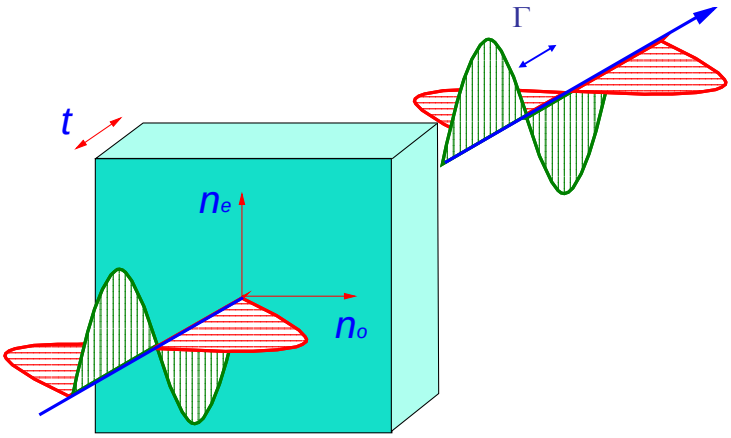
is extremely difficult and is defined as the near field diffraction pattern. The exact boundary of the Fresnel region will depend on the acceptable accuracy and can be found in Goodman.



The Fresnel region is very useful as it allows us to change the field as it propagates, giving rise to 3D images and distributions. It also is for Fractional FTs and the Talbot distance effects. The near field is also very useful as it has very rapid changes in spatial frequencies. This is used in high resolution phase masks and sensing applications. Near field is very useful for sensing applications due to the high sensitivity to the wavelet propagation. Fresnel region is also very useful as it allows the control of propagation in three dimensions in displays, tweezing and microscopy.

[*Not so well answered with most getting the regions, but also concluding that because the other regions are difficult to calculate, then they must not be useful. In fact most current applications using diffraction use these regions*]

Q2 a) [25%] Any material which passes light will have some effect on the that light as it passes through it. This normally takes the form of slowing down the wave and we define the refractive index *n* of the material as the ratio of the speed of light in a vacuum to the speed of light within the material. Most materials (such as water or window glass) exhibit *isotropic* behavior, which means that *n* is constant for any direction that the light travels through the material. Some materials however are anisotropic which means that *n* varies with the direction of propagation. Crystals such as Sodium Chloride have a cubic molecular structure. When light passes through these structures it sees no preferred direction and is relatively unaffected. If the crystal has structure such as hexagonal or triagonal, different directions of light will see very different crystalline structures. The effect of this anisotropy is called birefringence and is property which is exploited in retarders. In a birefringent material, each eigenwave sees a different refractive index and will propagate at a different speed. This leads to a phase retardation between the two eigenwaves which is dependent on the thickness t, of the birefringent material and the wavelength of the light.



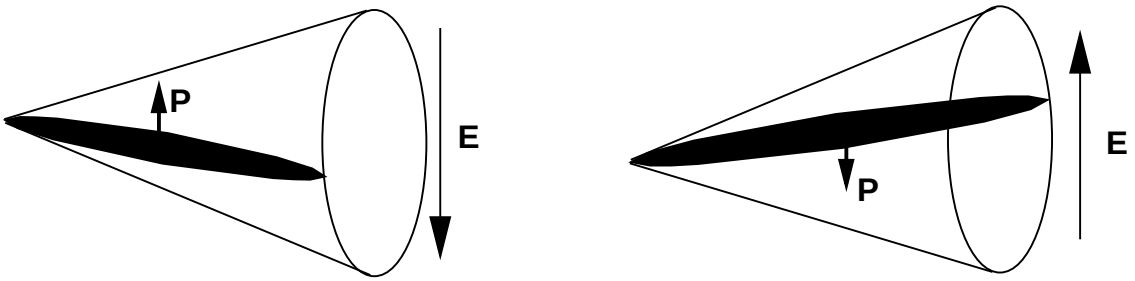
The preferred directions of propagation within the crystal are defined as the extraordinary (or fast) axis and the ordinary (or slow) axis. An eigenwave that passes in the same direction as the extraordinary axis sees a refractive index n_e and the eigenwave that passes along the ordinary axis sees n_o. For light of wavelength λ passing through a birefringent crystal of thickness t, we define the retardation Γ as.

