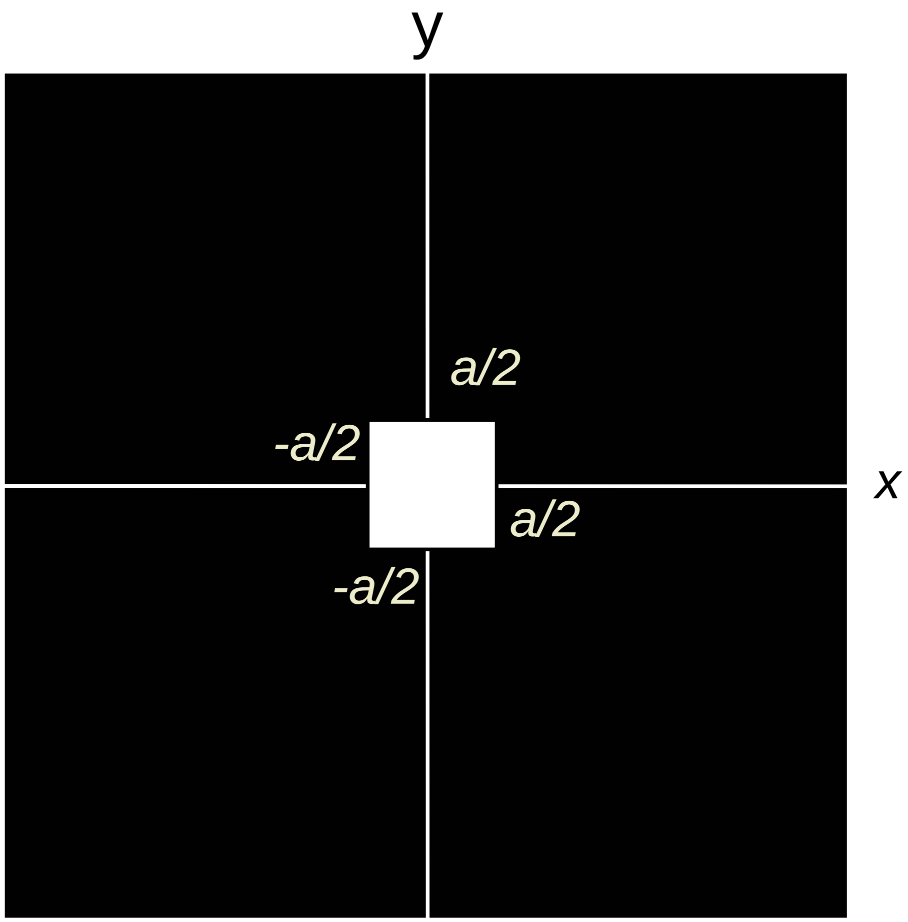
4B11 Photonic Systems

Tim Wilkinson

Lecture 2 ‘Holograms’

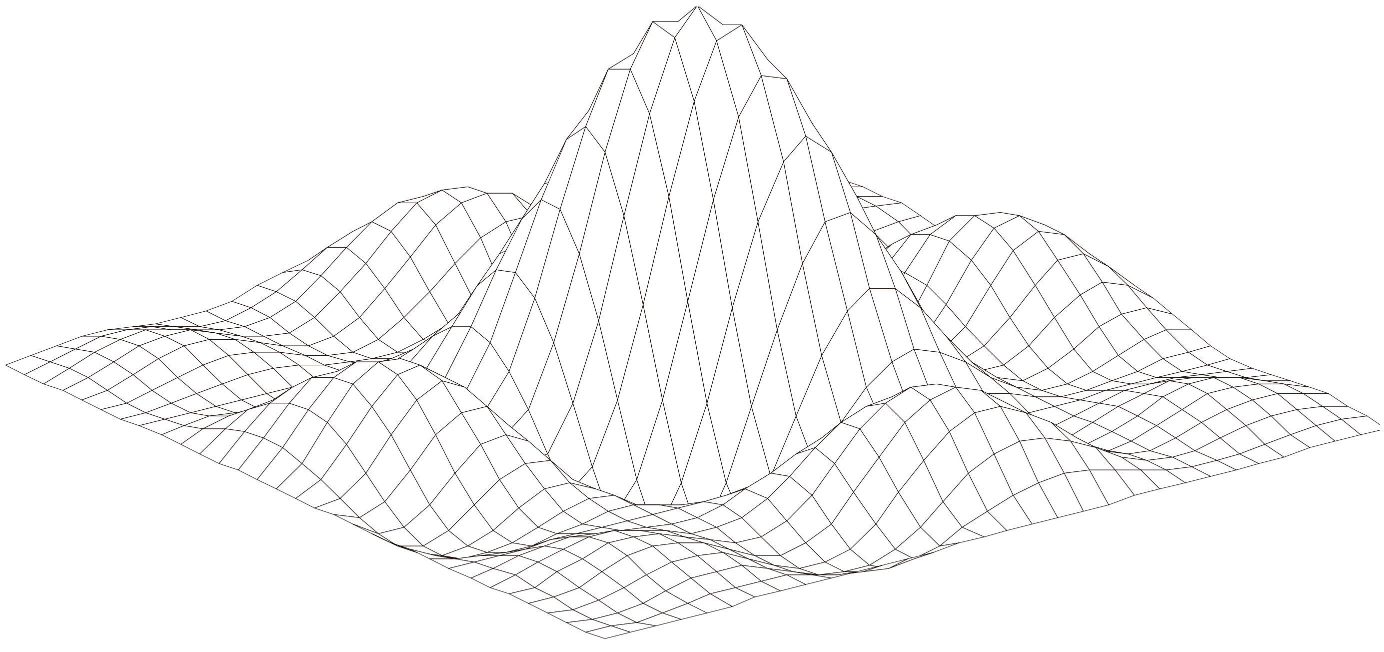
We want to calculate the far field or Fraunhofer region for a square aperture. This aperture can be represented in two dimensions as a ‘block’ function.

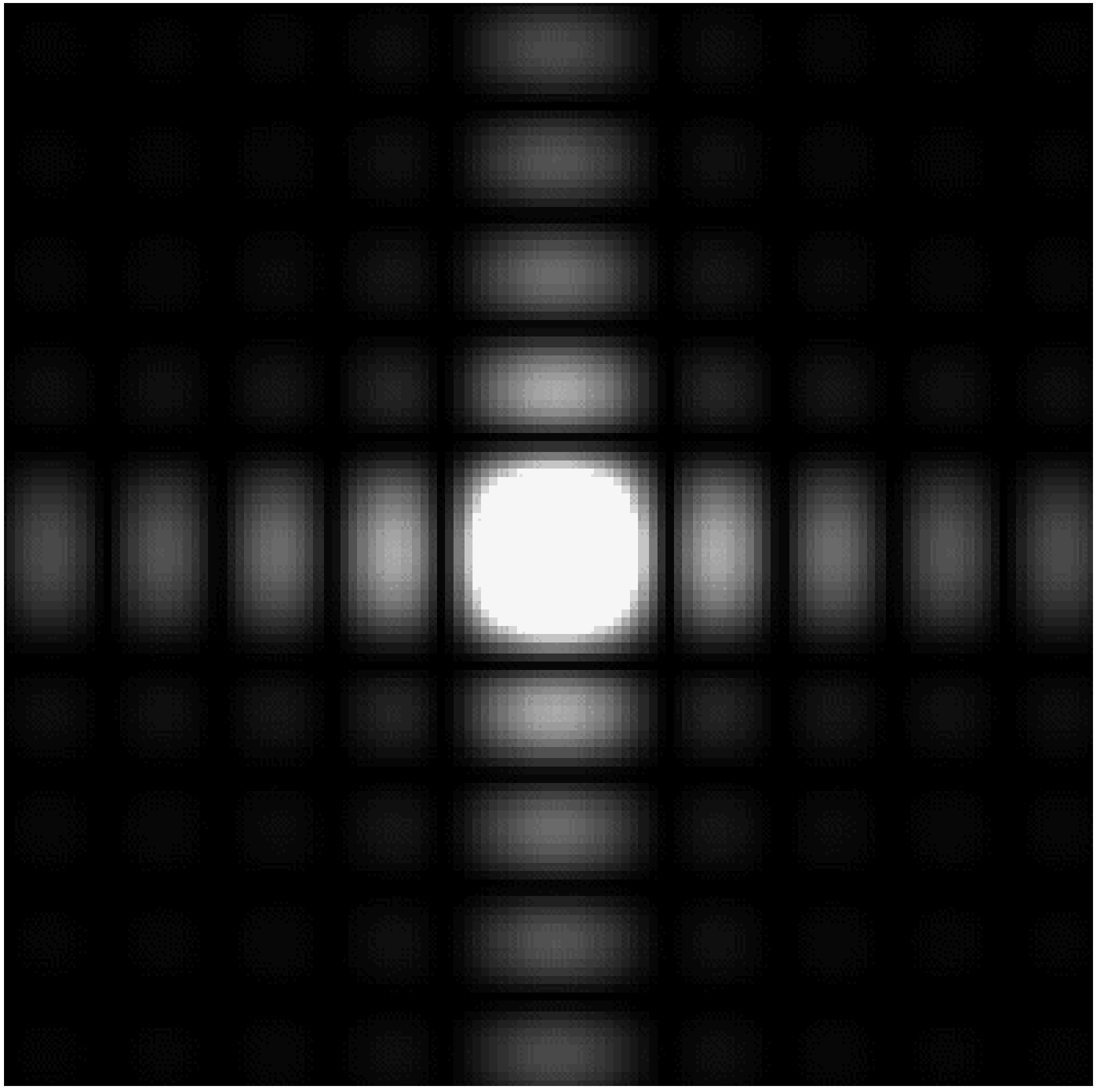


The far field of this aperture is its Fourier transform

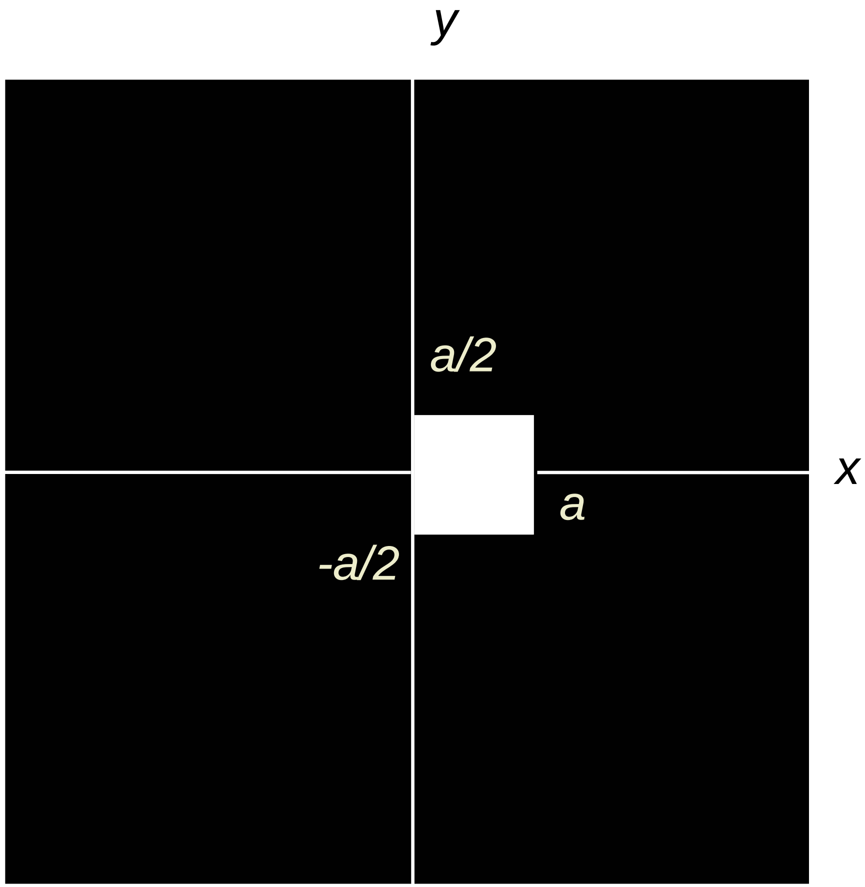
$$F(u,v) = Aa^2 \operatorname{sinc}(\pi a u)\operatorname{sinc}(\pi a v)$$

The far field is a 2-D sinc function





Now we will look at what happens when we shift the aperture by a distance *a*/2 from the origin of the plane



$$F(u,v) = Aa^2 e^{j\pi a u}\,\mathrm{sinc}(\pi a u)\,\mathrm{sinc}(\pi a v)$$

From this result we can see that the shift theorem for the 1-D FT also applies in 2-D.

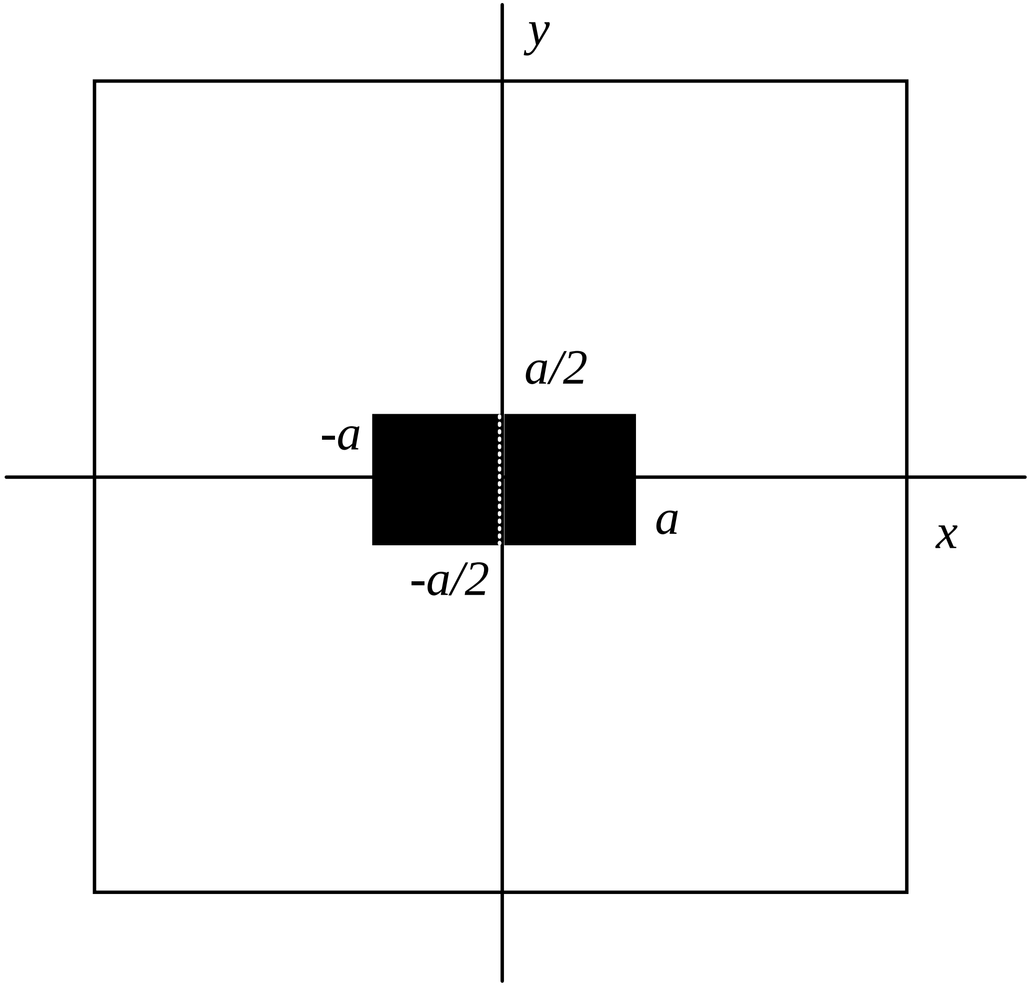
Hence an aperture shifted x_o, y_o from the origin adds an exponential phase term to the original 2-D sinc function of the centred aperture.

$$f\left(x - x_0, y - y_0\right) = F\left(u, v\right)e^{-j2\pi\left(x_0 u + y_0 v\right)}$$

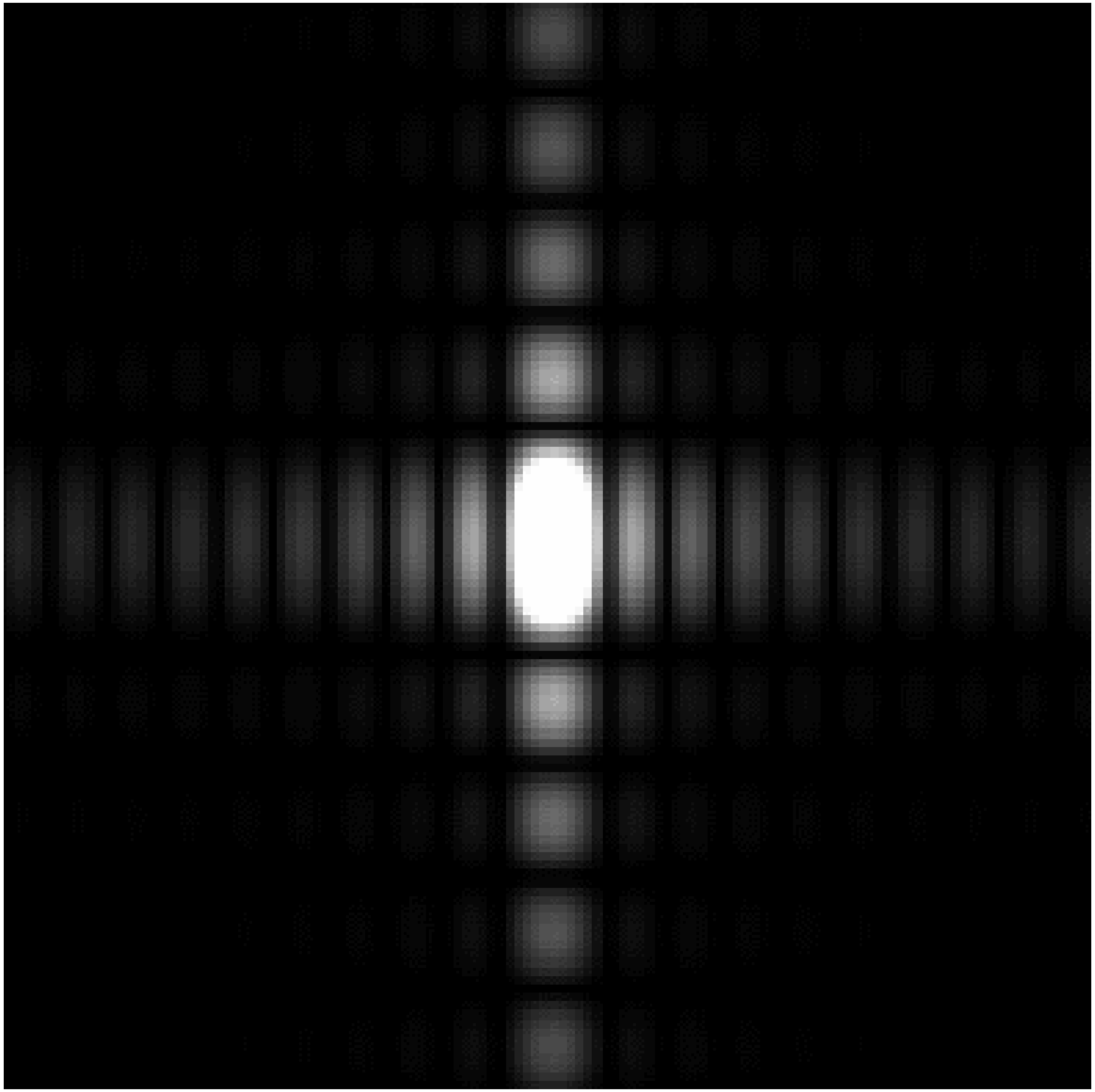
This is important as is demonstrates the shift invariant properties of the far field pattern. (The shape does not change, only the phase)

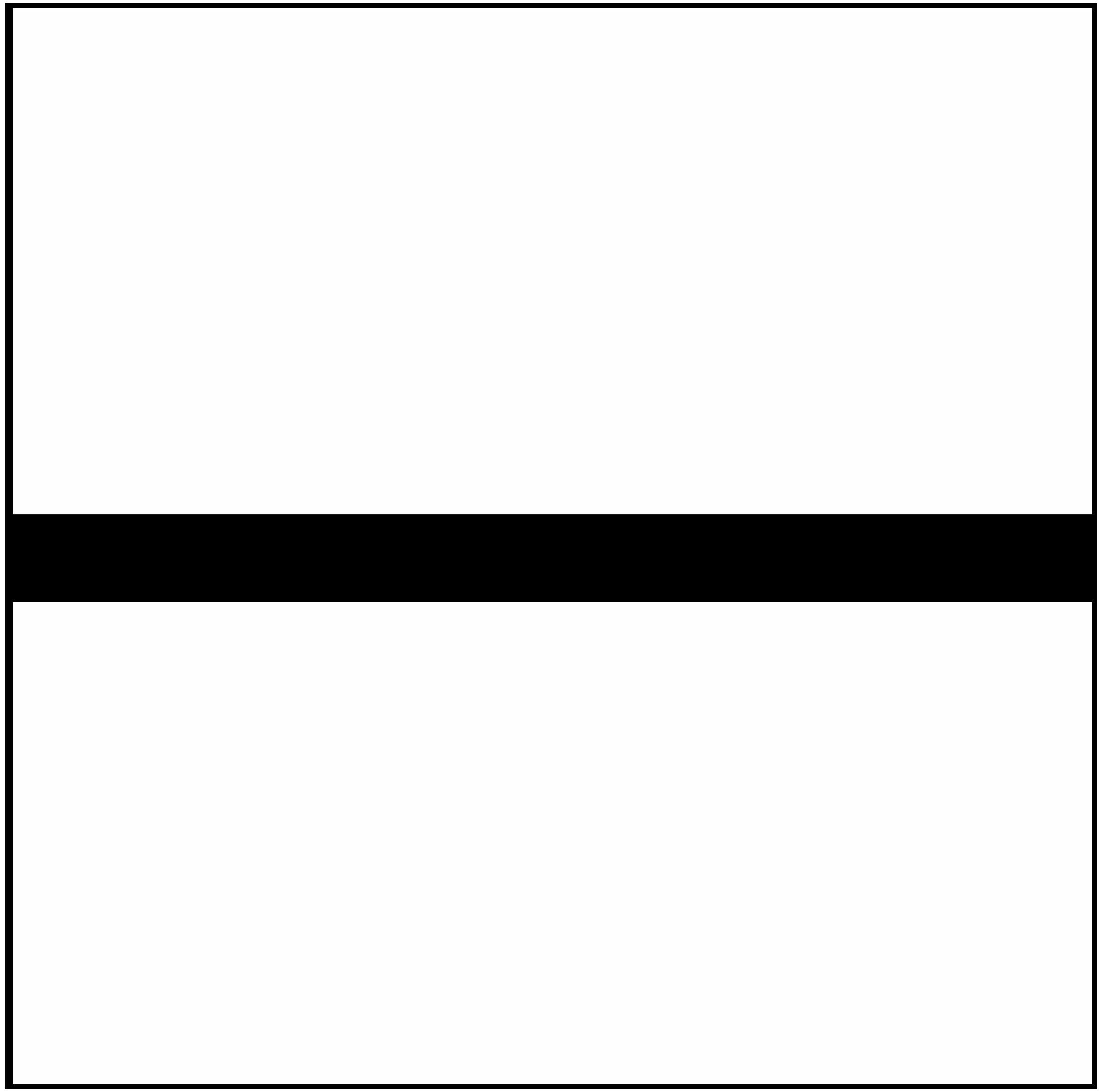
By altering the value of the amplitude A of each pixel, centred on a grid of interval b it is possible to add up the 2-D sinc functions and all the exponential phase terms due to the shifts to create an arbitrary 2-D distribution in the far field region.

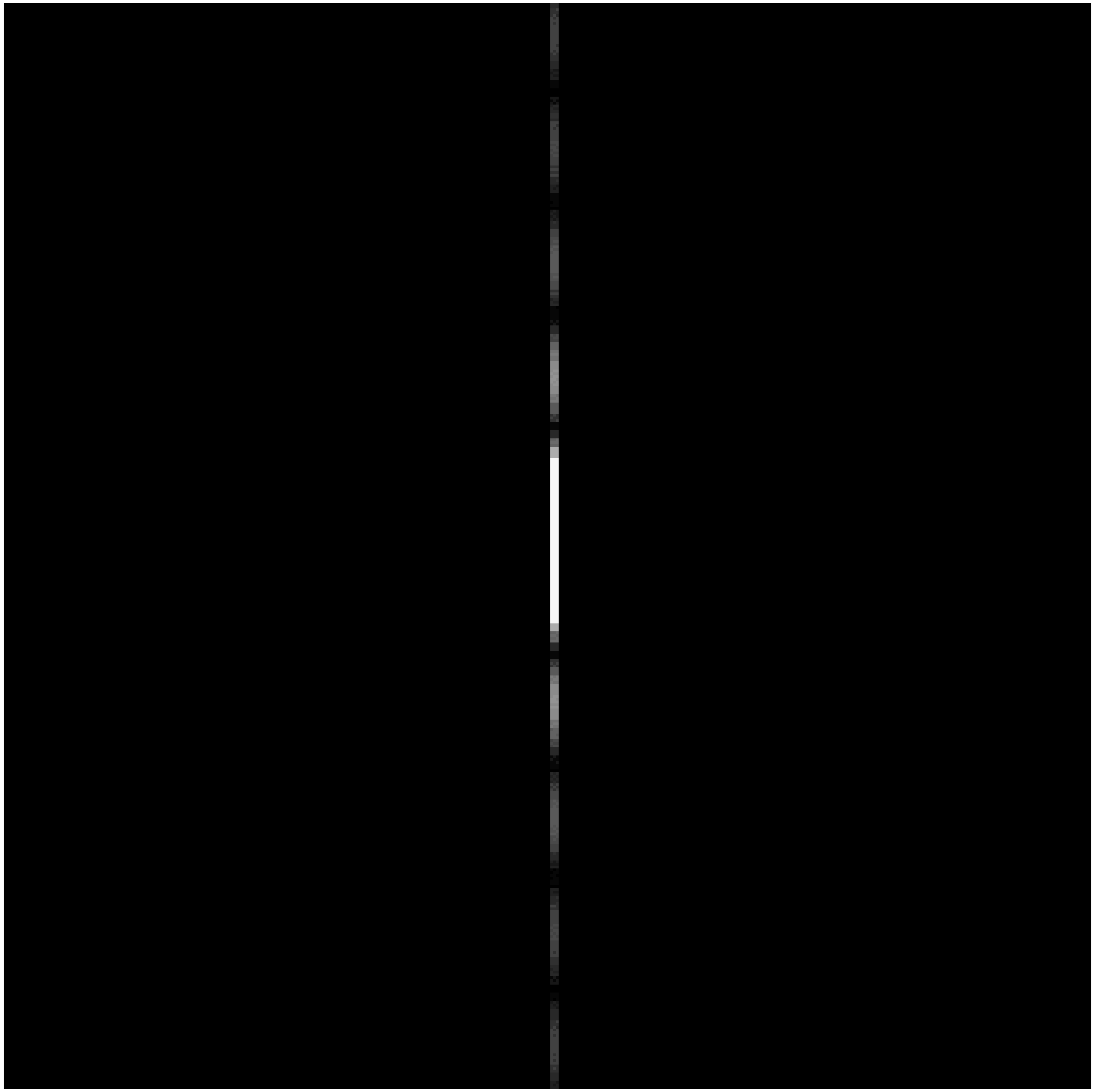
The exact structure of this distribution depends on the shape of the ‘fundamental’ **pixel** and the number and distribution of these pixels in the hologram.

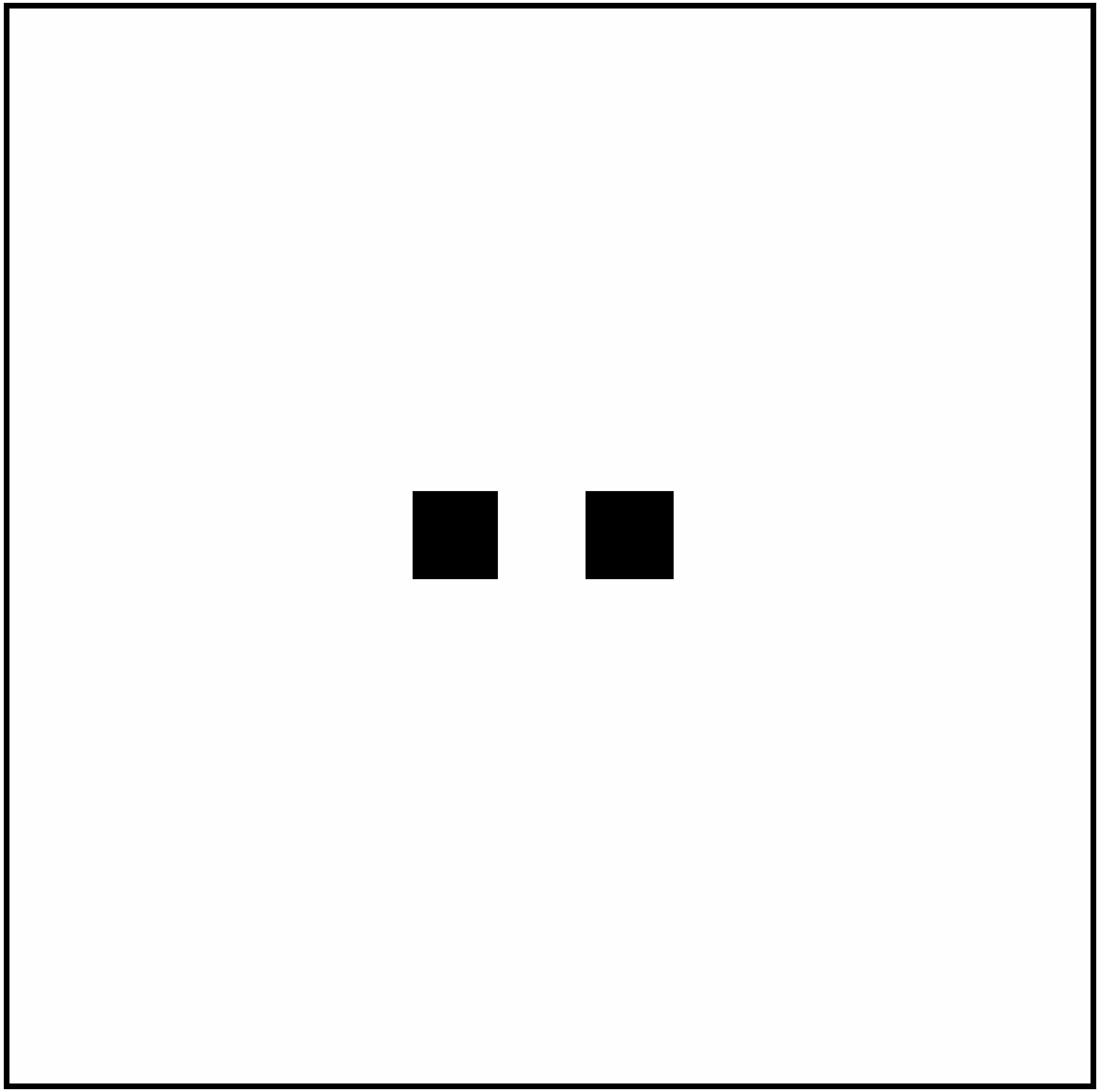


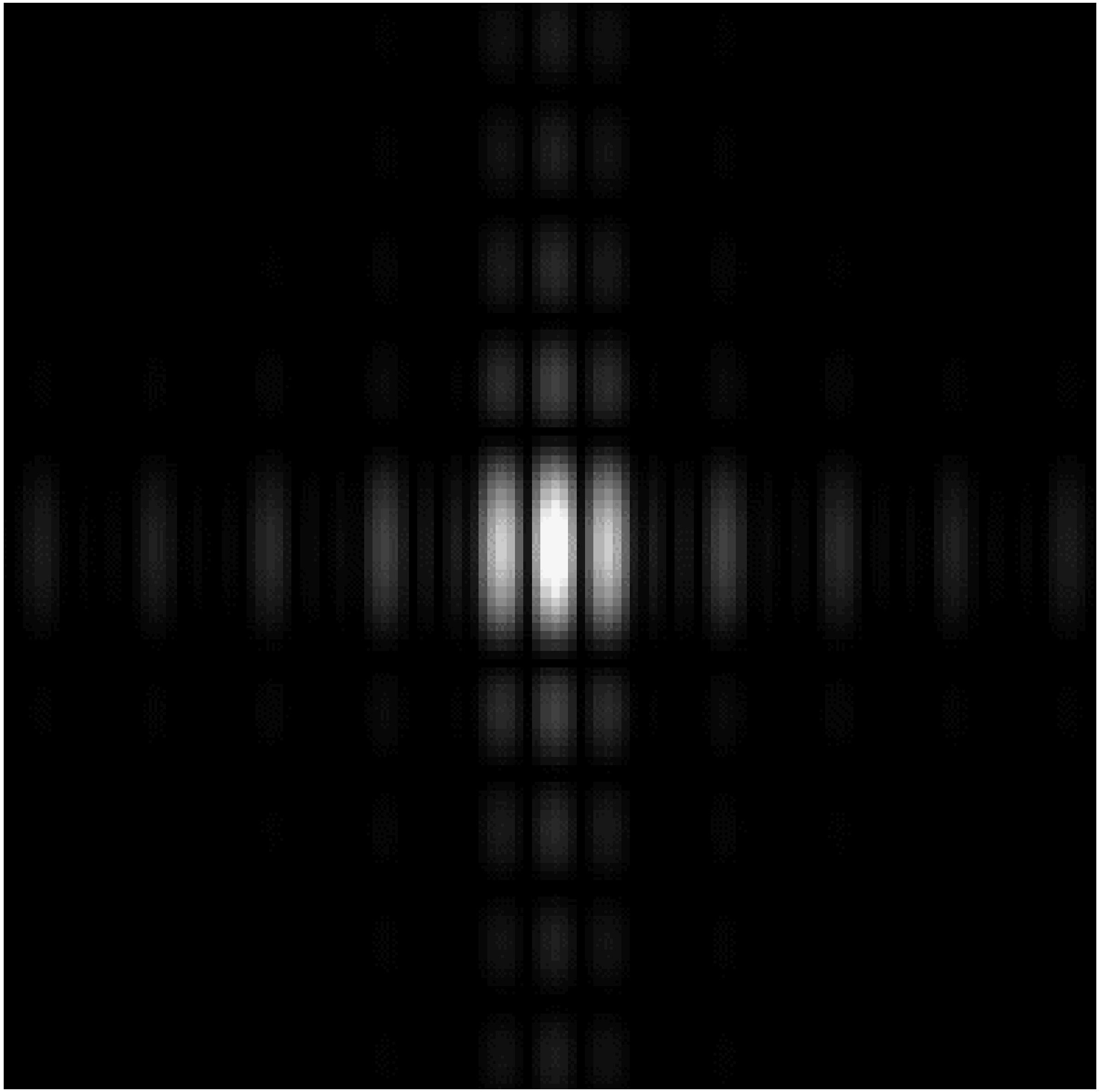
$$F(u,v) = 2Aa^2 \operatorname{sinc}(\pi 2au)\operatorname{sinc}(\pi av)$$