$$\Gamma = \frac{2\pi\ t}{\lambda}\left(n_e - n_o\right)$$

The main limitation is that the retardation only works for one wavelength

[*Good explanations although some confusion between phase and polarisation. Only a few got the limiting factor of wavelength*]

b) [25%] There are some mesophases, which have birefringent interactions with polarised light, due to the molecular order within the mesophase. This is the case with smectic mesophases, where the molecules are not only pointing in the same direction, but are also ordered along their length creating a layered structure. One of the most useful smectic mesophases is the smectic C (SmC) or ferroelectric phase as the molecules are highly ordered and form layers with the molecules tilted within each layer. When an electric field is applied across the molecules there is little interaction as the field is almost perpendicular to the main axis of the LC molecules. The smectic C structure can be improved by adding chirality to the molecular structure which adds an extra dipole perpendicular to the molecular axis of the LC material. This is often is referred to as chiral smectic C or SmC*.



The addition of this dipole due to the chirality of the molecules means that the molecules are in an ordered structure but they are free to move. An applied field across the molecules is now parallel to the extra dipole and the interaction will exert a force on the molecules causing them to rotate about their molecular axis. This rotation is usually about a cone of angles as

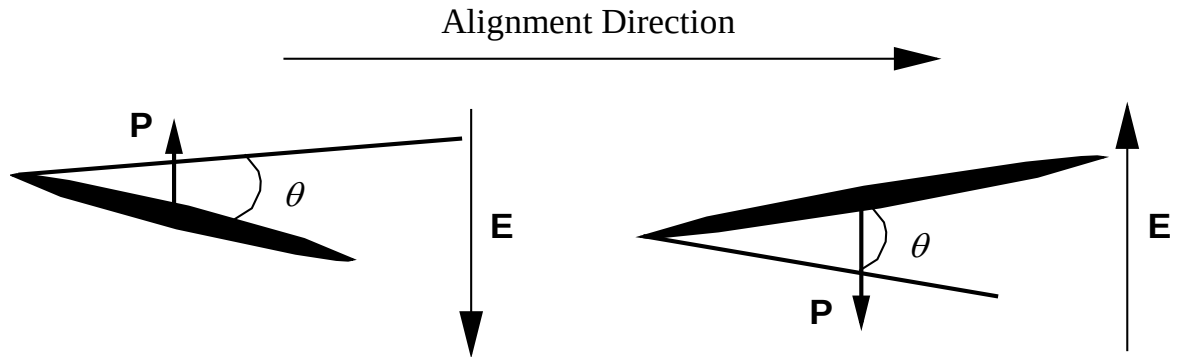
the molecules are designed so that one end is fixed in the layered structure. The dipole **P** of the molecule, which is perpendicular to its length is often referred to as the Spontaneous Polarisation and means that the overall bulk of the LC will have a net potential, hence these materials are called ferroelectric liquid crystals (FLCs).

When an electric field **E** is applied to the cell, there is an interaction between the **E** and **P**, which forces the molecule to move around the cone to a point of equilibrium. If the field is changed, the molecules move again.

[*Some good answers, but also a lot of waffle. No need to mentions surface stabilisation in this section*.]

c) [25%] The SmC* phase in thick (bulk) cells is not ferroelectric because in the equilibrium state the **P** dipoles of the molecules interact with each other forming a helix along the axis of the cell which results in no overall retardance. The fast and slow axes of birefringence align with the direction of each molecule and hence will vary across the area of the cell due to the helix.

If the FLC is restricted to a cell thickness of 2-5μm then the helix of along the cell is suppressed and the molecules are bounded into two stable states either side of the director cone. The angle between these two states is defined as the switching angle θ. This is referred to as a surface stabilised FLC geometry an creates a high degree of ferroelectricity and creates a large birefringent electro-optical effect. The penalty for doing this is that the molecules are only stable in the two states an therefore the modulation will only be binary. The up side to this binary modulation is that it can be very fast (~10μsec) and that the stability can lead to the molecules remaining in the two states in what is known as bistable switching.



To provide an overall direction for the molecules in the FLC cell, an alignment layer is included on the surface of the inner cell walls. This layer contains micro-grooves which run parallel across the surface of the cell and control the direction that the molecules point in and hence align the FLC across the surface of the cell. Molecules are aligned in a uniform direction across the cell and form the layers which are indicative of the SmC phase.

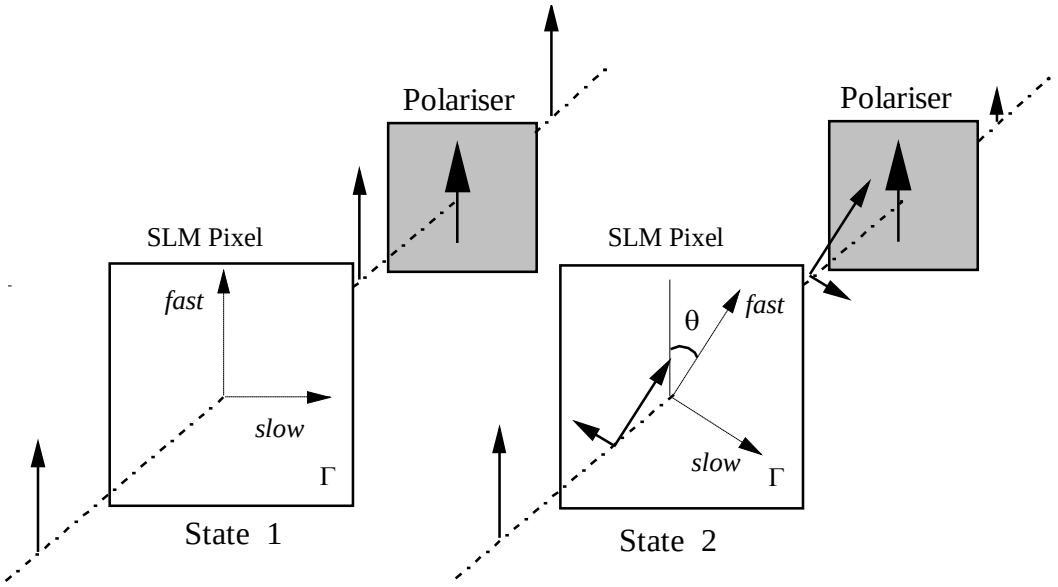
The FLC molecules switch through the angle θ in the plane of the cell when the electric field is applied across the cell. The motion of the molecule is very fast which makes FLCs very attractive for spatial light modulators (SLMs), however the motion of the molecule is only stable in two states which limits the modulation to be binary. The interaction between the

light and the FLCs molecules is dependent on the polarisation of the light and the orientation of the molecules. The director of the molecules acts like the fast axis of a retardation plate and the normal to the direction of the molecules is the slow axis. Hence the FLC displays birefringence across the cell which is aligned in the direction of the molecules. This means that the FLC crystal acts like a switchable waveplate whose fast and slow axes can be in two possible states separated by the angle θ and whose retardation depends on the thickness and birefringence of the FLC. A pixel with retardation Γ at an angle θ can be represented using Jones matrices.

Two main penalties for SSFLC are binary modulation and the extra complexity of having to DC balance the FLC.

[*Most candidates got this right, although once again a lot of waffle. Not many listed the limitations, especially the limits of DC balance*.]

d) [25%] A shutter will use binary intensity modulation



If the light is polarised so that it passes through an FLC pixel parallel to the fast axis in one state, then there is no change due to the birefringence and the light will pass through a polarisor which is also parallel to the fast axis. If the pixel is then switched into state two, the fast axis is rotated by θ and the light now undergoes some birefringent action. We can use Jones matrices to represent the optical components.

State 1

$$\begin{pmatrix} V_x' \\ V_y' \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-j\Gamma/2} & 0 \\ 0 & e^{j\Gamma/2} \end{pmatrix} \begin{pmatrix} 0 \\ V_y \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ V_y e^{j\Gamma/2} \end{pmatrix}$$

State 2

$$\begin{pmatrix} V_x' \\ V_y' \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{-j\Gamma/2}\cos^2\theta + e^{j\Gamma/2}\sin^2\theta & -j\sin\frac{\Gamma}{2}\sin(2\theta) \\ -j\sin\frac{\Gamma}{2}\sin(2\theta) & e^{j\Gamma/2}\cos^2\theta + e^{-j\Gamma/2}\sin^2\theta \end{pmatrix} \begin{pmatrix} 0 \\ V_y \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ V_y\left(e^{j\Gamma/2}\cos^2\theta + e^{-j\Gamma/2}\sin^2\theta\right) \end{pmatrix}$$