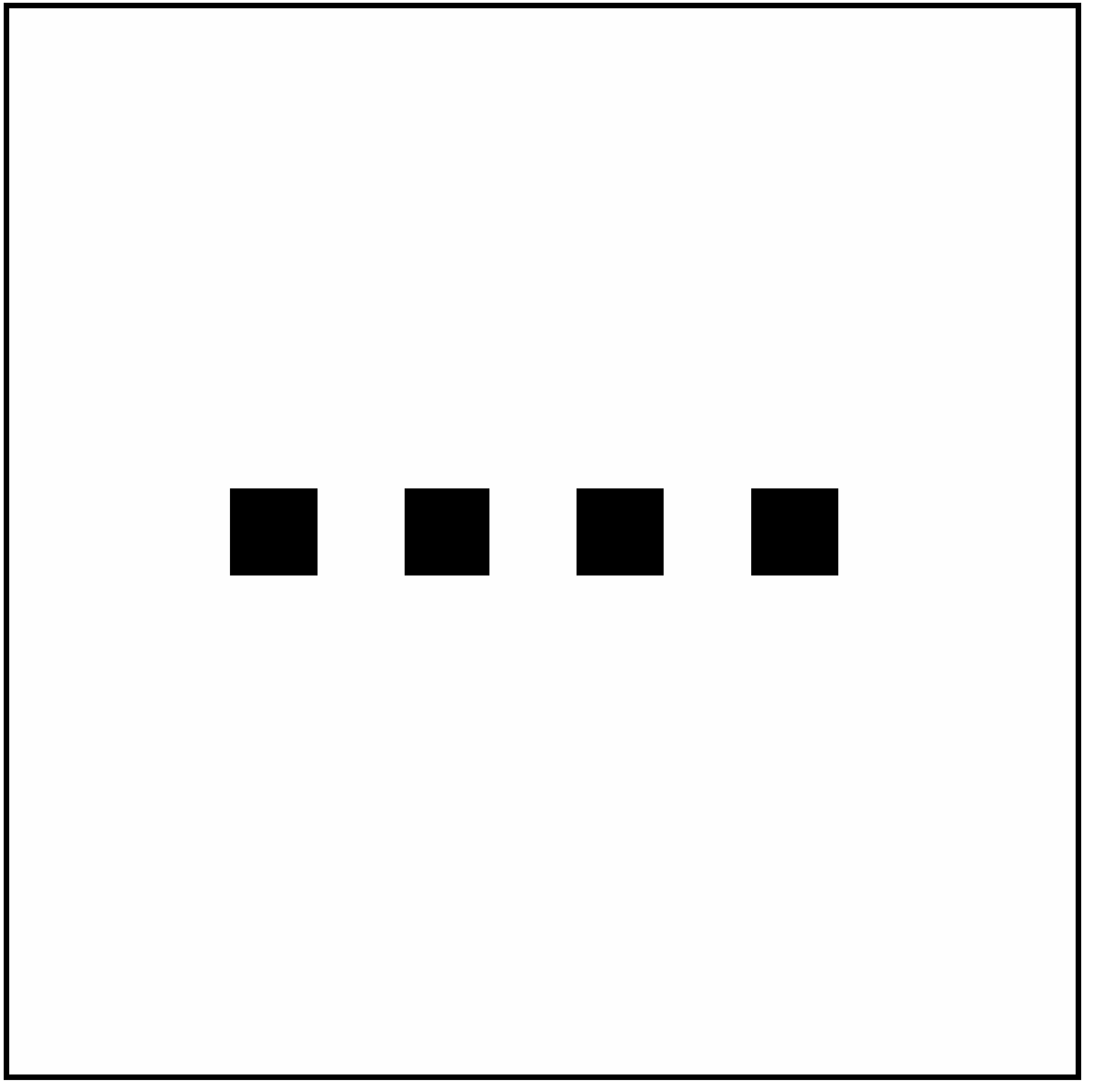


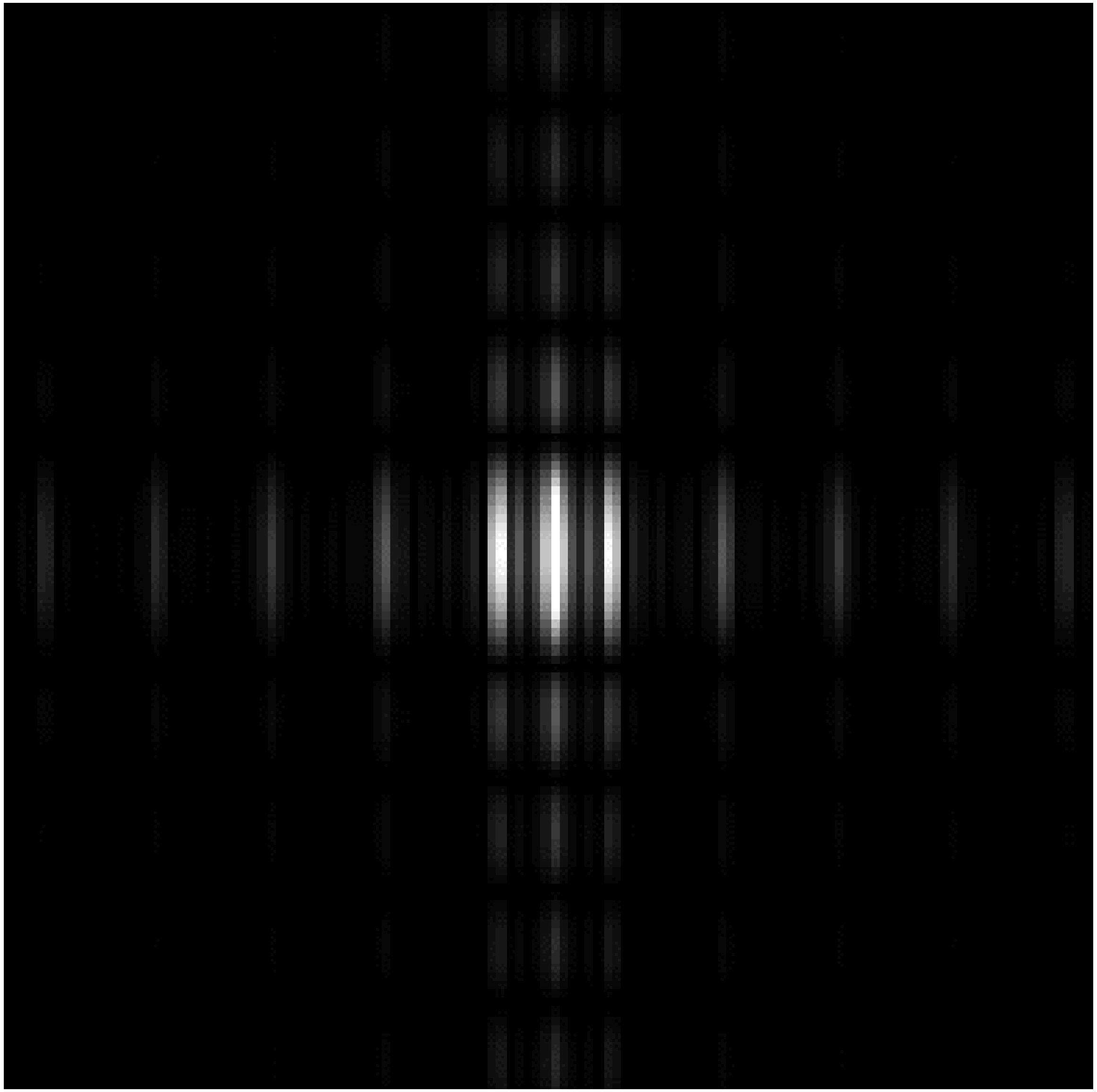


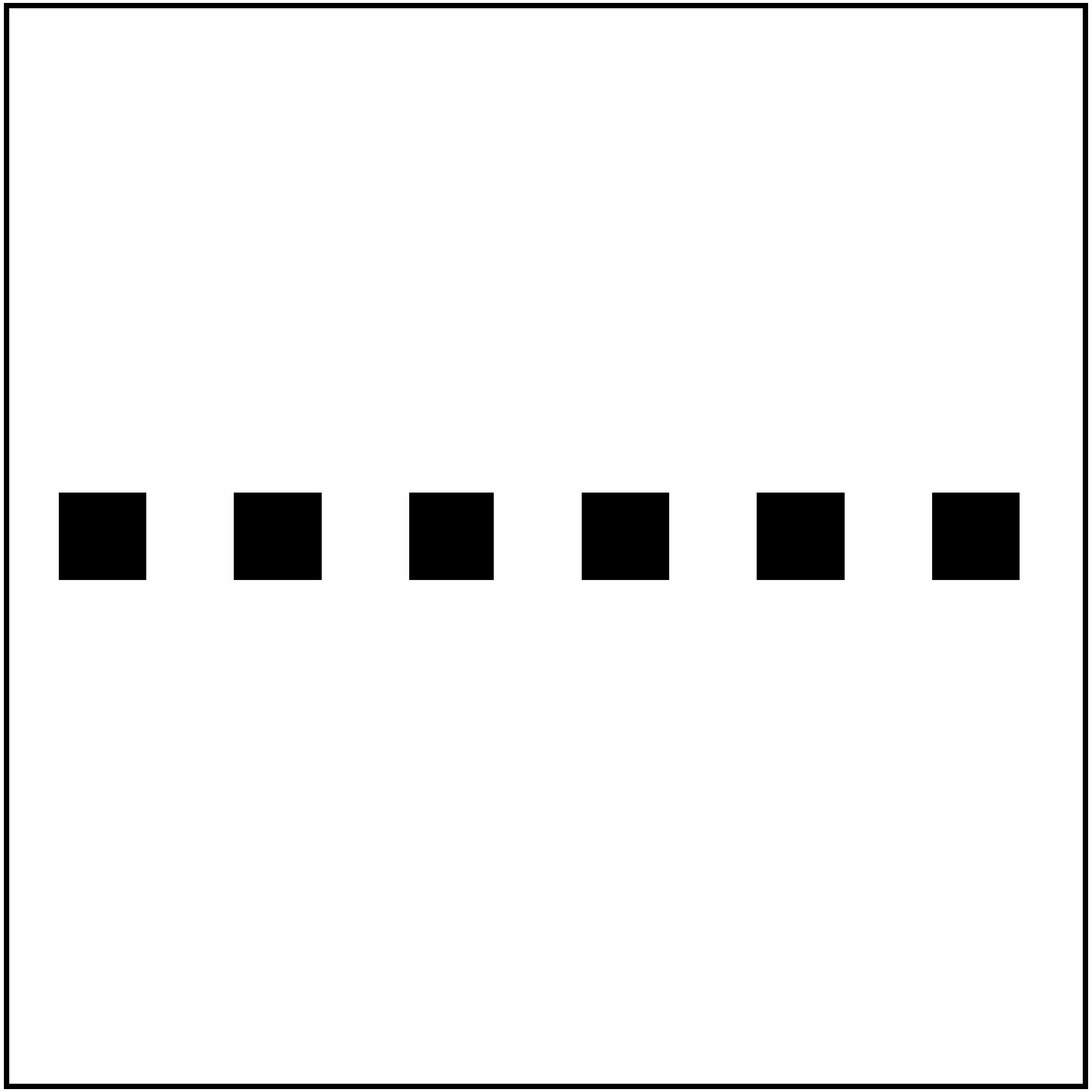


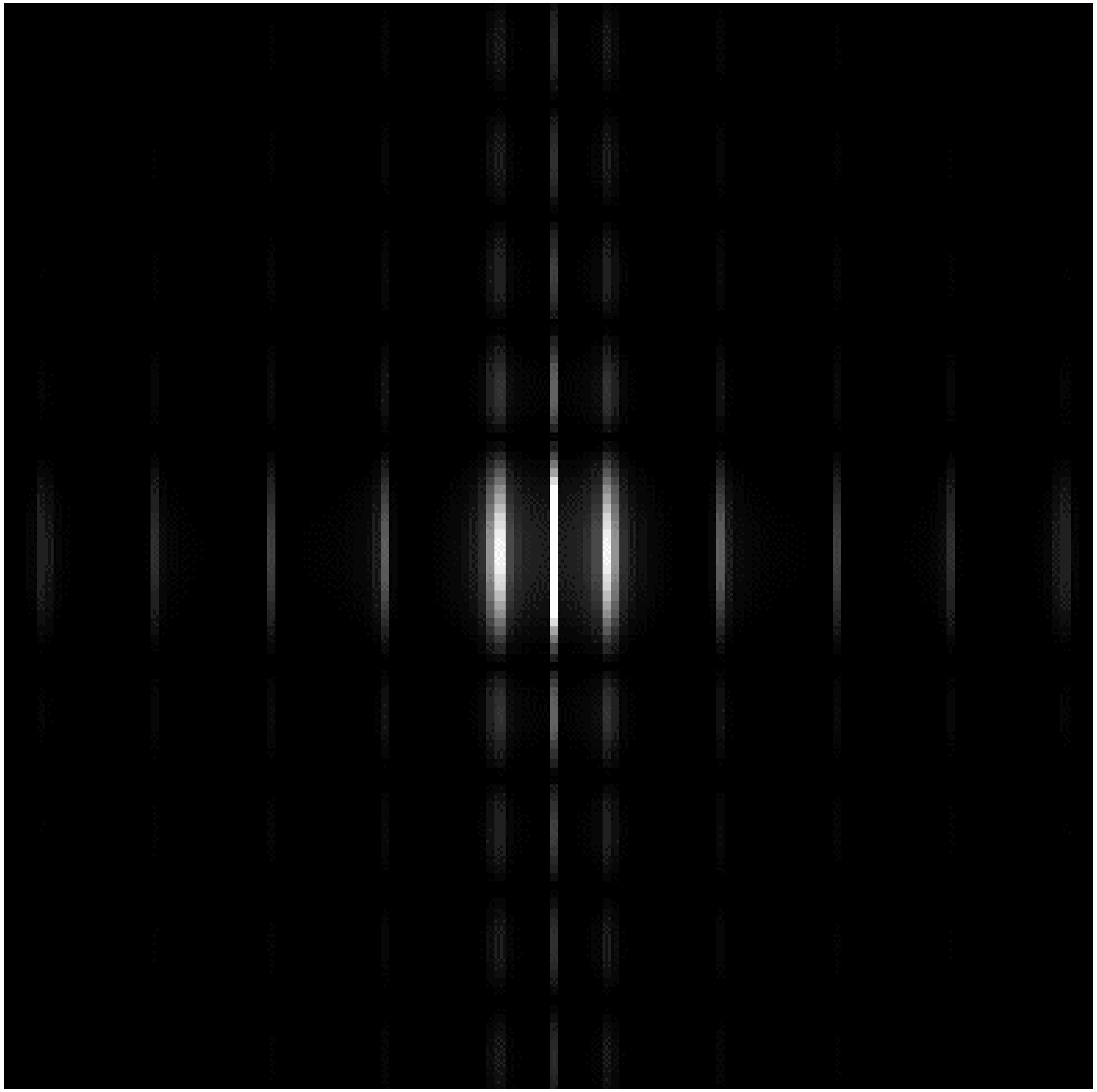


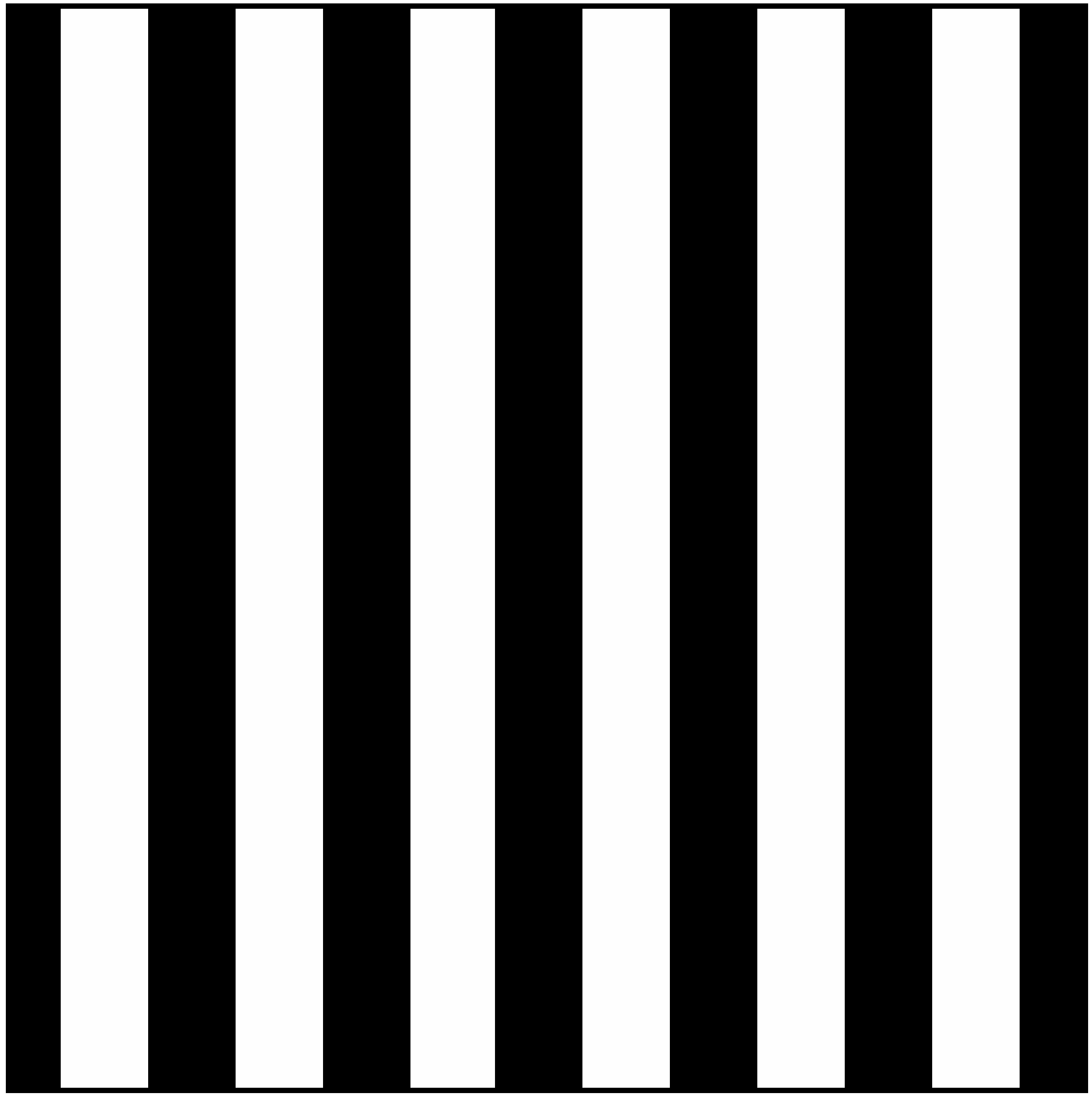


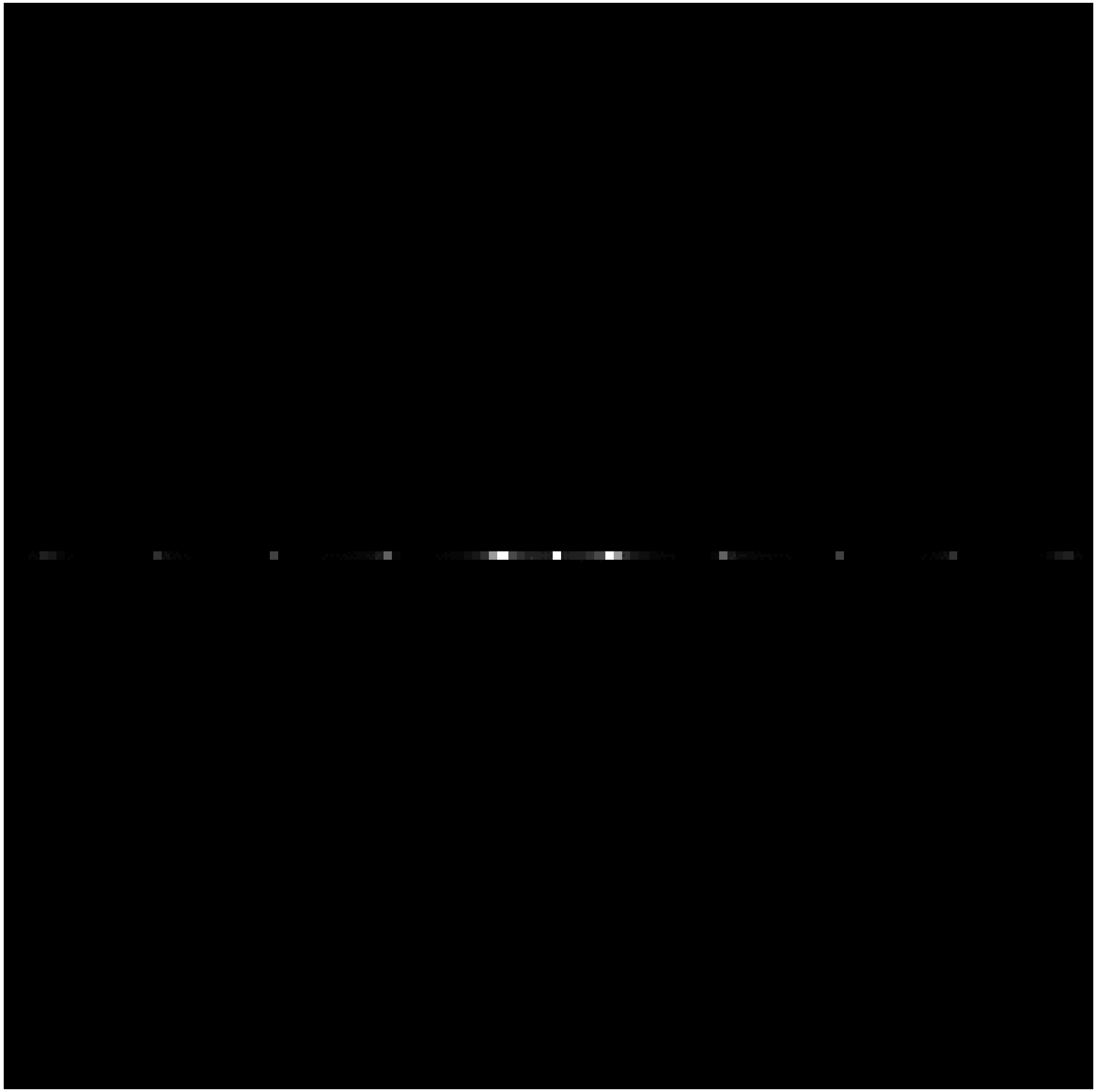


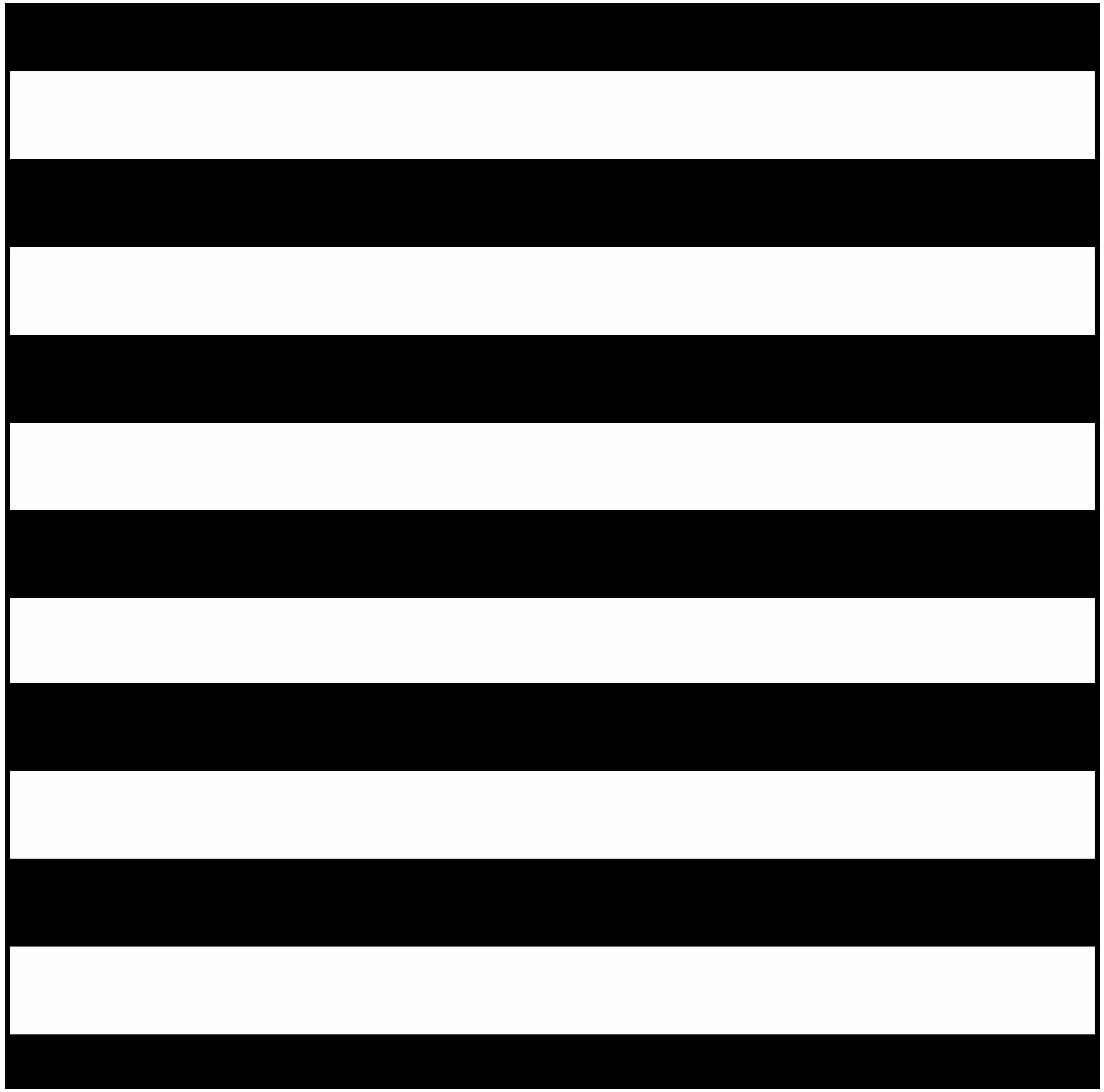


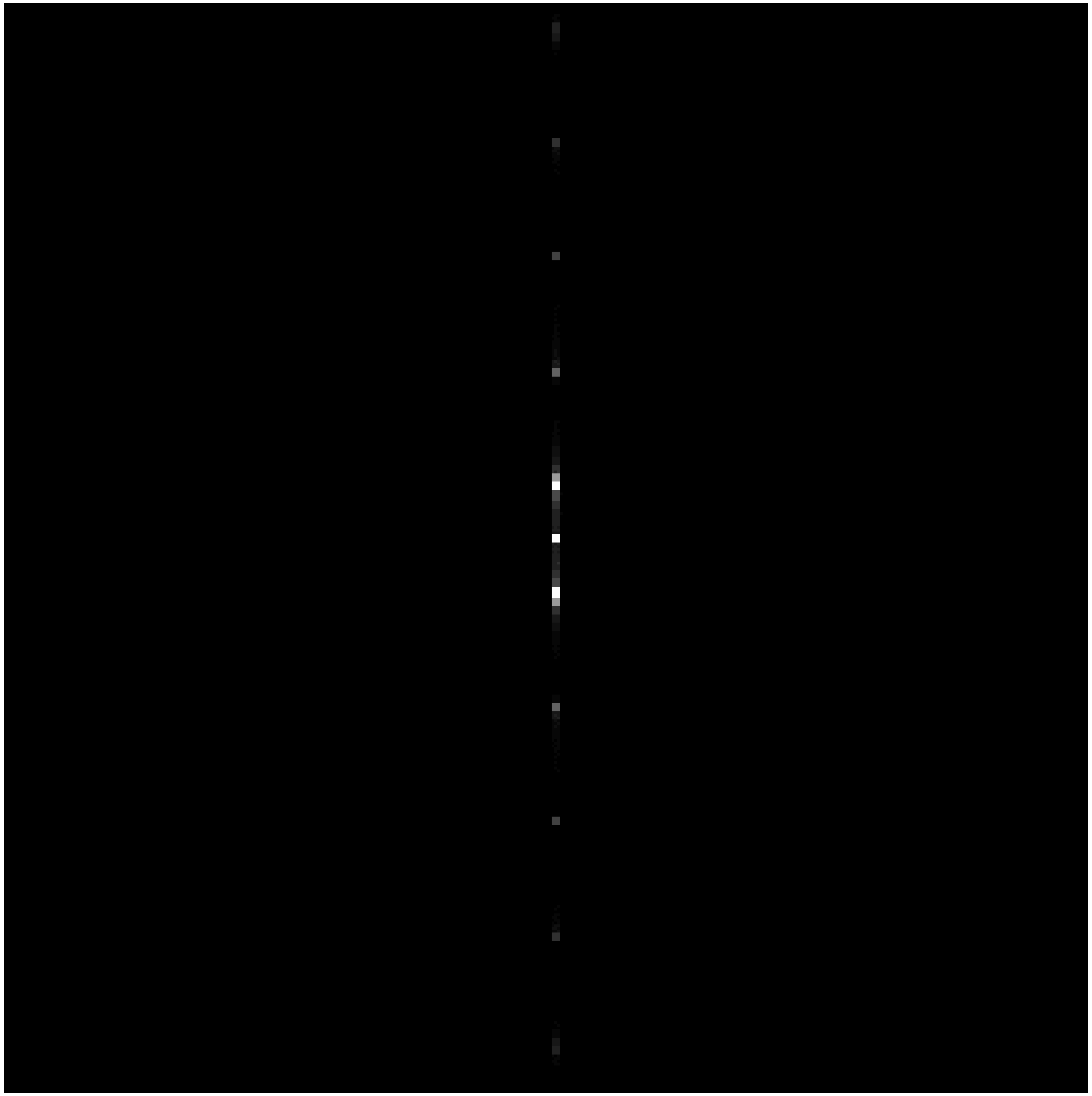




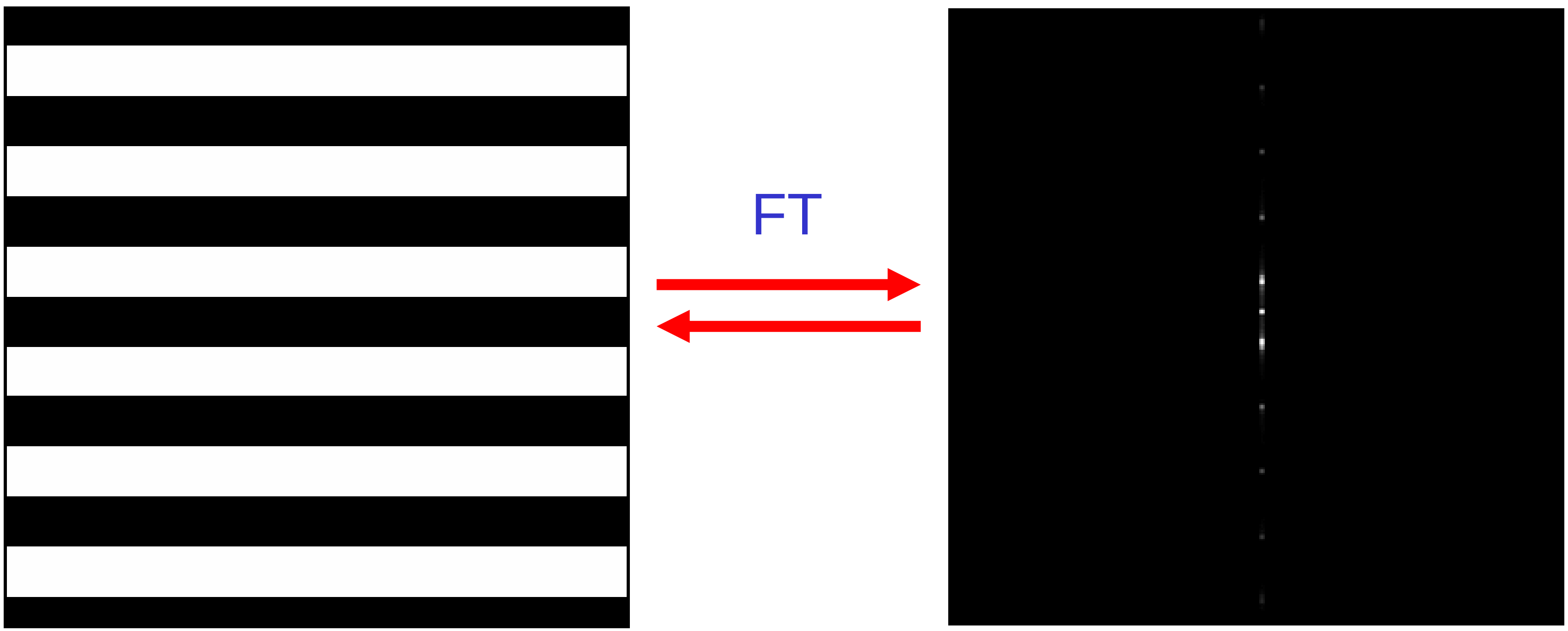








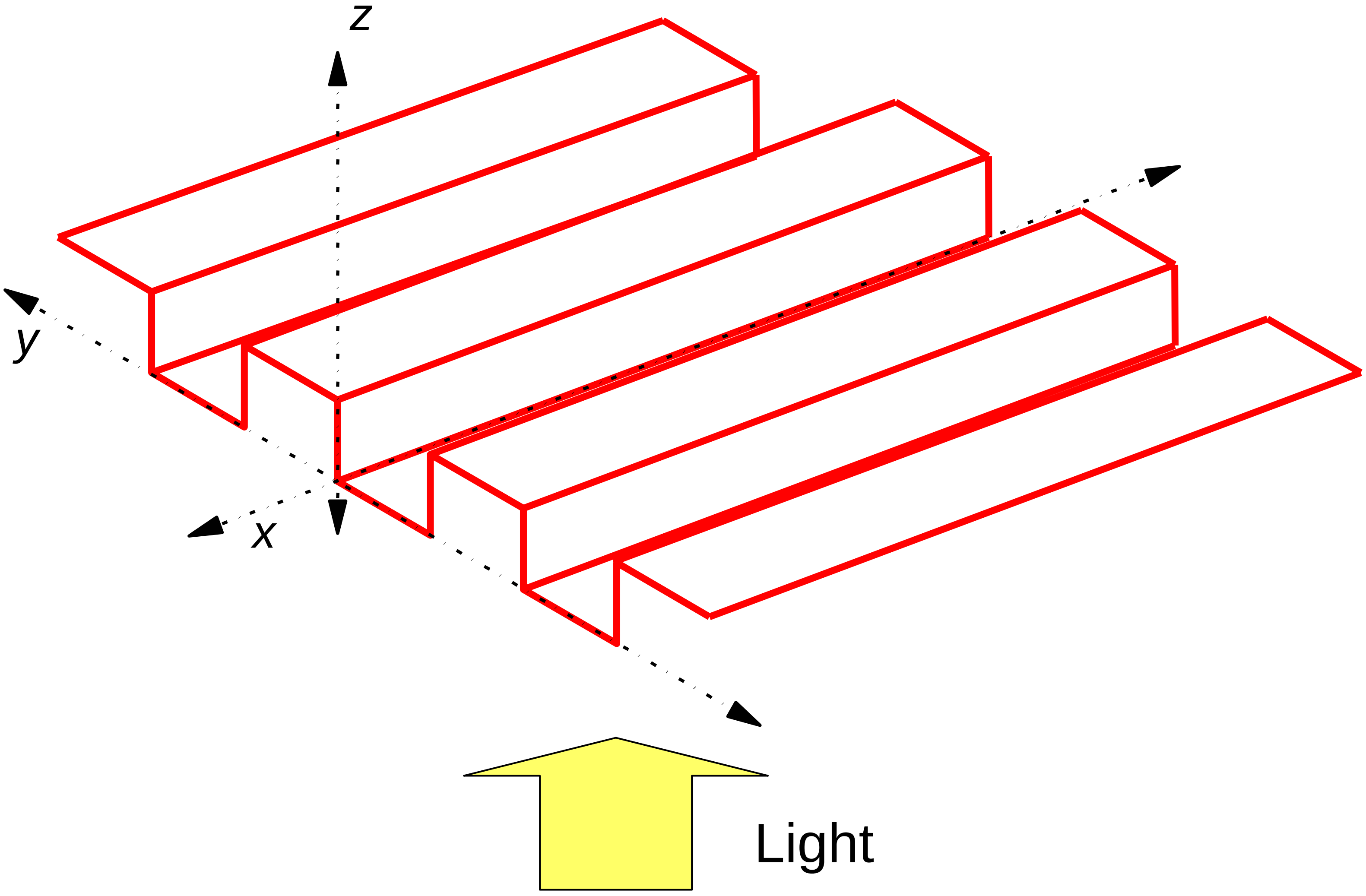
Terminology



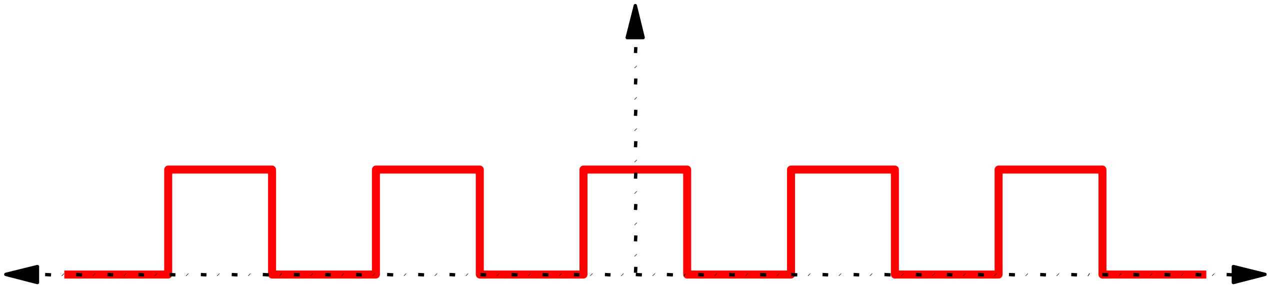
Hologram (or grating)

Replay field

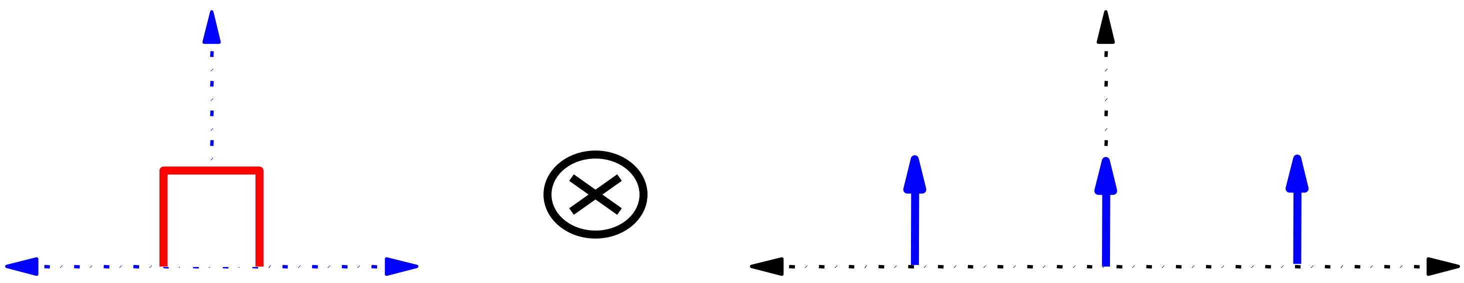
Another way of analysing a 2-D grating or square wave



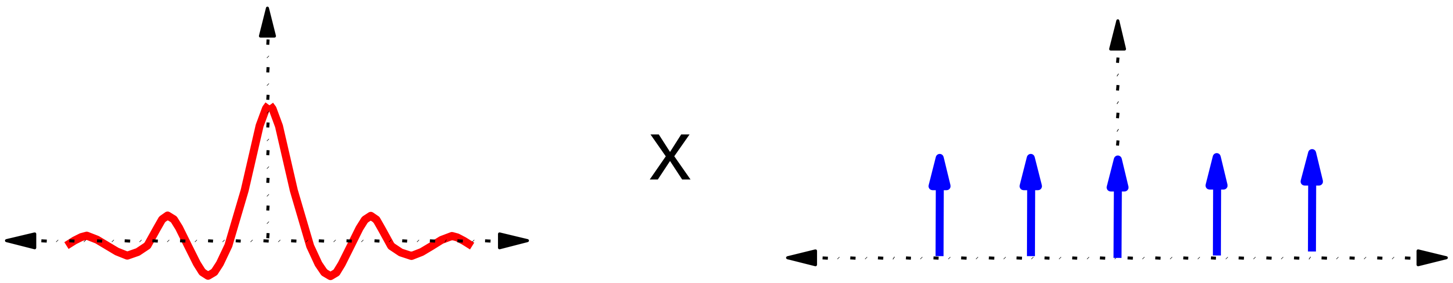
If viewed from the end (as a 1-D function)



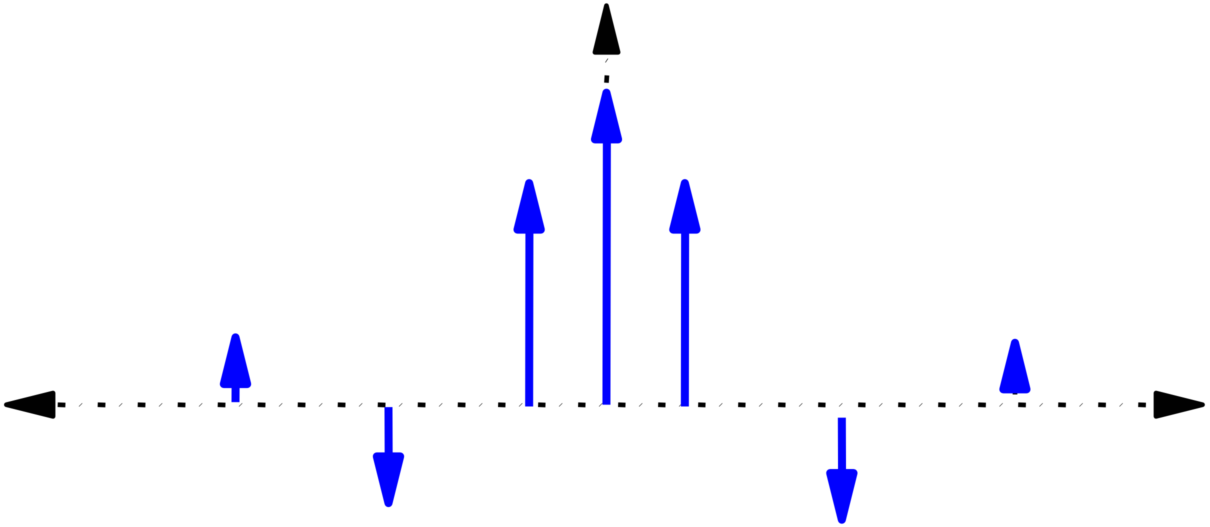
Which can be expressed as a convolution of two functions

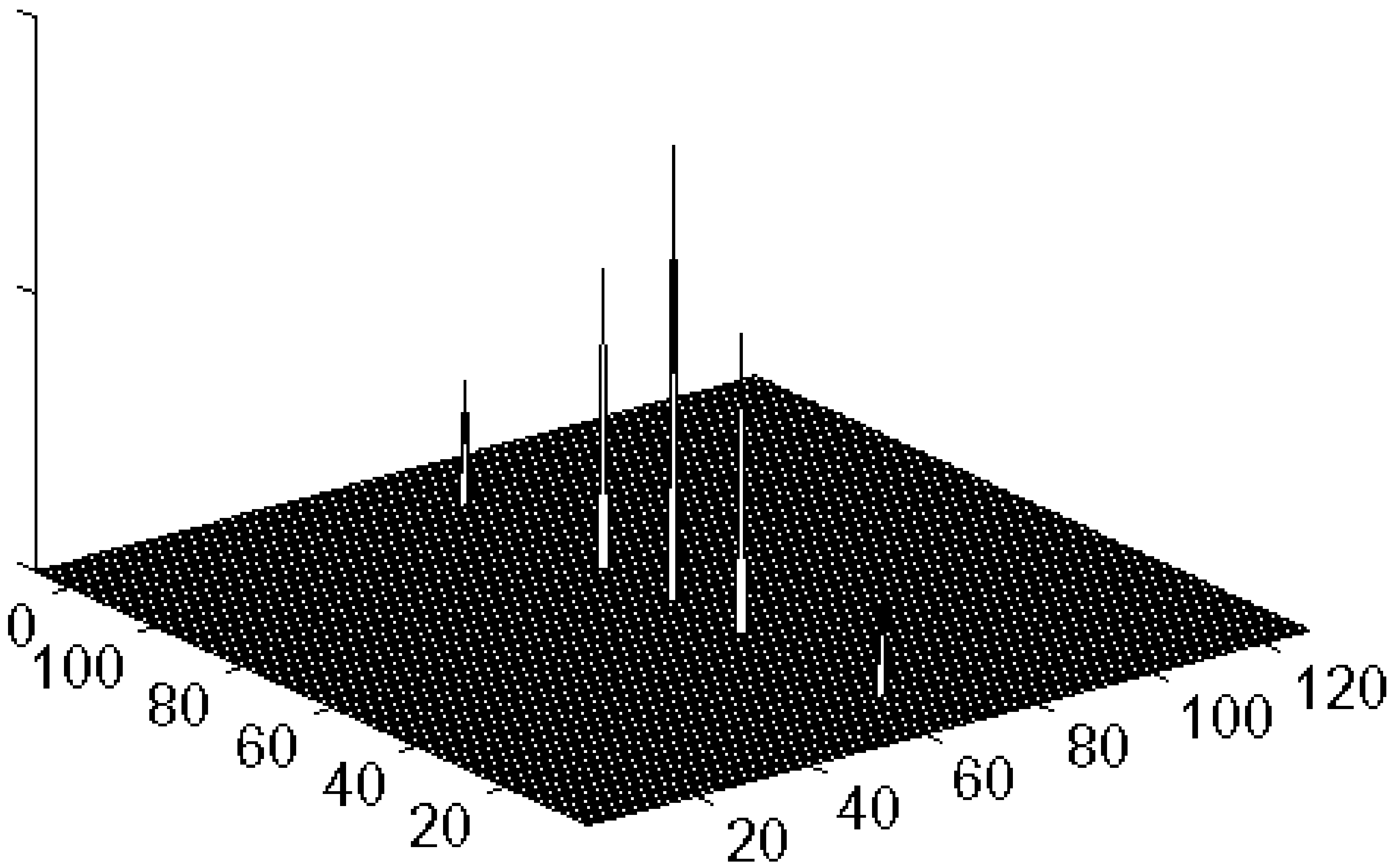


After the Fourier transform



Gives the final result



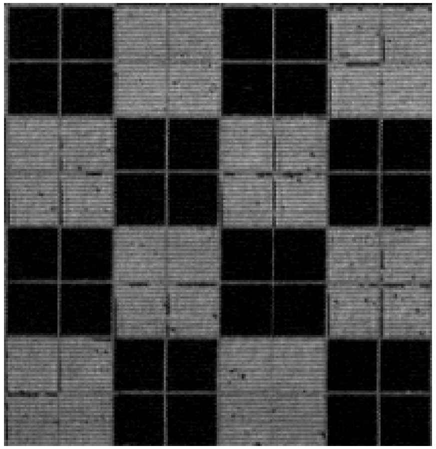

0
100
80
60
40
20
20
40
60
80
100
120

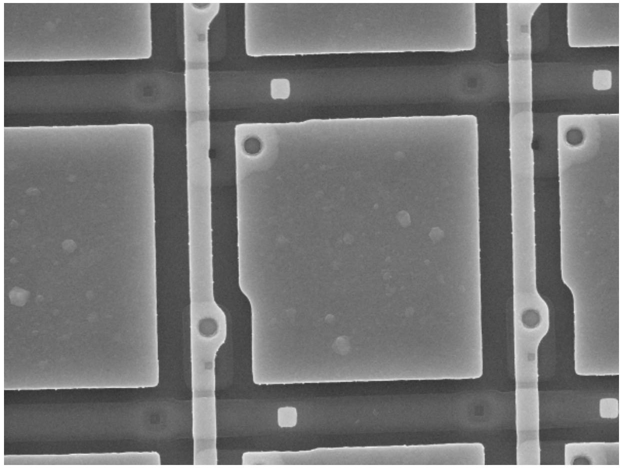
Effects of deadspace

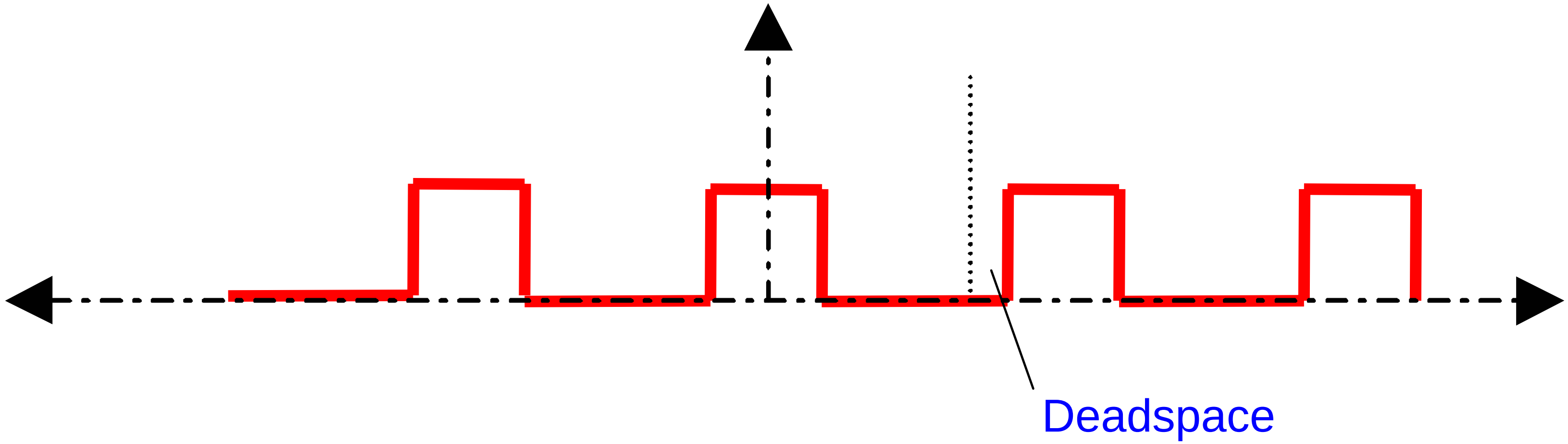
Up till now we have assumed that the pixel pitch always matches the finite width of the pixels themselves.

This is not the case in reality as each pixel must have a gap between itself and its neighbour to prevent short circuits and to allow for transistors etc to be used to drive the pixels.

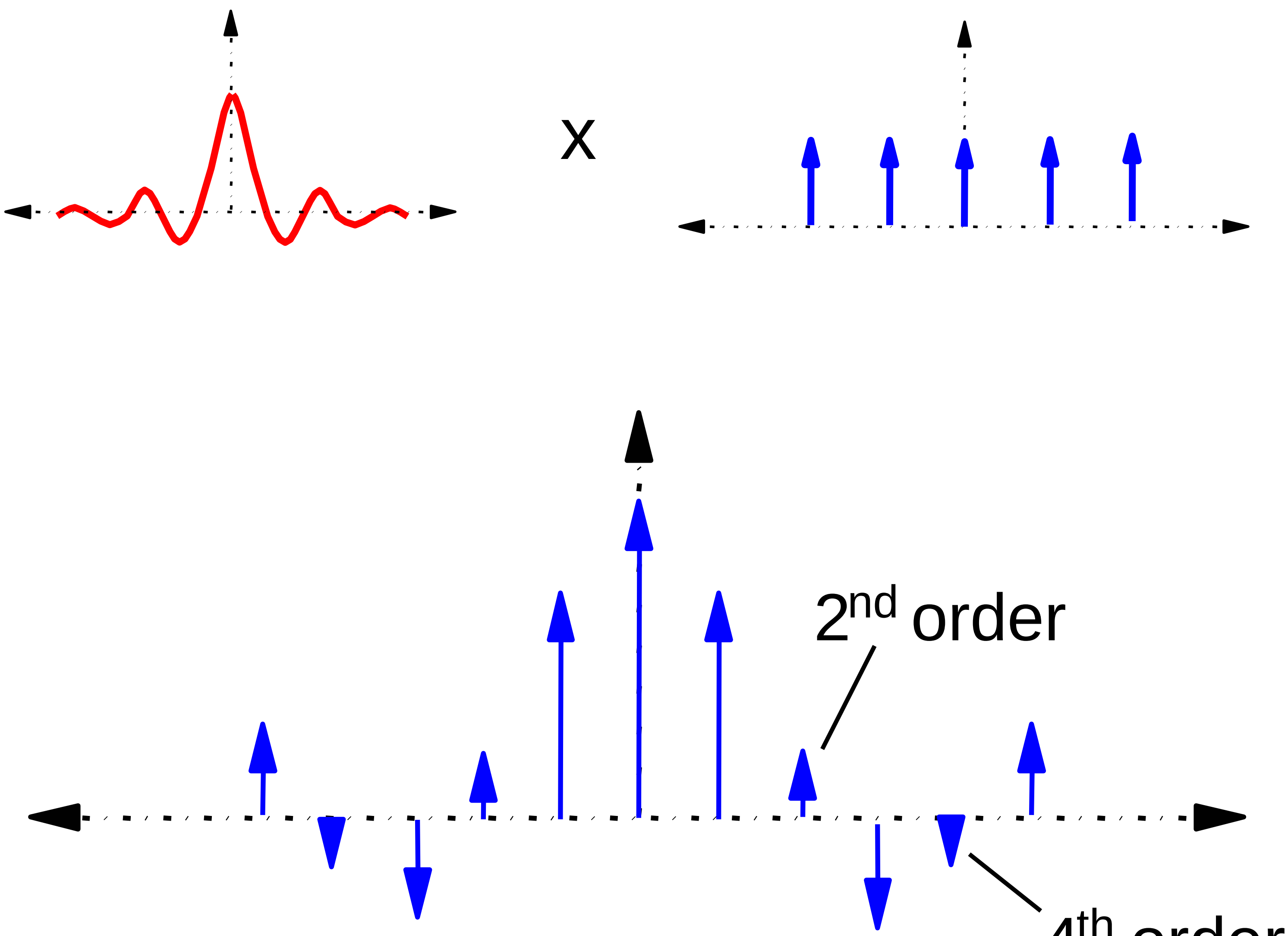
This region between the pixels is referred to as the deadspace and the ratio of pixel size to pixel pitch is called the fill factor







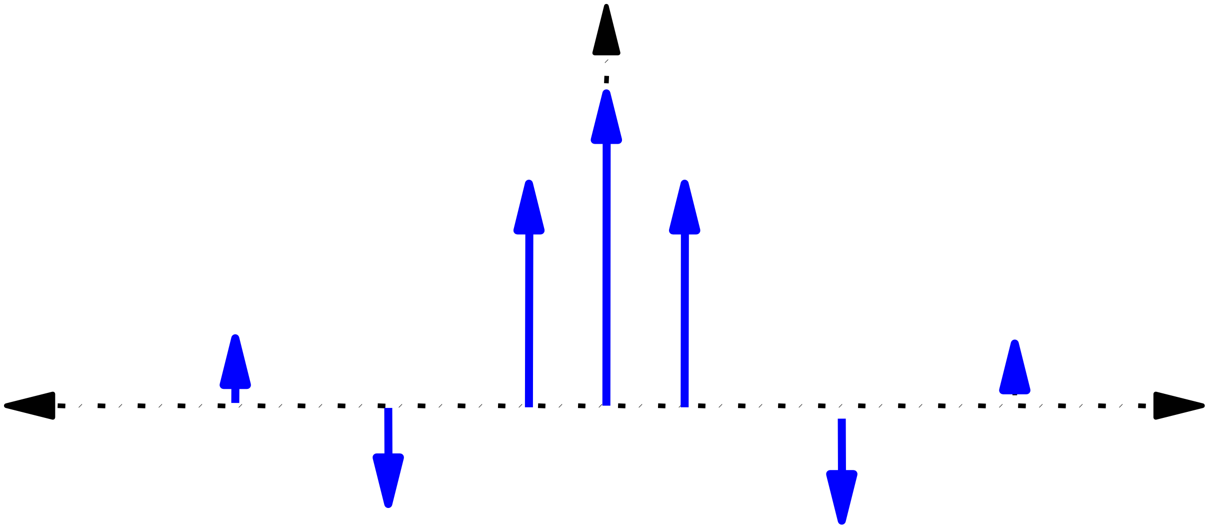
We can no longer assume that the zeros of the sinc envelope due to the fundamental pixel shape will be in the same spatial frequencies as the delta functions due to the repetition of the pixels, hence the second order is no longer suppressed by the sinc envelope.


x
2nd order
4th order

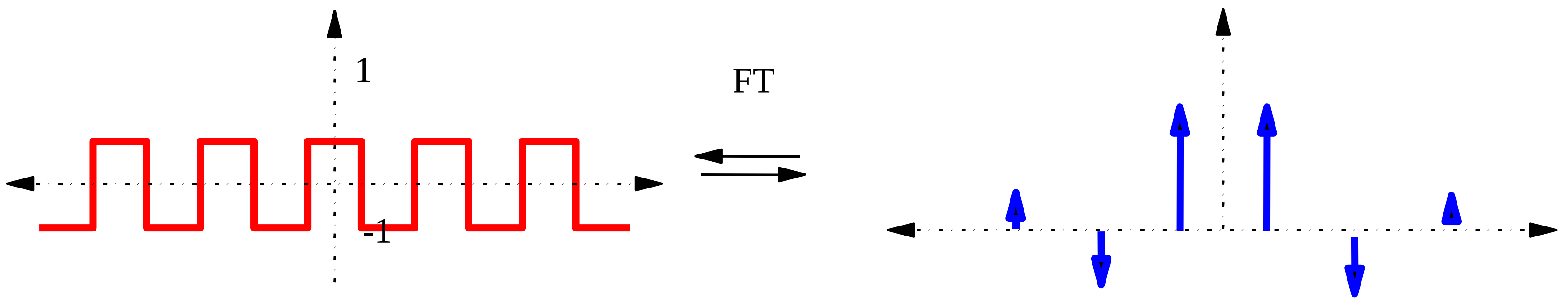
Modulation schemes

If we have amplitude pixels, $A \in [0,1]$, then the zero order (the energy at the central point or origin) of the replay field can only be zero if all values of A are zero.

This is because the point at the origin of the replay field (often called the zero order) is proportional to the average of the pixels in the hologram.



A better modulation scheme would be to have $A \in [+1,-1]$. If there are the same number of pixels set to +1 as set to -1, then the average will be zero and the zero order will be zero too.



Hence, with binary phase modulation ($A \in [+1,-1]$), the pixel in the centre of the replay field (the zero order) can be defined by the structure of the hologram.