If the thickness of the FLC is set so that $\Gamma = \pi$, then the light in the direction of the slow axis will be rotated by 180º. This leads to a rotation of the polarisation after the pixel, which is partially blocked by the following polarisor. Maximum contrast ratio will be achieved when state 2 is at 90º to the polarisor and the resulting horizontal polarisation is blocked out. This will occur when

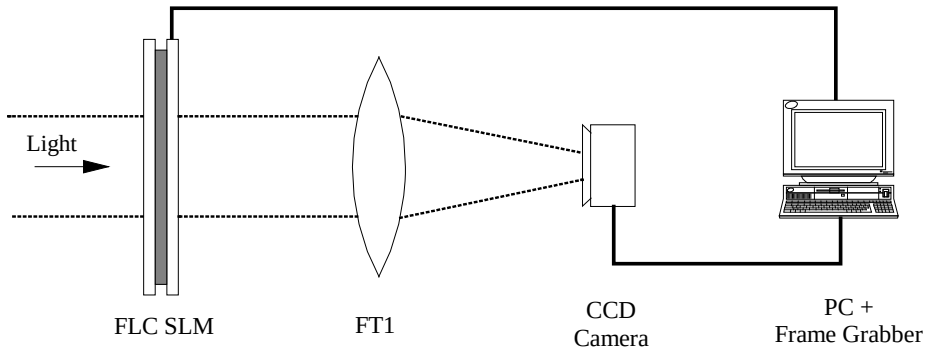
$$V_y\left(e^{j\Gamma/2}\cos^2\theta + e^{-j\Gamma/2}\sin^2\theta\right) = V_y\left(j\cos^2\theta - j\sin^2\theta\right) = 0$$

Hence, the optimum FLC switching angle for a FLC is $\theta = 45^o$.

Applications include shutters for optical switches and laser cut outs for safety systems. The FLC is capable of very high contrast and also very high speed modulation (usec).

[*Surprisingly poor answers for what is basically book work. Most avoided the maths and many talked about phase rather than intensity. Several also used the OASLM structure or drew cell diagrams for LCoS. Several did a full design for a shutter switch.*]

Q3 a) [40%] The JTC works on the basis of a non-linearity working on the spectrum of the input objects to create the product of the two Fourier transforms. This was modeled as a square law detector, but this gives undesirable broad peaks. A much better correlation peak is obtained when the degree of non-linearity is increased as high as possible. A simple square root function on the spectrum gives good narrow peaks, but the best performance is when the spectrum is thresholded. Hence the JTC was originally built using an optically addressed FLC SLM. A very powerful layout of the JTC is to exploit the symmetry about the OALSM or non-linearity. If the optical system is split at this point, then the JTC just becomes two Fourier transforms and in fact can be done with a single laser, SLM and camera by doing two passes through the Fourier transform lens. This is know as the 1/f JTC.

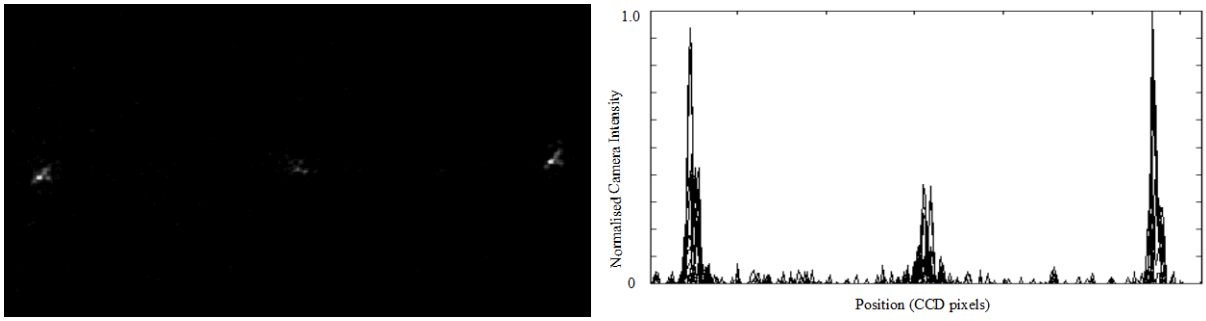


The input and reference images are displayed side by side on a FLC SLM as in a full JTC. The SLM is illuminated by a collimated laser beam and the images are Fourier transformed by a single lens in its focal plane. This spectrum is then imaged onto a CCD camera. The spectrum is then non-linearly processed before being displayed onto the SLM again to form the correlation information. The 1/f JTC is a two-pass system, using the same lens to perform the second Fourier transform.