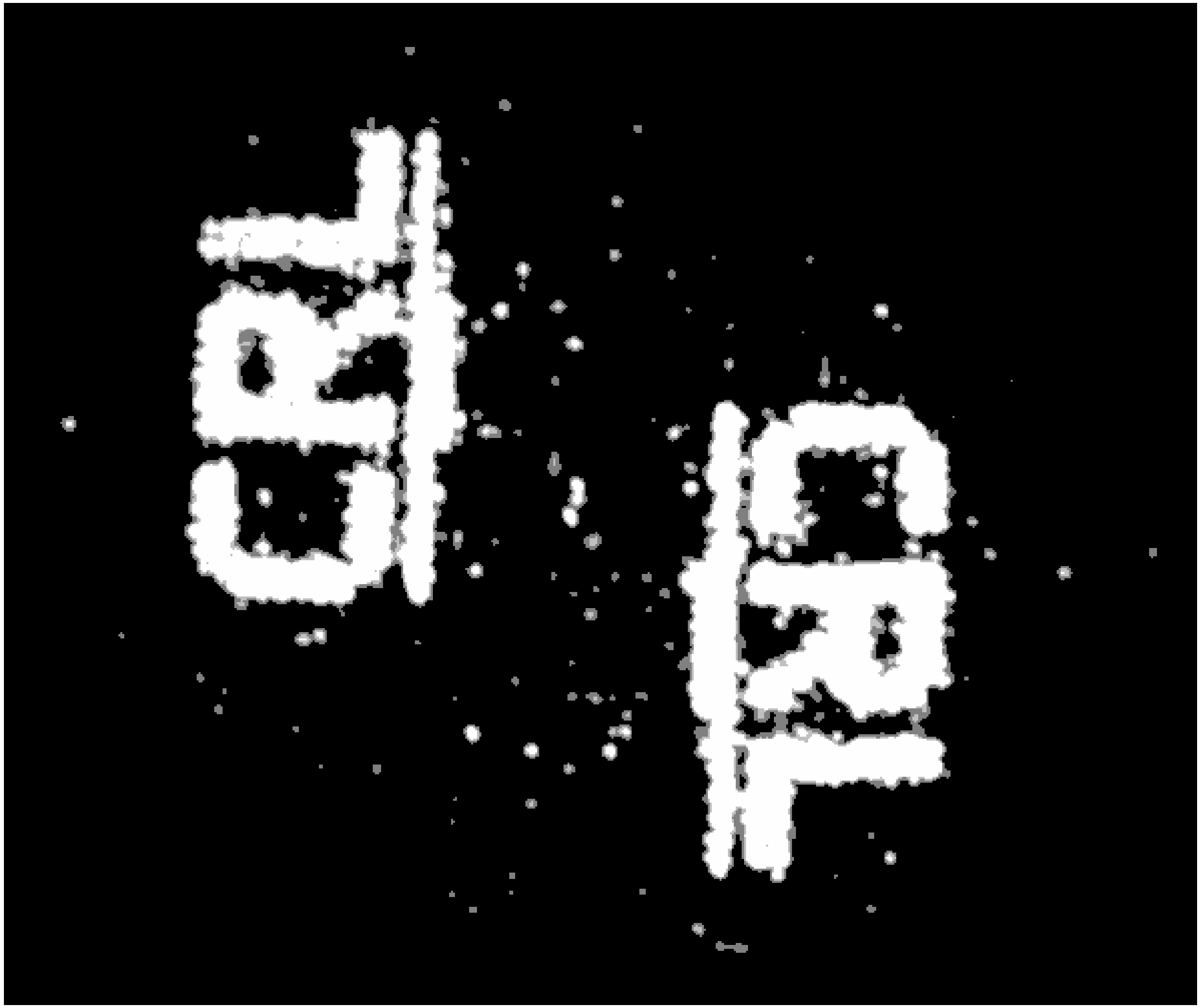
A drawback of both these binary modulation schemes is that the hologram will always be a real function.

$$F_T\left[h(x,y)^*\right]=H(-u,-v)$$

If function *h(x,y)* is real then *h(x,y)* = h(x,y),* hence we cannot differentiate between *H(u,v)* and *H(-u,-v)* which means that both must appear in the replay field.

Any replay field generated by a purely binary phase or amplitude hologram will always have 180º rotational symmetry.

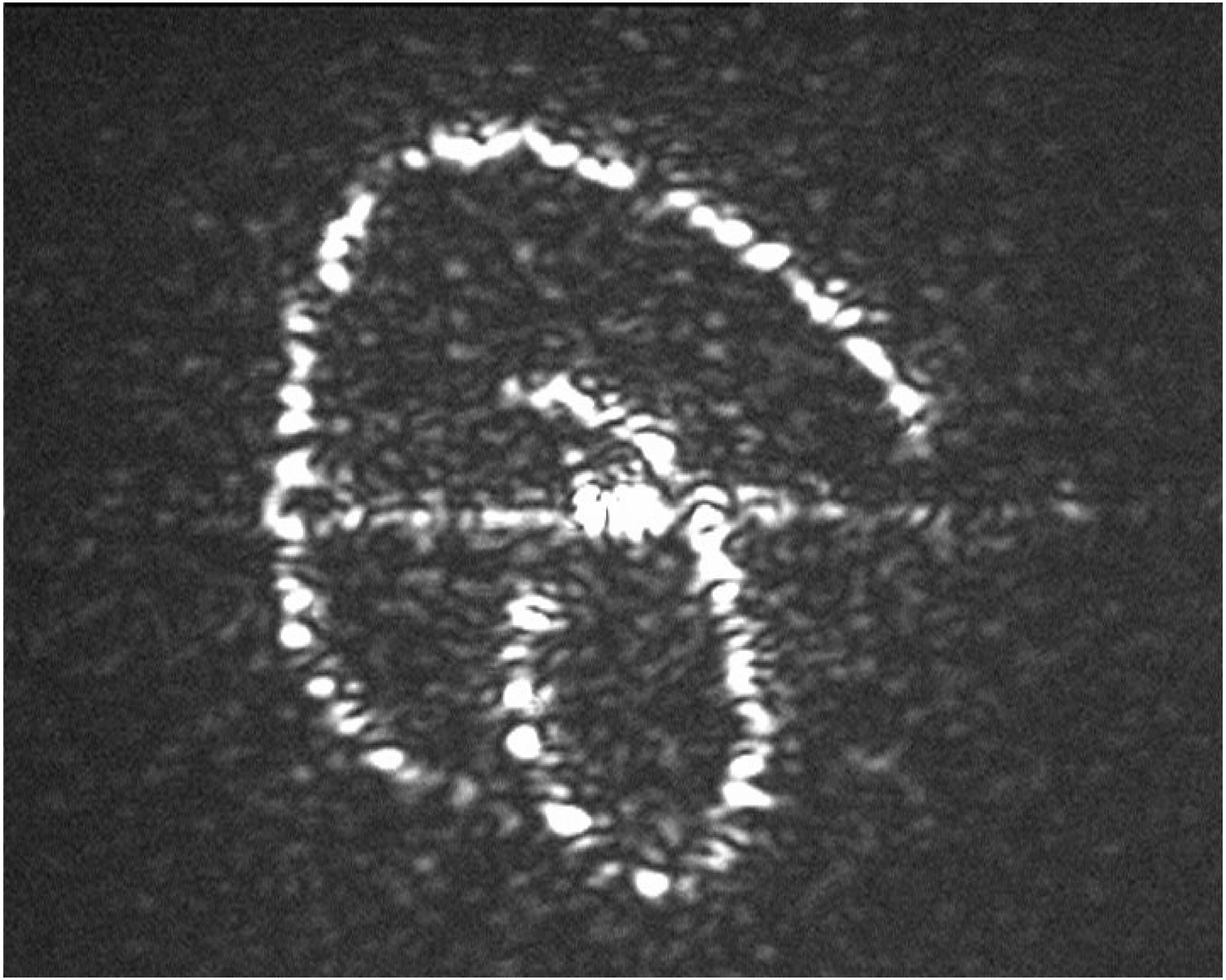
This symmetry restricts the useful area of the replay field to the upper half plane of the sinc envelope.


CRL
CRL

Performance is better when A is no longer restricted to binary values. An example of this would be a four level or quaternary hologram where $A \in [+1, j, -1, -j]$.

This allows us to break symmetry and control noise properties...

A suitable PhD topic…



Hologram scaling and apodisation

In order to display a grating or computer generated hologram (CGH) we need to be able to understand the effects that optical components will have on the replay field.

We have already seen how sampling and finite pixel size effect the replay field by adding outer orders and a sinc envelope.

We can now look at the scaling (dimensions) of the replay field and the effects of having a finite number of pixels (apodisation).

The spatial coordinates (u,v) are related to the original absolute coordinates used earlier in the diffracted aperture (α,β) by the relationship.

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

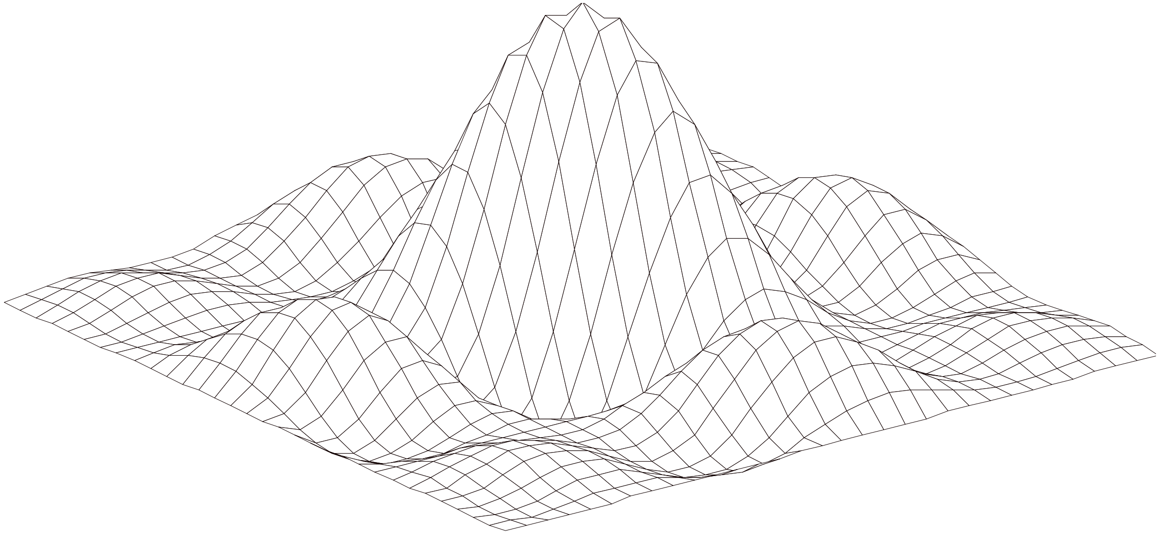
Hence, the scaling of the FT is proportional to f, and as the focal length shortens, the FT shrinks in dimension.

We can use the above relationships to directly calculate the positions of peaks in the envelope of the hologram's replay field.

A hologram comprises of an *N*x*N* array of square apertures (pixels) with a pixel pitch Δ having an amplitude *A* given by either amplitude modulation [0,1] or phase modulation [+1,-1].

The envelope due to the fundamental pixel which covers the far field diffraction pattern (or FT) of the hologram is just a 2-D sinc function (where *a* = Δ).

$$A\Delta^2 \mathrm{sinc}(\pi\Delta u)\,\mathrm{sinc}(\pi\Delta v)$$



It forms the *envelope* function for the replay field of the hologram.

The useful information of the replay field is contained in the central first lobe of the sinc function, so we can calculate the width of the replay field as where the first zero of the sinc function occurs ($\pi\Delta u = \pi$, $\pi\Delta v = \pi$).

We want the coordinates in terms of [α,β] so we get the limits of the central lobe.

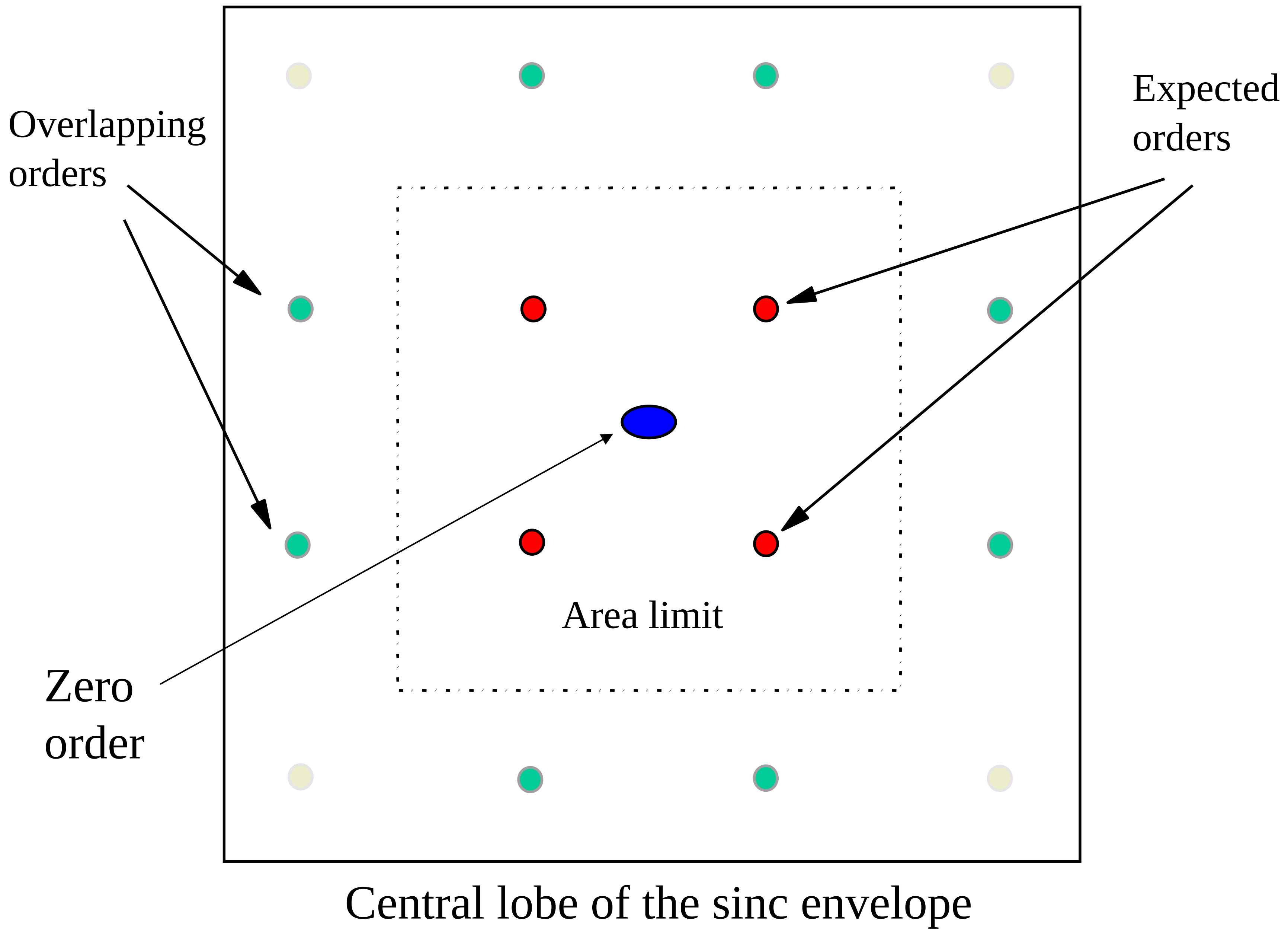
$$\alpha_M = \frac{f\lambda}{\Delta} \qquad \beta_M = \frac{f\lambda}{\Delta}$$

$2\alpha_M$ and $2\beta_M$ tell us the width of the central lobe of the sinc envelope, due to the pixel pitch Δ. From this we can assume that an NxN pixel hologram will generate NxN spatial frequency 'pixels' in the replay field.

This is an approximation, as the FT actually generates a continuous function in the replay field. Hence, the replay field will have an approximate spatial frequency pixel of pitch.

$$\alpha_0 = \frac{2 f \lambda}{N \Delta} \qquad \beta_0 = \frac{2 f \lambda}{N \Delta}$$

Note, 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 [α_M/2, β_M/2] if overlapping hologram replay patterns are to be avoided. (Nyquist sampling theorem)


Overlapping
orders
Expected
orders
Area limit
Zero
order
Central lobe of the sinc envelope


This is a test of
OSPR at 1280x1024.
This is a test of
OSPR at 1280x1024.
This is a test of
OSPR at 1280x1024.

Apodisation

So far, the illumination of the hologram (and the hologram) has been uniform and the hologram and lens extended 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.

This is often referred to as **apodisation**.

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 a intensity distribution which can often be expressed as a Gaussian beam profile or function.

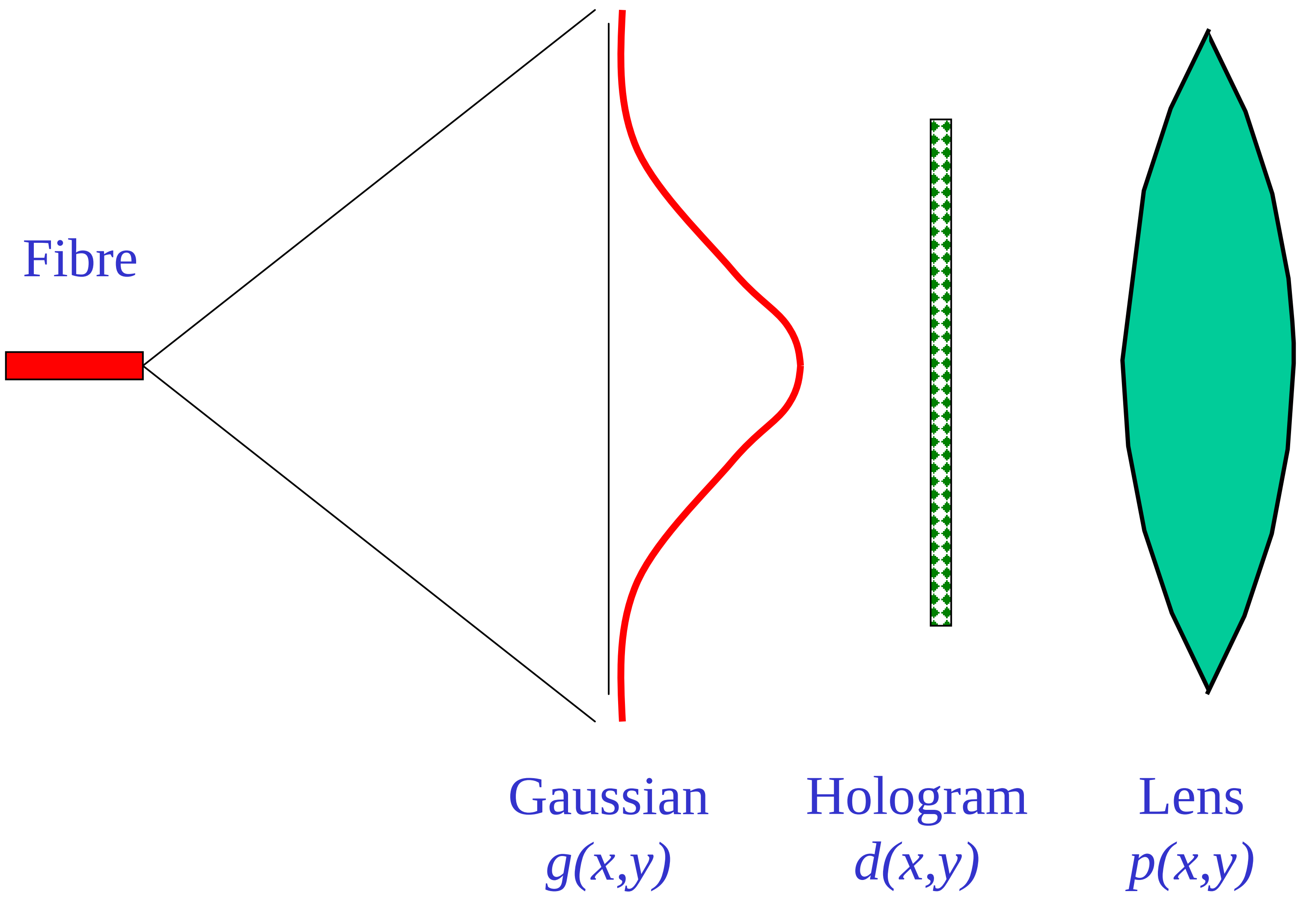
$$g(x,y) = A_G e^{l\left(x^2+y^2\right)}$$

The most useful property of the Gaussian function is that its FT is another Gaussian with a different scale factor (i.e. size).

If a lens is illuminated with a collimated Gaussian source, then the focal plane will also contain a Gaussian profile which is scaled by the focal length.

The actual structure of the replay field of a hologram is even more complex, as the Gaussian beam profile of the source is also apertured by the hologram and lens.

Apodisation



The entire illumination system can be modelled 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.

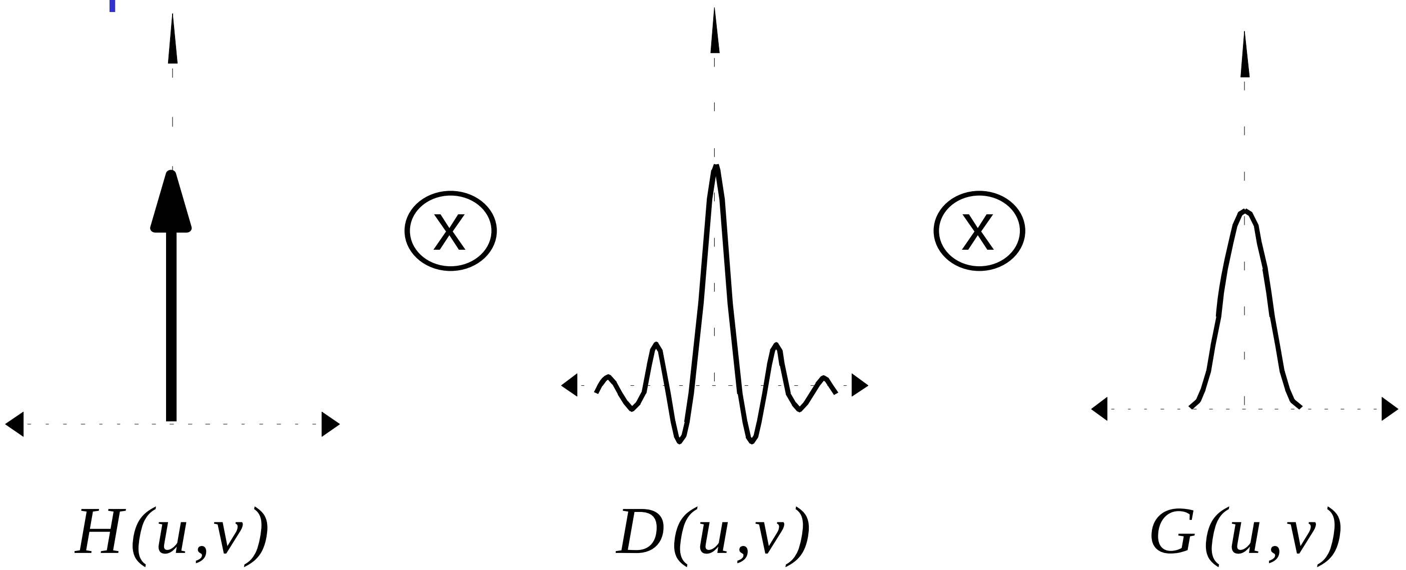
$$F(u,v) = G(u,v) \otimes D(u,v) \otimes P(u,v)$$

The ideal hologram replay field *H(u,v)* is designed as an array of delta functions in desired positions.

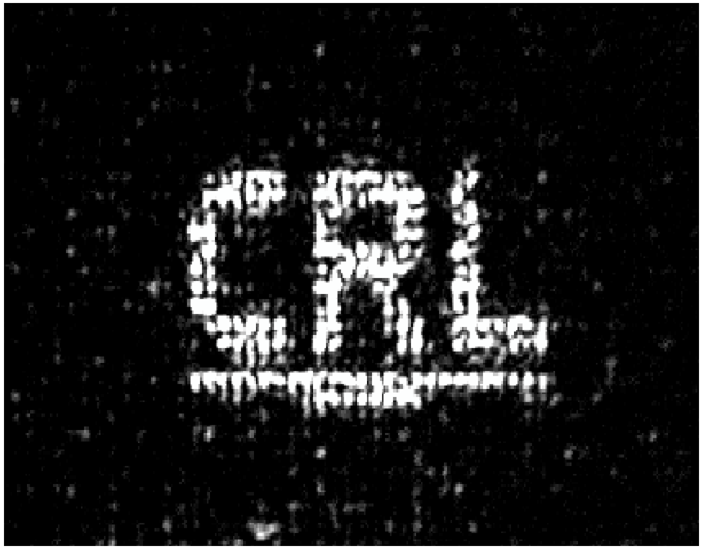
The lens aperture *p(x,y)* is a large circular hole, so the FT *P(u,v)* is a first order Bessel function (a circular sinc function).

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 convolve a Gaussian profile.



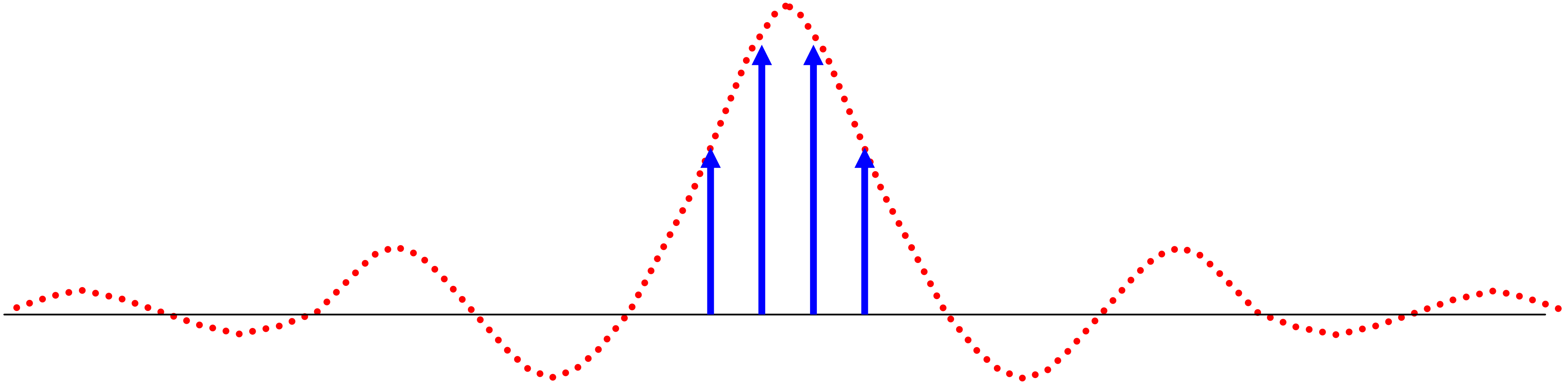
This means that the replay field will not look exactly as expected, spots which are placed next to one another will interfere due to the tails of the Gaussian, sinc and Bessel functions and the individual desired sharp ‘spots’ become ringed blobs with finite width.



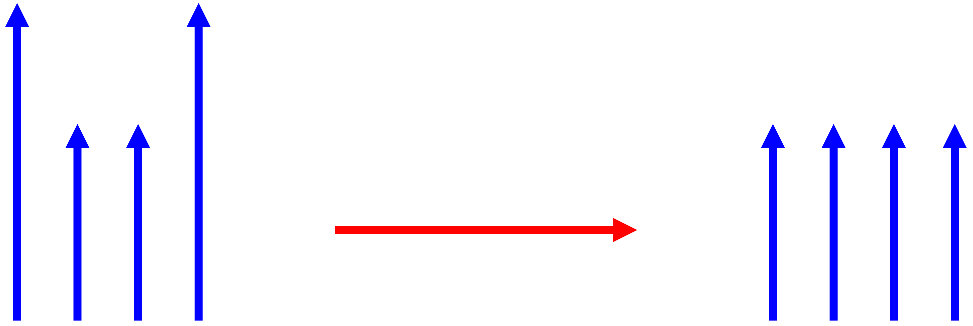
Sinc envelope compensation

Don’t forget that all this occurs under the envelope of the sinc function over the entire plane due to the square shape of the hologram pixels.

There is a pronounced effect due to the overall sinc envelope, with the outer spots in the corners being lower in intensity than the centre spots.



The effects of the sinc envelope can be countered by compensating the height of the outer spots so that the overall effect cancels and leaves the spots all equal heights



Wavelength manipulation

The manipulation of spatial frequencies in the replay field implies that the wavelength is held constant, however as can be seen in the normalized spatial frequency variables [u,v], the spatial frequency is also a function of wavelength ($k = 2\pi/\lambda$).

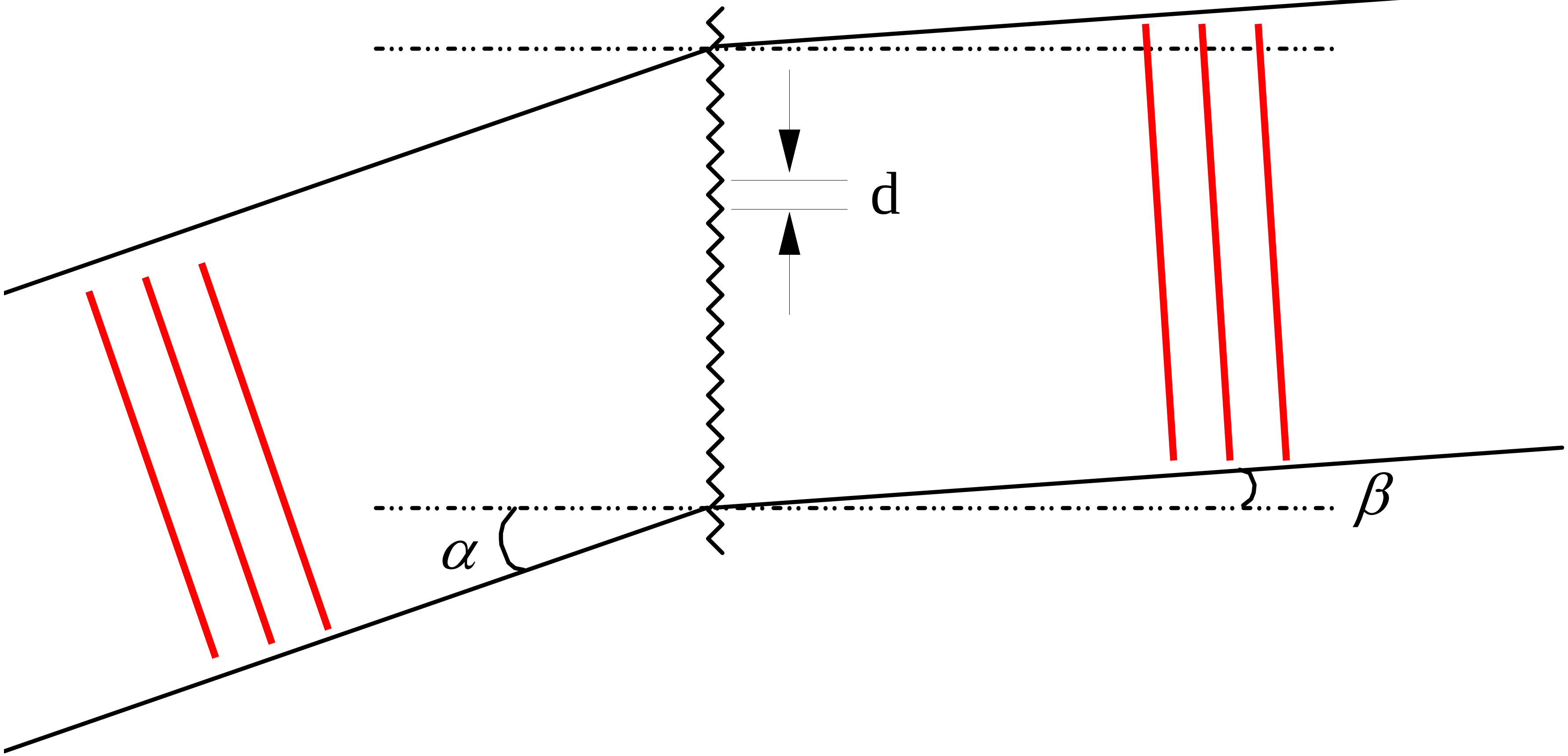
$$u = \frac{k\alpha}{2\pi f} \qquad v = \frac{k\beta}{2\pi f}$$

In Fourier holograms, spatial and wavelength properties are interchangeable.

If a hologram or grating is created to generate fixed positions or orders in the replay field, then if the wavelength is varied, then the positions of the orders from the hologram will vary.

This is known as a wavelength sensitive or dispersive hologram or grating.

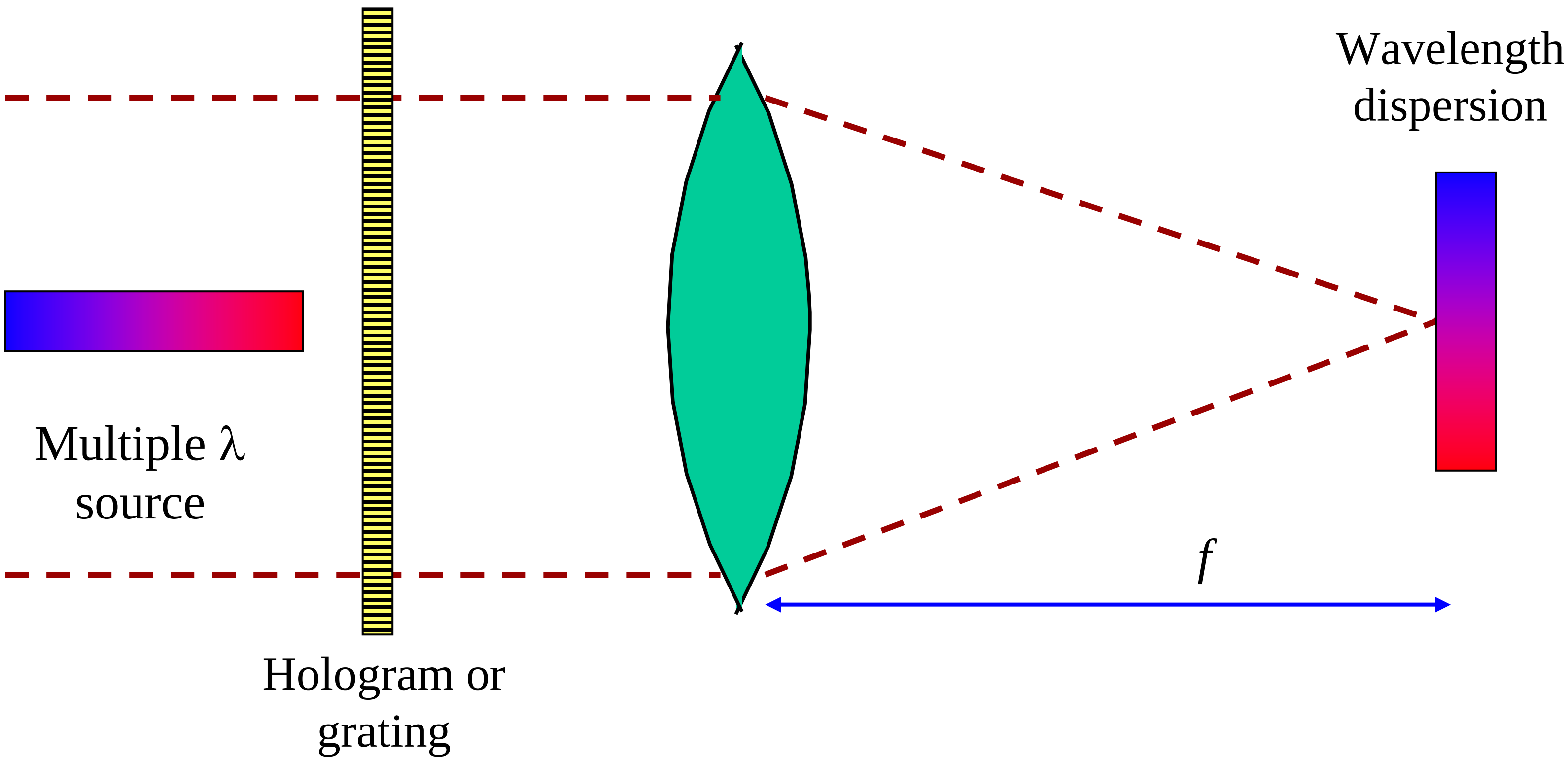
This concept can also be expressed by classical grating theory which relates the pitch of the grating *d* to the angle at which the light that passes though it is diffracted.


d
α
β

This shows how the diffracted angle of the light is dependent on the grating pitch and the wavelength.

$$\sin\alpha - \sin\beta = \frac{m\lambda}{d}$$

By placing a positive focal length lens after the grating, we can view the far field we see that the diffracted angle is converted into the position of the diffracted order (as we would expect from a grating).


Wavelength
dispersion
Multiple λ
source
f
Hologram or
grating

Hence, wavelength and position will vary with the grating pitch in the far field.

If we monitor a fixed point in the far field, then the wavelength will scan across that point with the changing pitch, making a wavelength filter.

If the grating pitch d, were increased as an integer, we would only have a small number of fixed wavelengths that could be tuned.

However, by using one dimensional holograms, we can select any point in the output plane along the single axis which means we can select a range of angles and hence a range of wavelengths.

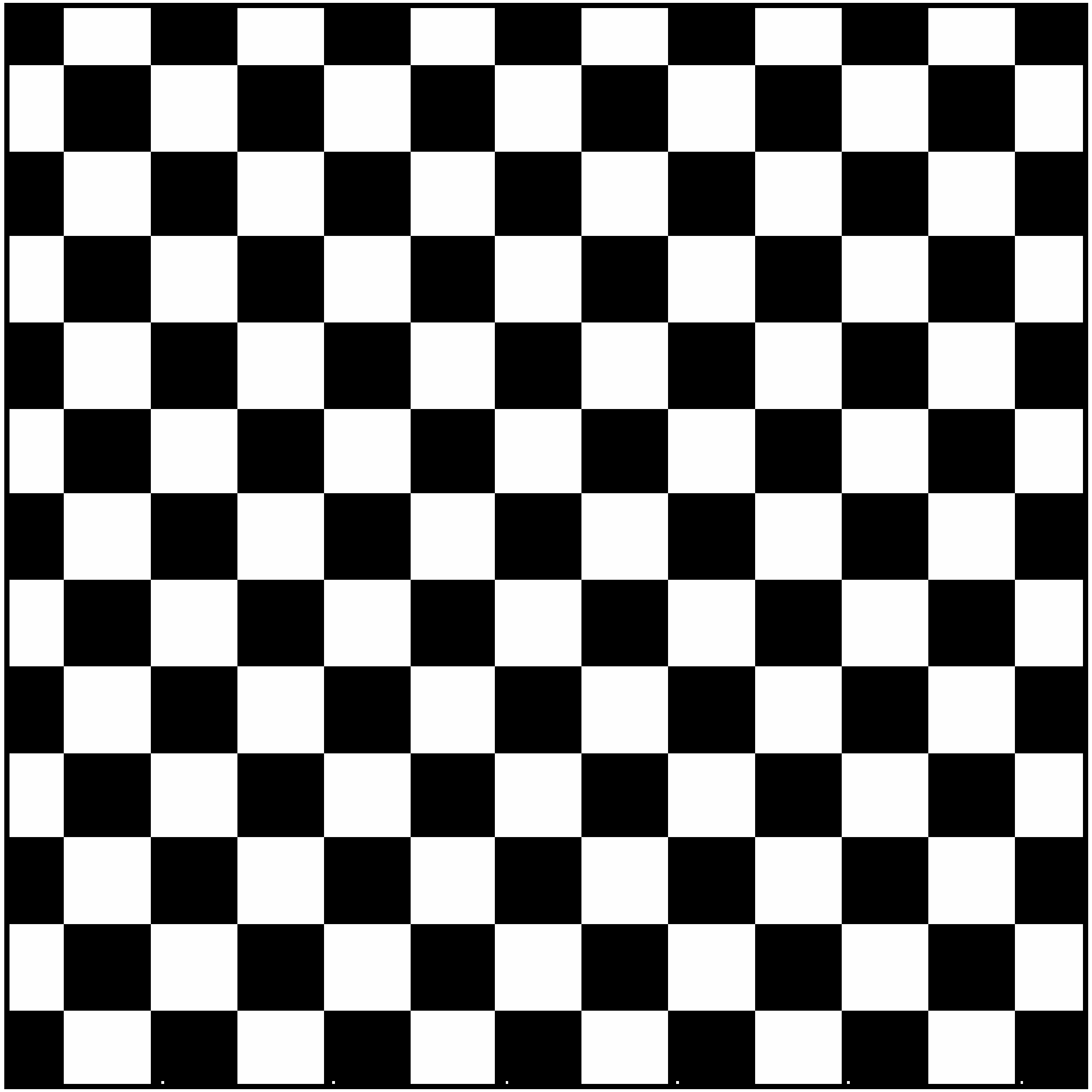
The golden rule of holograms

The hologram pixel shape and pitch defines the overall envelope of the replay field

The hologram aperture or overall size defines the spatial resolution or ‘spatial frequency pixels’ in the replay field.

The replay field scales with wavelength

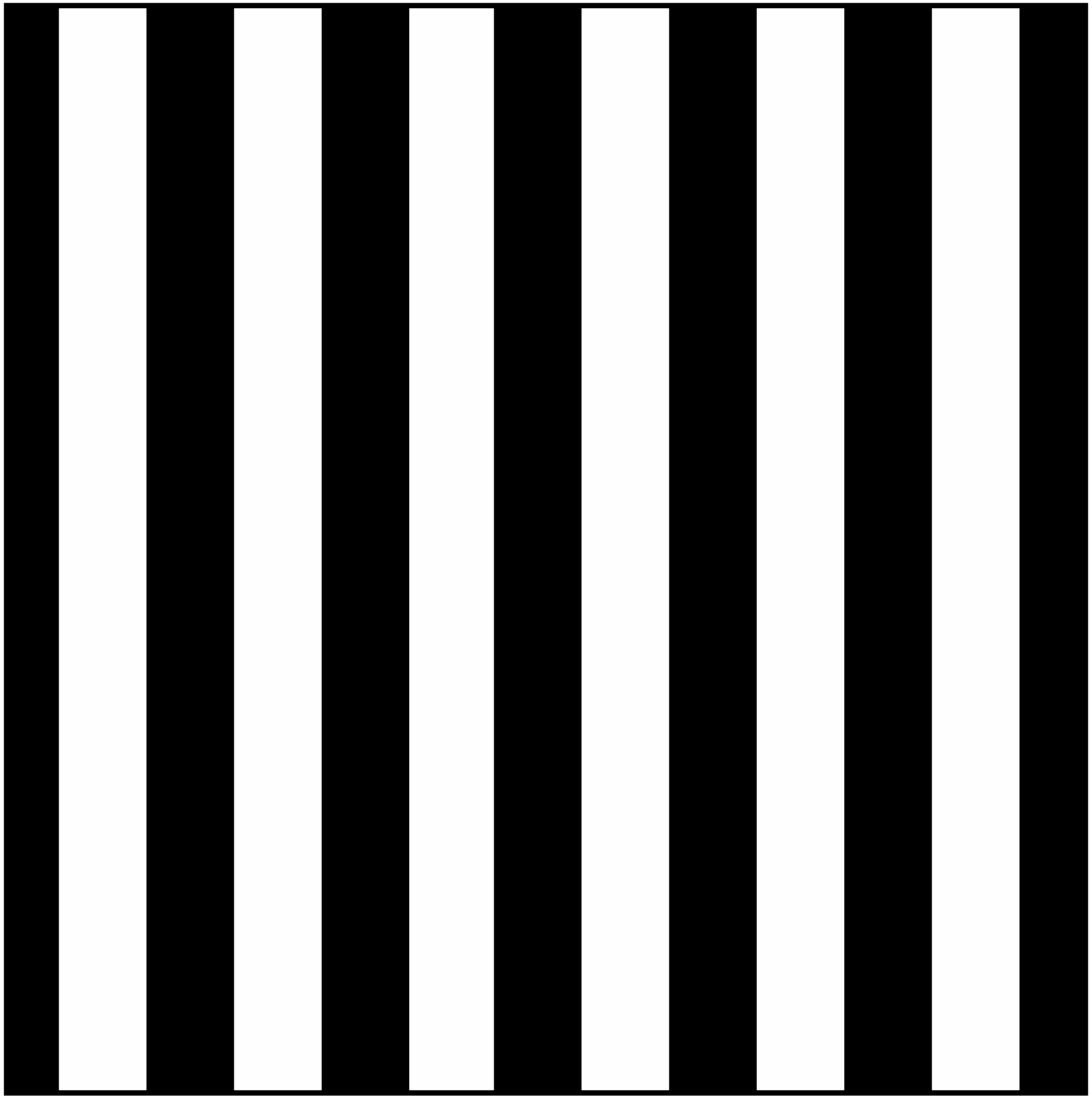
How do we calculate more complicated patterns in the replay field?



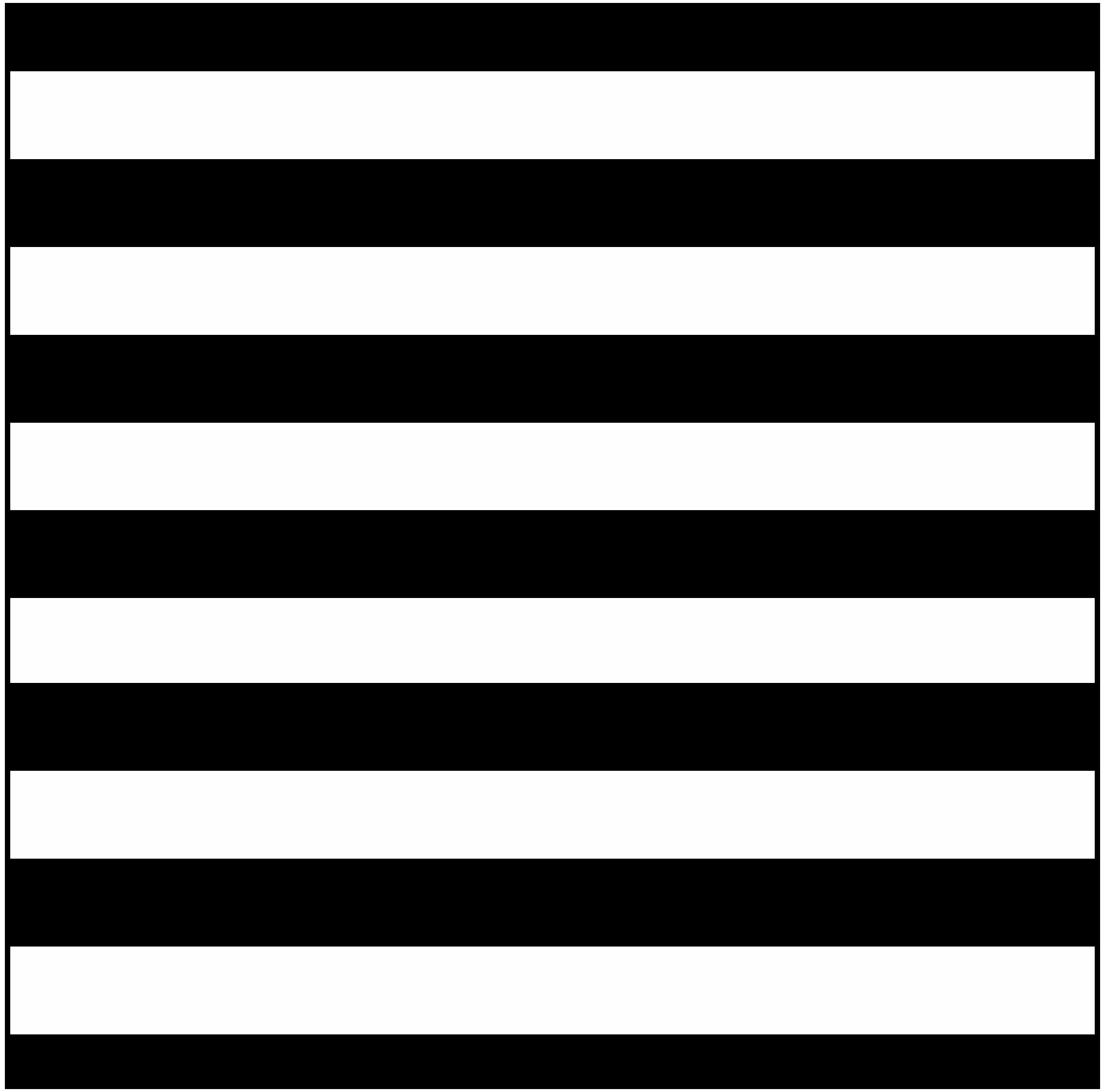
Now we can look at a chequerboard pattern of pixels on an equally spaced grid (*A* is restricted to binary values, such as 0 or 1).