[*Well answered book work section. Only a few candidates pointed out the simpler mechanical properties and several drew unnecessary full JTC diagrams. No one mentioned the speed penalty*]

b) [30%] If the spectrum from a traditional JTC was directly Fourier transformed, the result would be the two symmetrical correlation peaks characteristic of the JTC along with a huge zero order located in the centre of the output plane. The quality of the correlation peaks and the zero order can be improved by non-linearly processing the spectrum that also suits available FLC SLM technologies. Some of the best results have been reported by using binary thresholds on the spectrum to improve the correlation peaks. A binarised spectrum produces good sharp correlation peaks and reduced zero order. If the binarised spectrum is converted to binary phase modulation [-1,+1], then the zero order can be reduced to around the height of the correlation peaks.

The success of the edge enhancement scheme on the JPS comes from the fact that the image captured will have noise in the dark areas where the JPS intensity is not high enough to be detected. This noise is also edge enhanced to give binary phase noise. The Ft of noise is noise, hence there is no change to the central DC in the final correlation plane. If the JPS is edge enhanced to that there are equal numbers of the two binary states the n there will be no zero order at all.



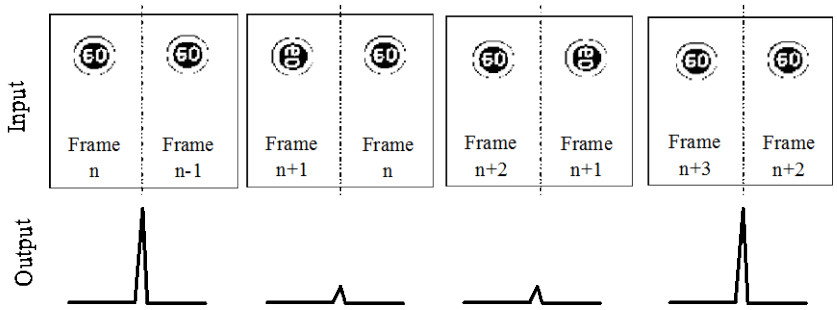
A variety of processing schemes have been tried. A 3x3 convolution binarisation scheme gives reasonable results, but the processing time is slow due to the complexity of the filtering algorithm. The success of this algorithm is due to fact that the 3x3 convolution is a form of edge enhancement, which enhances the spectrum and the noise in the dark areas. The best scheme used to binarise the spectrum was based on a nearest neighbour average comparison. The pixel to be binarised, P_{jk} is thresholded based on the average of its four nearest neighbours.

$$P_{j,k} = \begin{cases} +1 & \text{if} \quad P_{j,k} \geq \frac{1}{4}\left(P_{j-1,k} + P_{j+1,k} + P_{j,k-1} + P_{j,k+1}\right) \\ -1 & \text{Otherwise} \end{cases}$$

[*Some good answers with several candidates pointing out the benefits in terms of zero order. No one mentioned the benefits of speed with the binary phase modulation.*]

c) [30%] A very important observation is that the traditional 1/f JTC can be used as a 'recogniser' or comparator, where the reference image is unknown or unspecified. This completely reverses the role of the correlator and opens up a whole plethora of applications that revolve around object tracking and motion analysis. Rather than having a pre-defined target or reference image, the input is made up from a sequence of frames from a video source.

One elegant scenario is when a correlator is used to compare sequential frames in a video stream in a production line. In such an application, the current frame is the unknown and the previous frame is the 'reference'. Events that occur from frame to frame can now easily be tracked such as in an industrial inspection system as shown below. The current frame and previous frame are synchronised with the progress of objects through the system. If the sequence does not change, then the output correlations remain from frame to frame, however, when a change occurs (in the example a rotated roadsign), then the correlation between frames is interrupted. Moreover, the cycle of distortion can be detected by looking at the sequence of disturbances about the first detected defect.