The chequerboard can be generated by the XOR of a 2D grating with itself rotated by 180º (as discussed earlier).

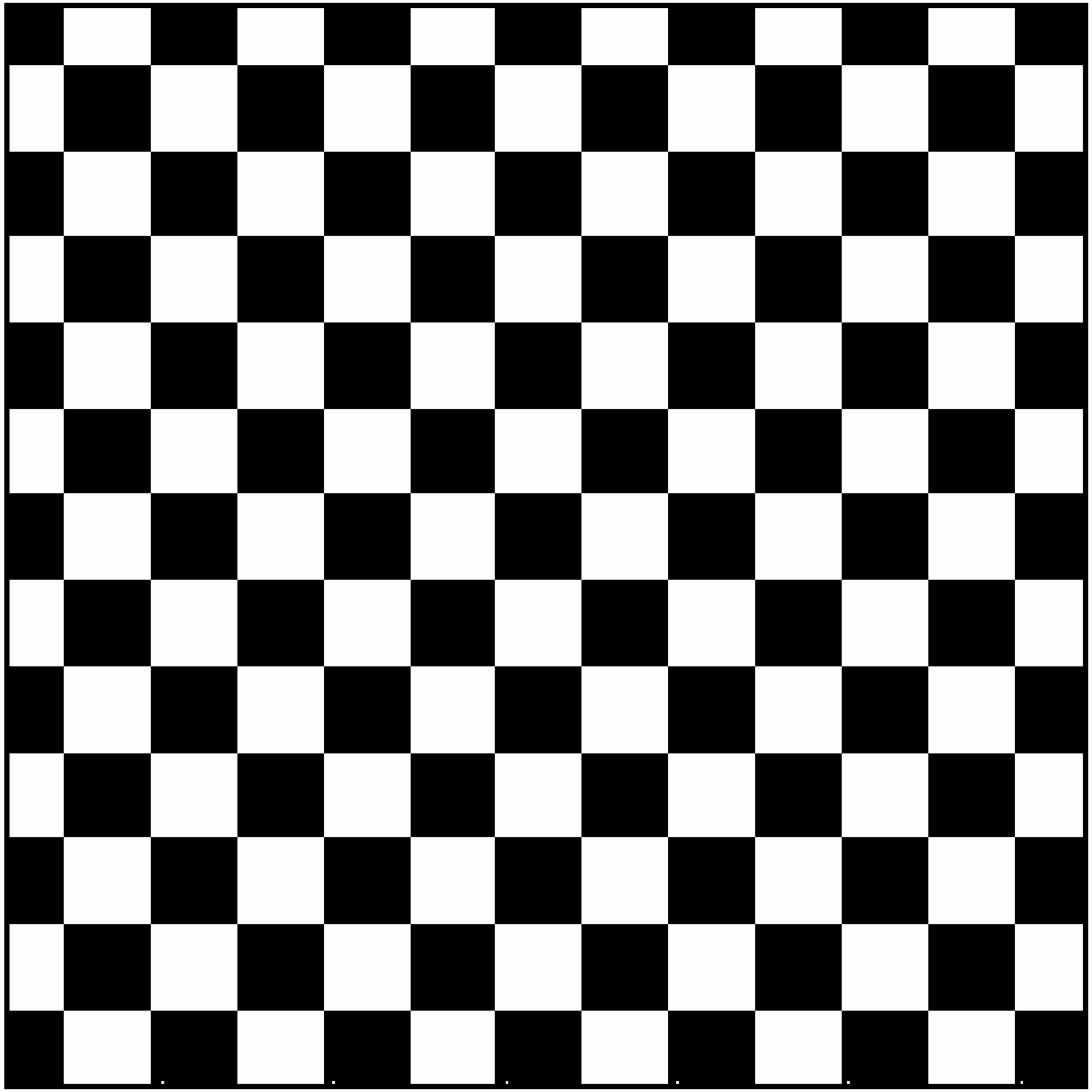
Hence, the replay field will be made from the convolution of the FT of the two gratings

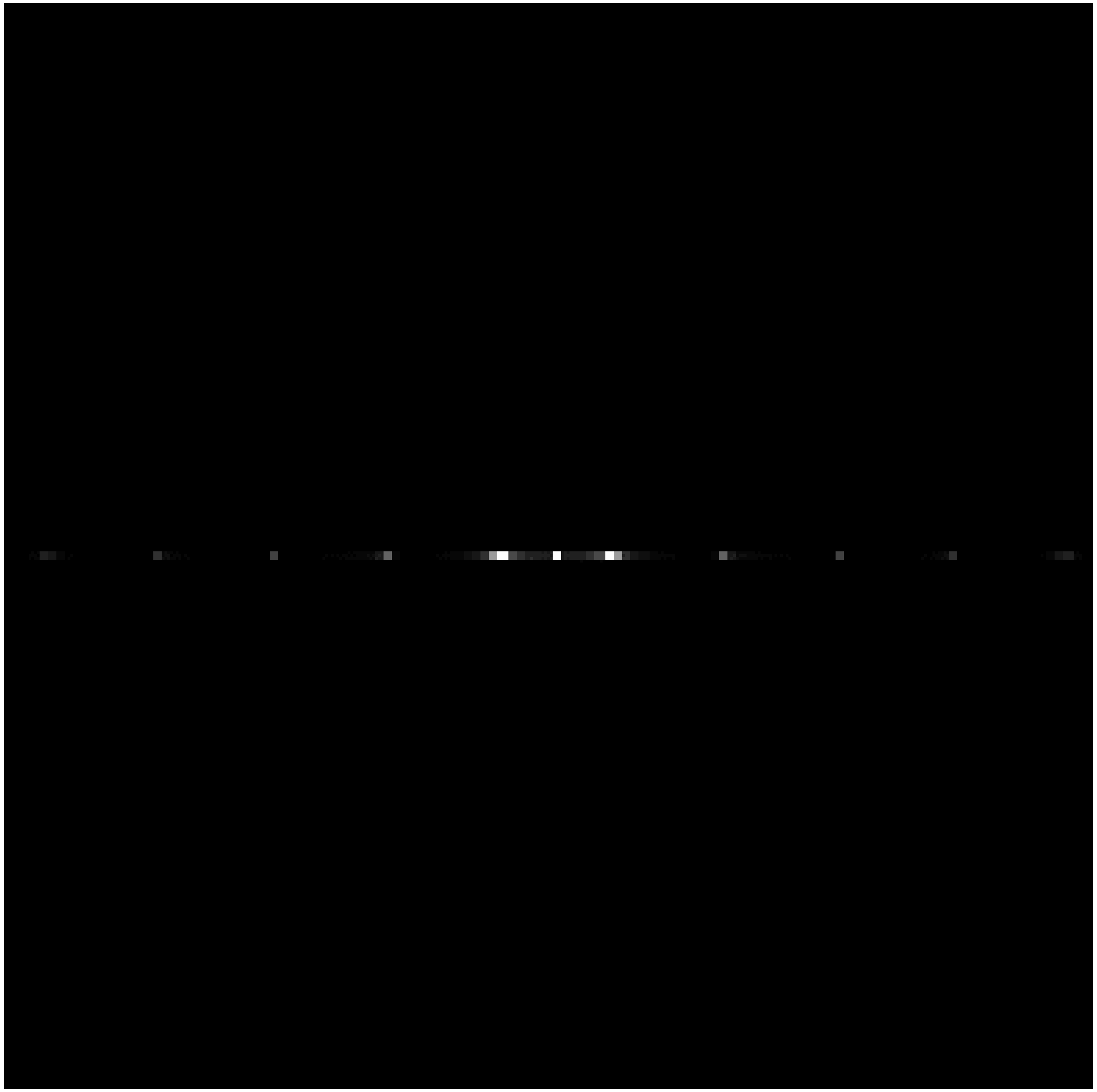


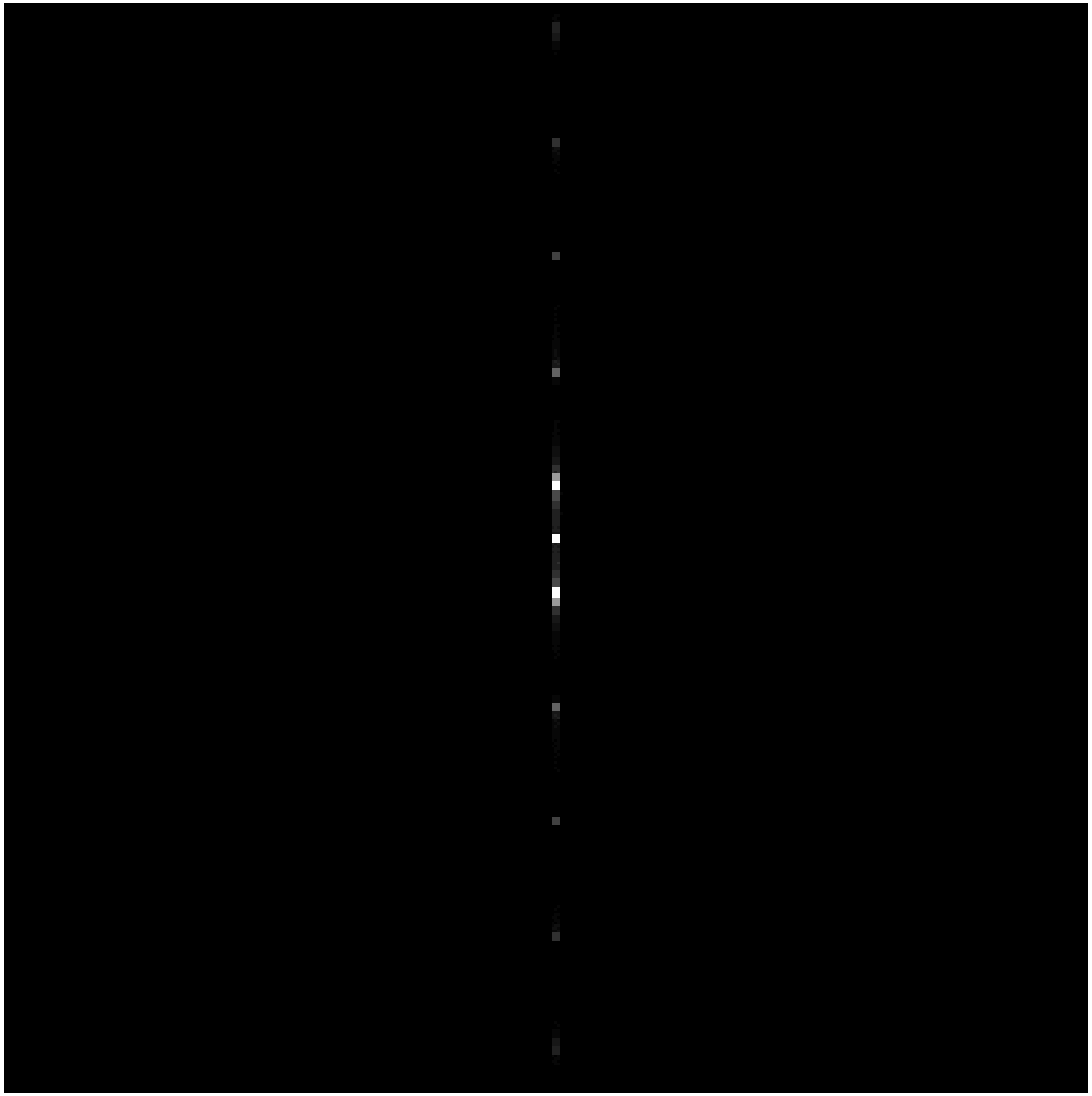
X

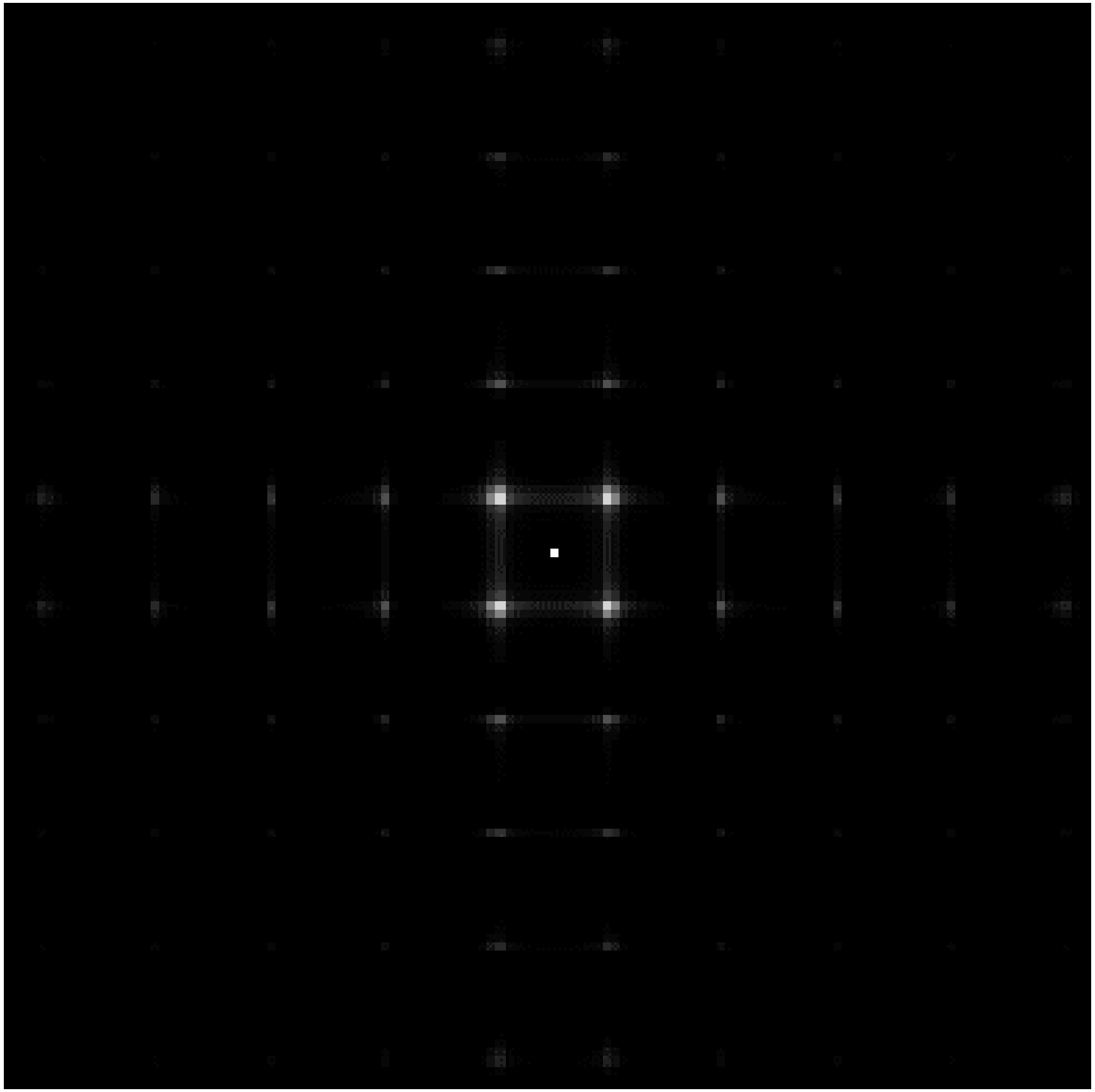


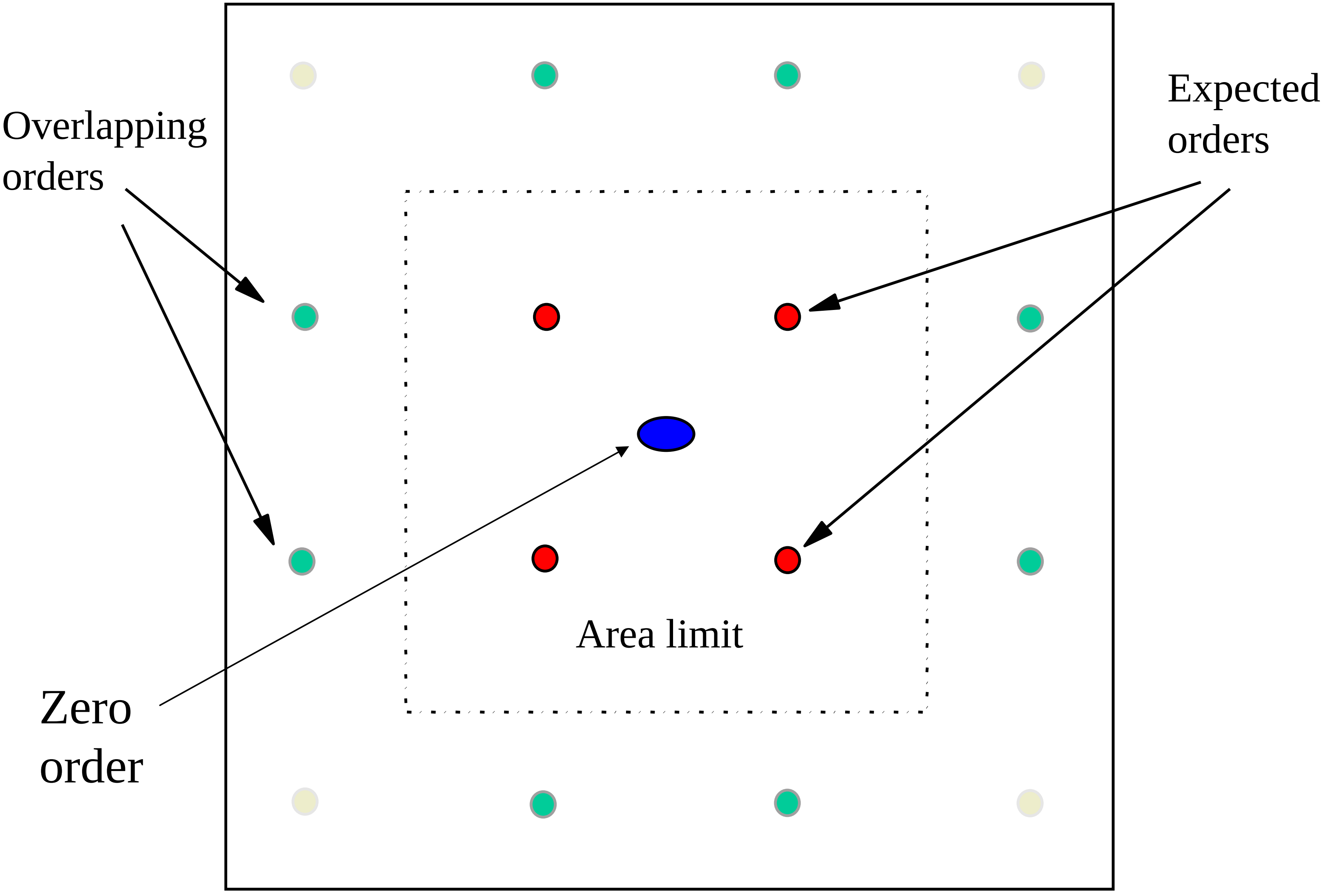
=







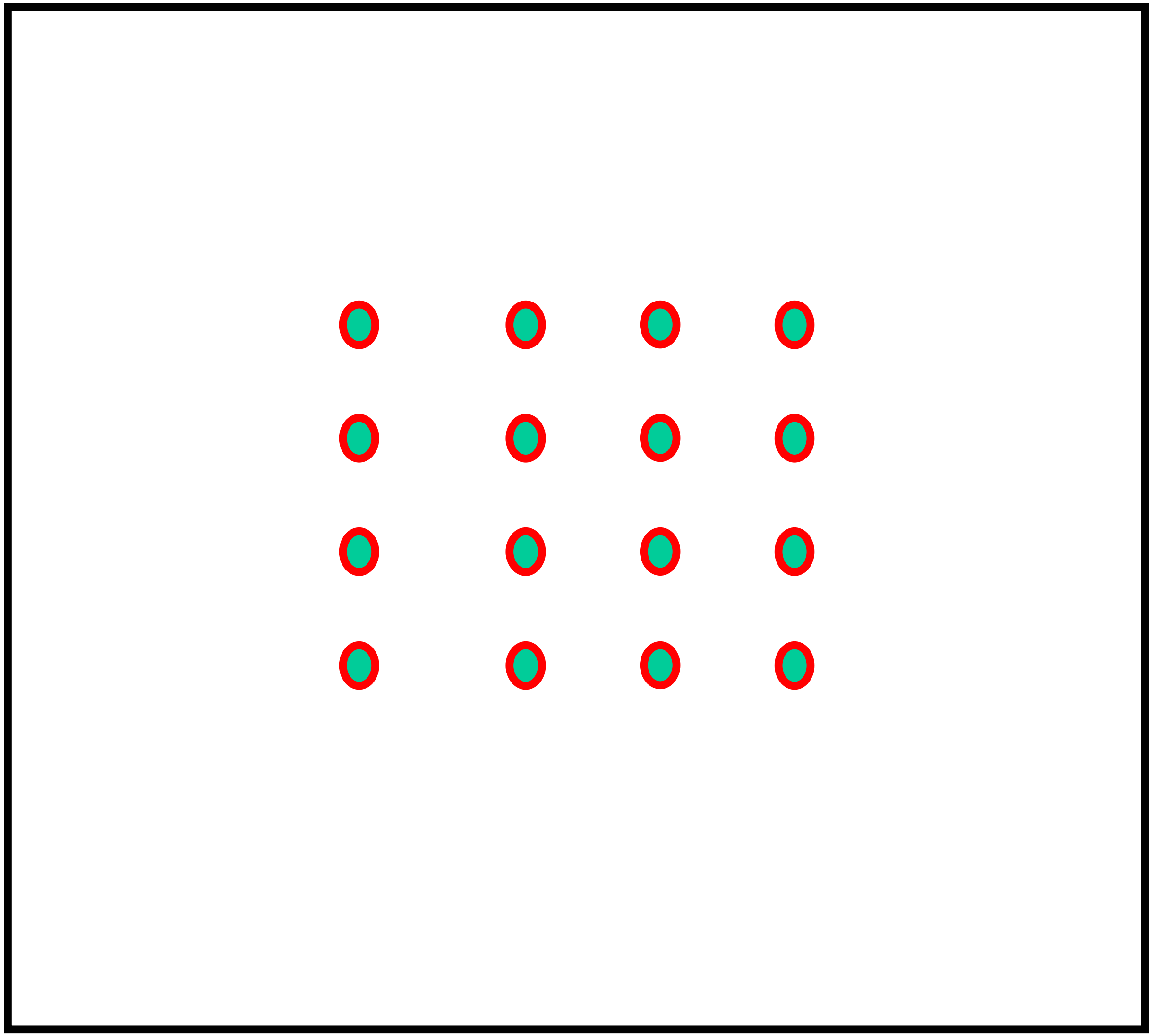


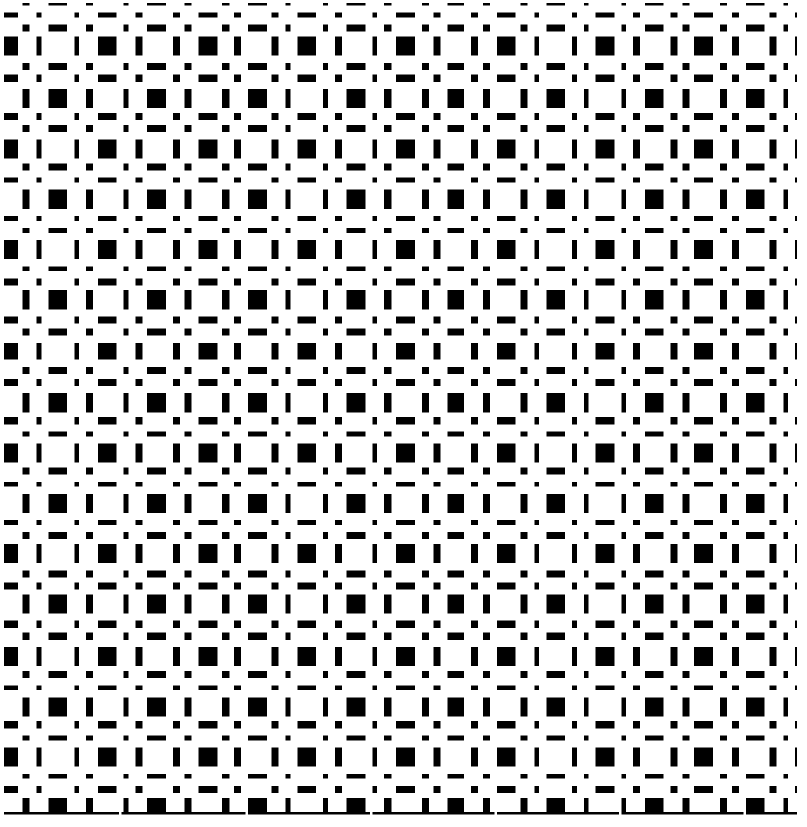

Overlapping
orders
Expected
orders
Area limit
Zero
order

There is no simple way of generating holograms except for simple cases such as gratings and chequer boards.

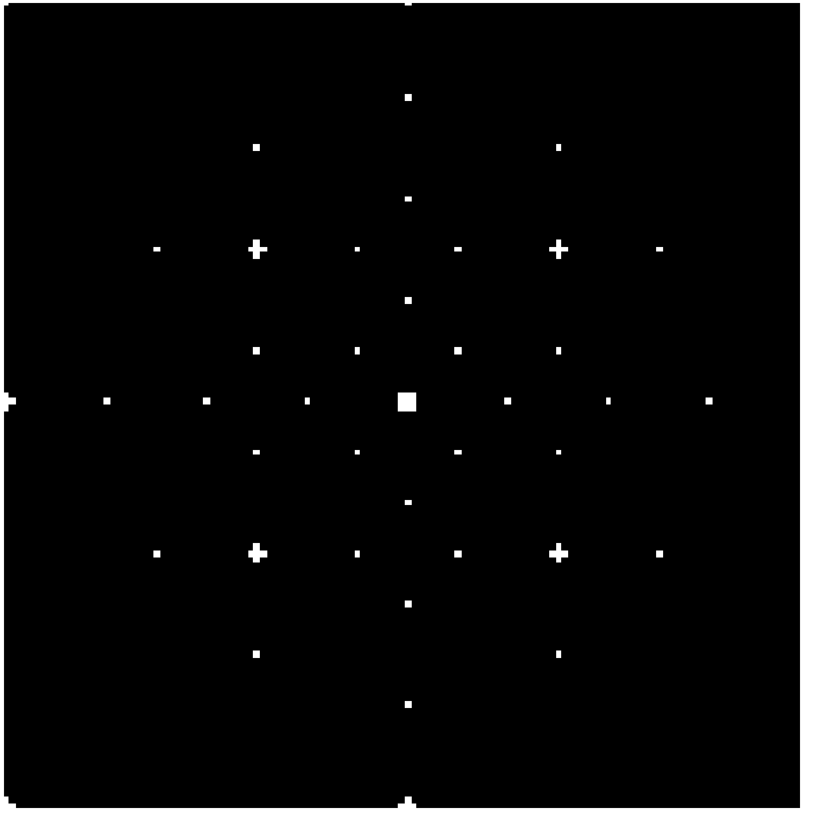
In order to create a hologram which generates an arbitrary replay field we need a more sophisticated algorithm. We must use optimisation techniques such as **Gerchberg-Saxon**, **direct binary search, simulated annealing, error diffusion** or **the genetic algorithm**.

To illustrate the basic concept of these algorithms we shall look at a simple example based on direct binary search, which is a simplification of simulated annealing.





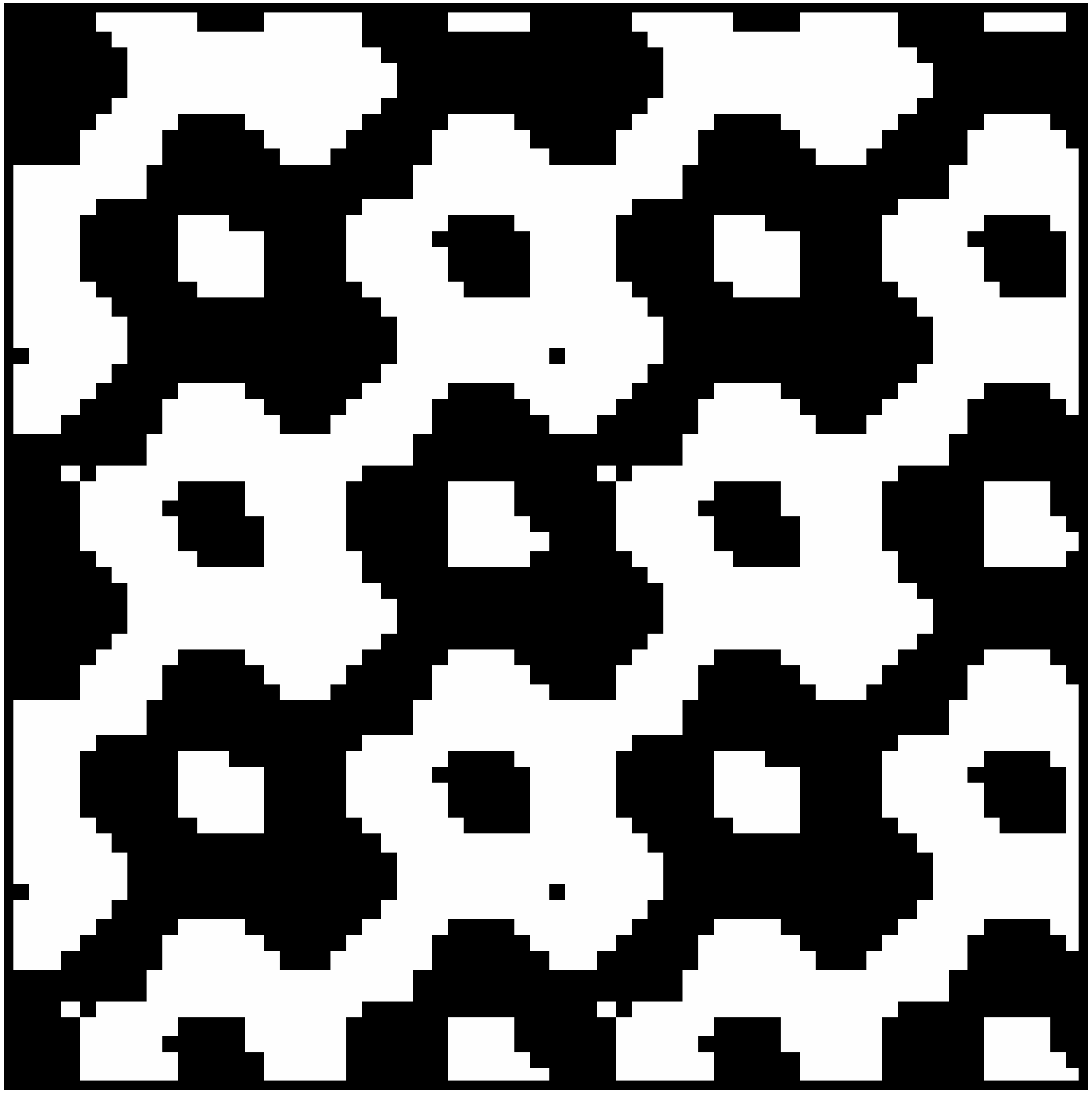
Hologram

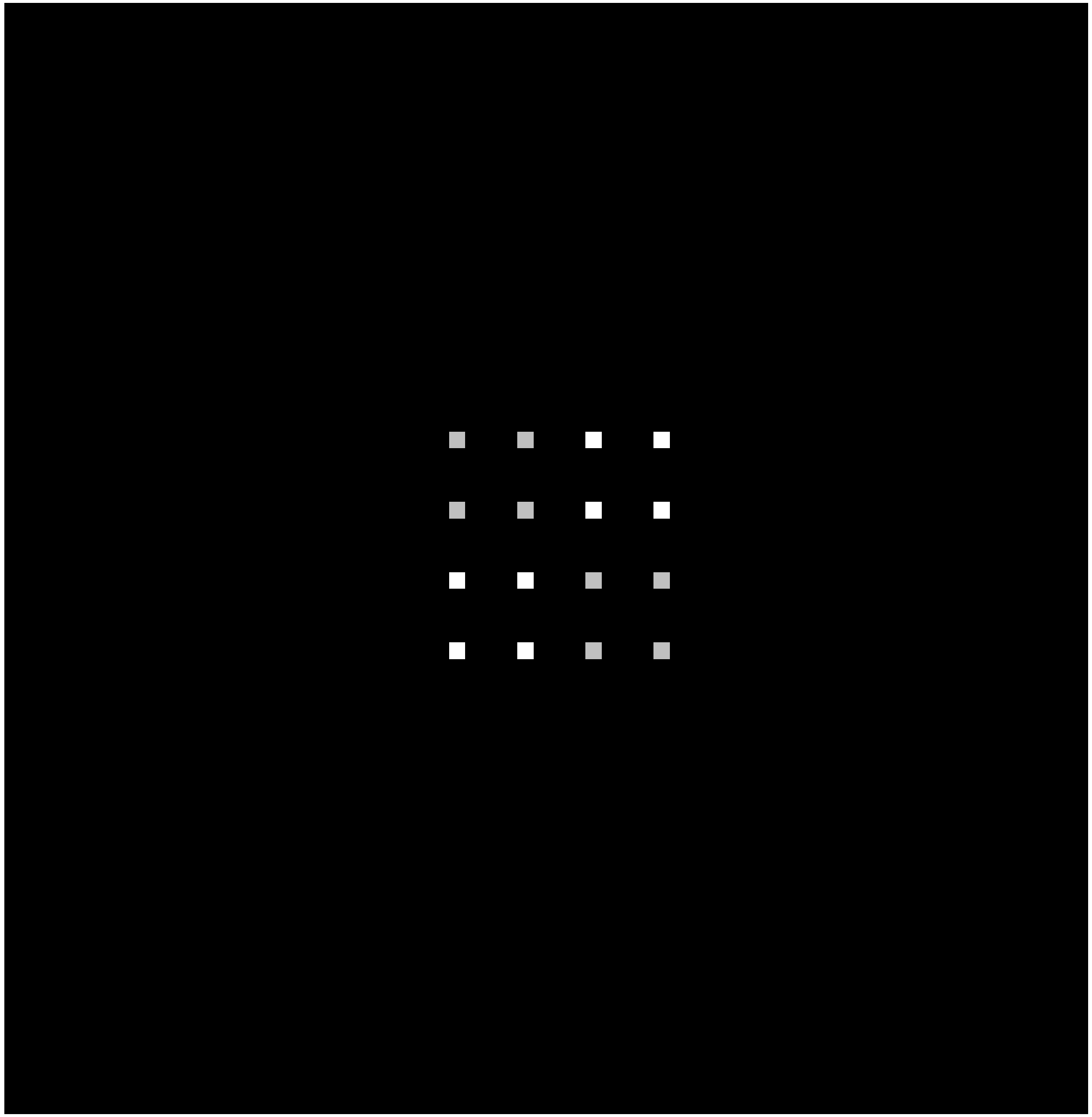


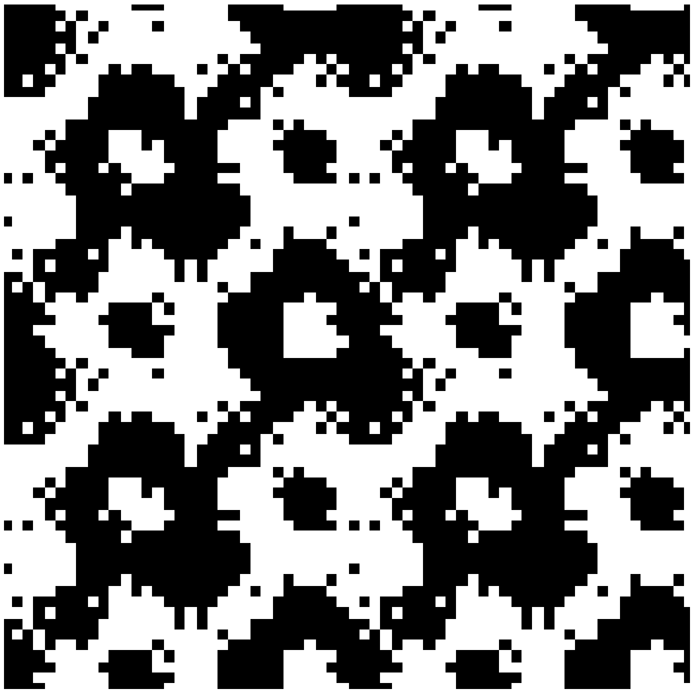
Replay Field

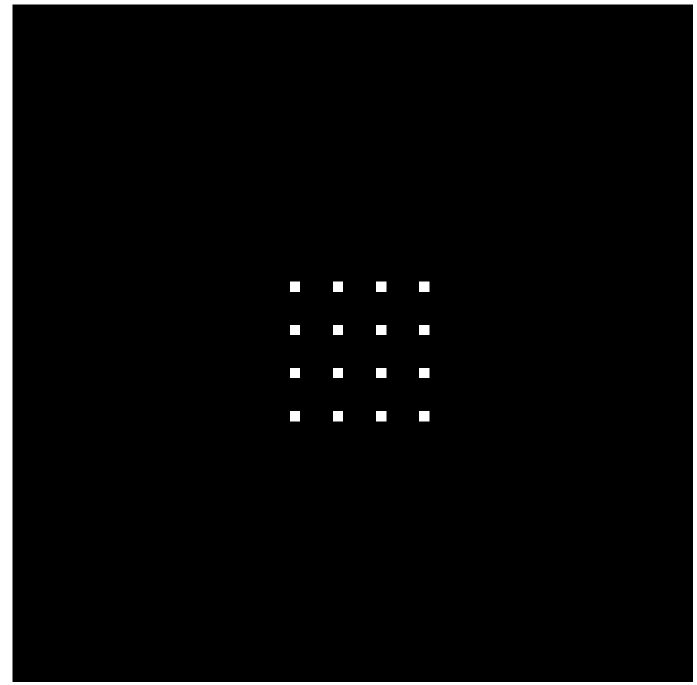
Direct binary search

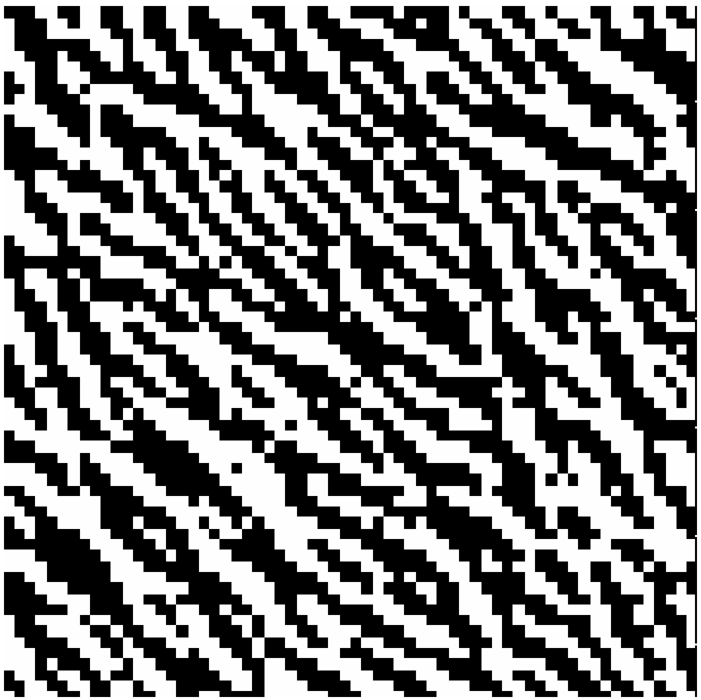
1) Define an ideal target replay field, T (desired pattern)
2) Start with a random array of binary phase pixels.
3) Calculate its replay field (FT), H_0.
4) Take the difference between T and H_0 and then sum up to make the first cost, C_0.
5) Flip a pixel state in a random position.
6) Calculate the new replay field, H_1.
7) Take the difference between T and H_1 then sum up to make the second cost, C_1.
8) If $C_0 < C_1$ then reject the pixel flip and flip it back.
9) If $C_0 > C_1$ then accept the pixel flip and update the cost C_0 with the new cost C_1.
10) Repeat steps 4 to 9 until $|C_0 - C_1|$ reaches a minimum value.

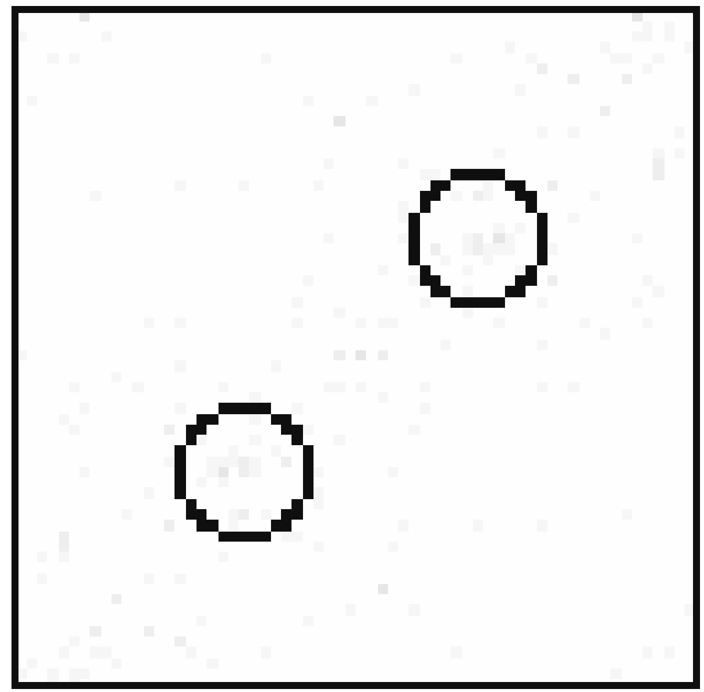


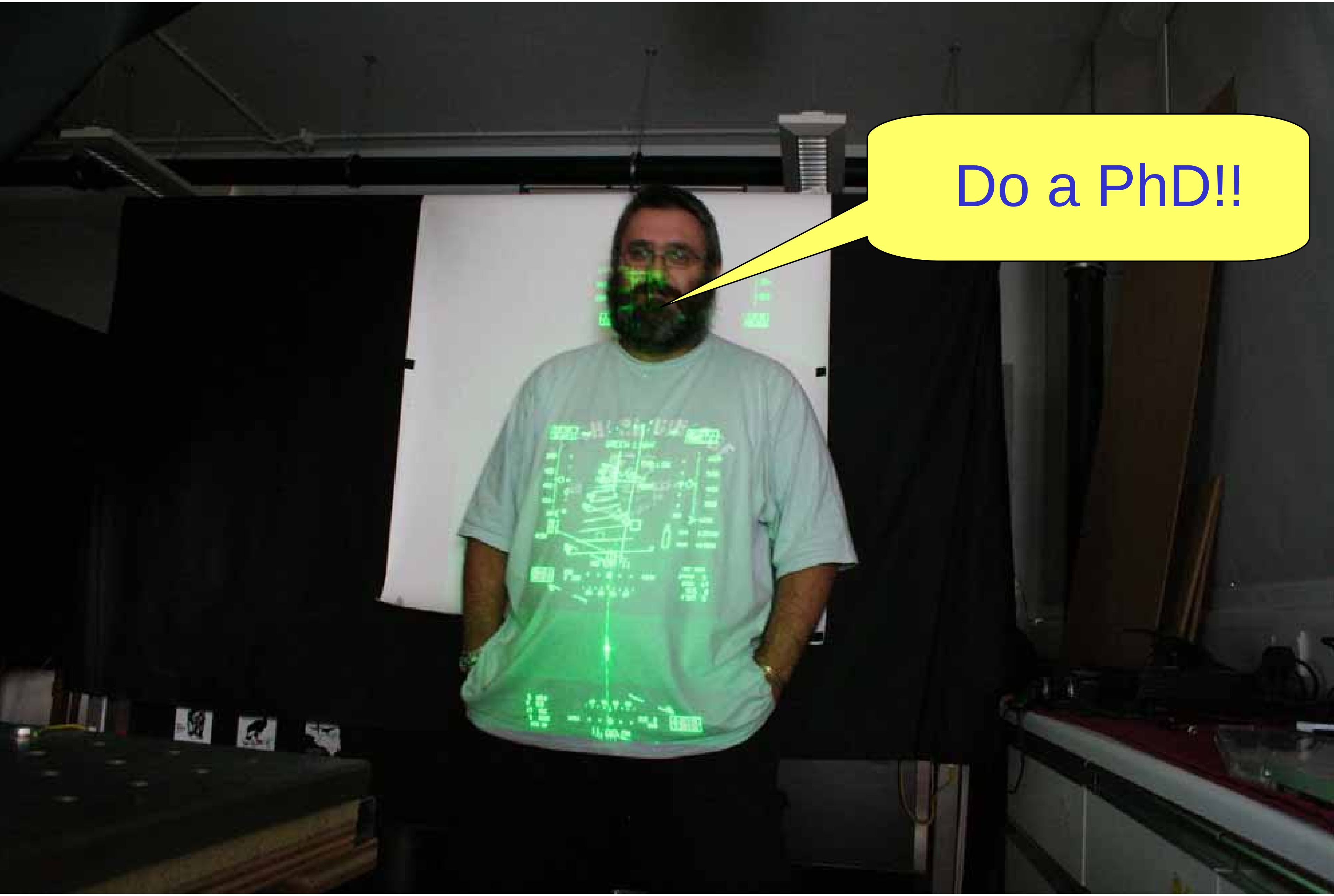











Do a PhD!!