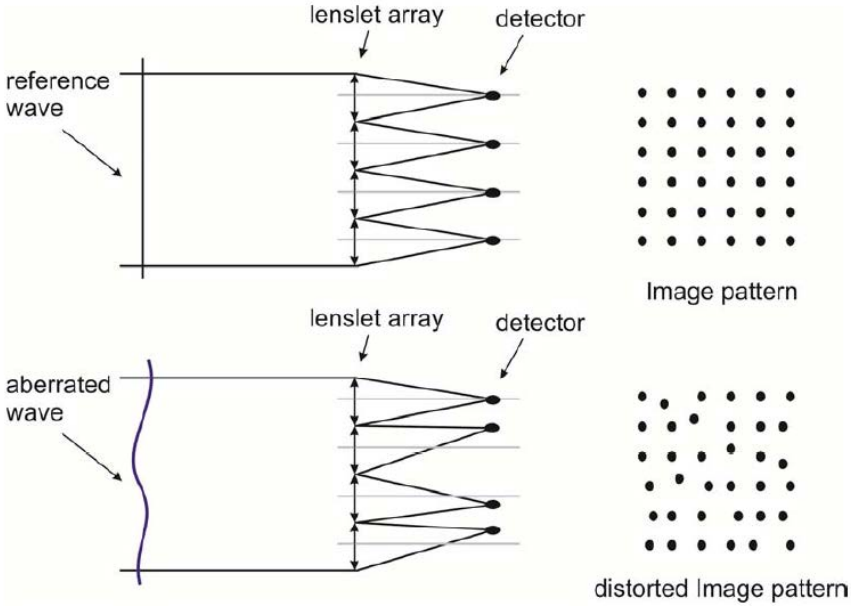


This frame by frame comparison system can be further extended to allow the correlator's inherent shift invariance to be exploited and a new dimension of object tracking added. If we use the comparator to compare two identical frames, then there will be a strong correlation, which will be centred in the middle of the output plane. If the object in frame *n*, shifts to the left, then the correlation between the two frames will also shift, and we can track the position of the object in frame *n* relative to its old position in frame *n-1*. Furthermore we can keep tracking the relative position of the object from one frame to the next, as long at it stay within the frame and does not rotate or change scale.

[*Varied answers here with some unrequired diagrams of the Matched filter. Not many explained the advantages of not having to calculate the filter well]*

Q4 a) [30%] One 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.

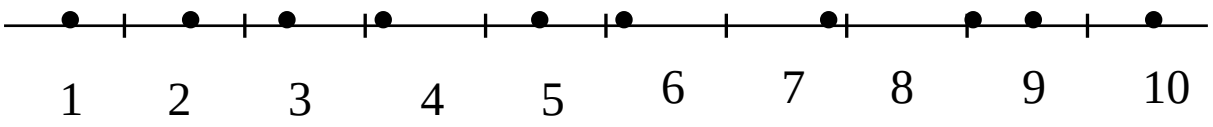


The main limitation comes from the zones defined by the lenslet array. Each lenslet samples a region of the incident wave. The spatial frequencies sensed by each zone will depend on its aperture. The more zones, the higher the frequencies of aberration detected, but also the higher the resolution of 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.

[*Mostly answered well, but most did not get the limitations due to the stroke of the wavefront and the overlap of spots.*]

b) [20%] 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 figure 2, we can estimate the approximate slope across each zone.

Zone	1	2	3	4	5	6	7	8	9	10
Slope(pi)	0	0	0.5	1.5	0	1.5	-1.5	-2	0	0



[*Quite a few did not realise that the slope of the wavefront changes sign after the peak, hence the spots will move in the opposite direction. Ie it's a bipolar system.*]

c) [30%] 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 correct for an aberration g(x,y) we have to include it in the algorithm steps above. Step 2, multiply the initial hologram by g(x.y). Step 6, when calculating the new replay field after the pixel flip, include the corresponding distortion at that pixel due to the aberration.

[*Well answered section*]

d) [20%] The function of the lenslet array can be performed by the hologram as Fresnel elements, These can be focussed onto a camera or a camera can be focused onto the replay field. If an SLM is used to display the hologram, then it can be programmed to display an hologram of an array of spots, one for each zone. Each zone is then imaged, and corrections are added to the hologram to see how the spot in that zone moves. Once the spot has been centered in the zone, the correction factor for the hologram can be read off as an estimate of the aberration.

[Good answers in this section showing a good understanding of the principles of a CGH.]