Convolution

23RD NOVEMBER 2020

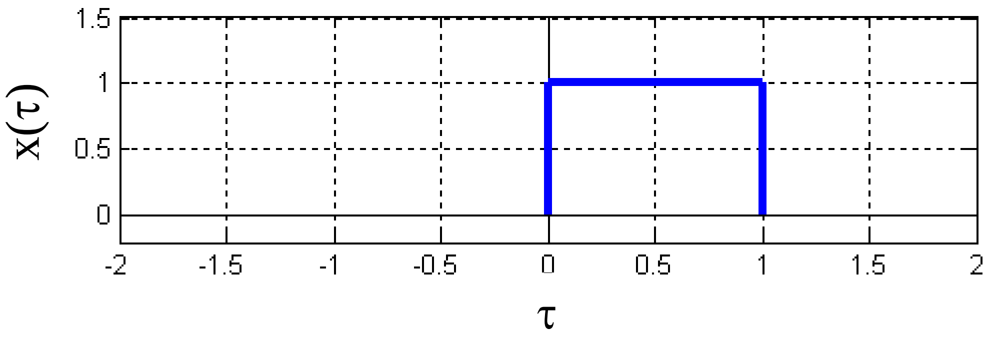
What is the convolution?

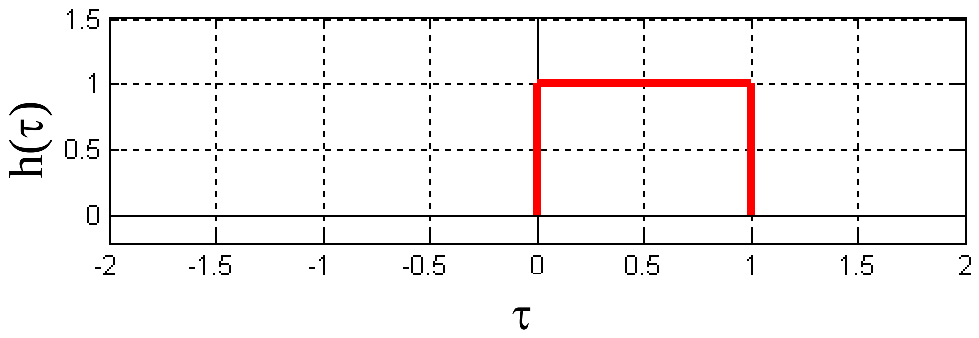
- **Convolution** is a mathematical operation on two functions (*f* and *h*) that produces a third function (*g*) expressing how the shape of one is modified by the other.
- The term *convolution* refers to both the **resulting function** and to the **process of computing it**.
- It is defined as the integral of the product of the two functions after one is reversed and shifted. And the integral is evaluated for all values of shift, producing the convolution function.

$$g(X) = f(x) \otimes h(x) = \int_{-\infty}^{\infty} f(x)h(X - x)dx$$

Example 1 :
1-D Convolution of two square pulses

- Consider two square pulses in 1-D time domain



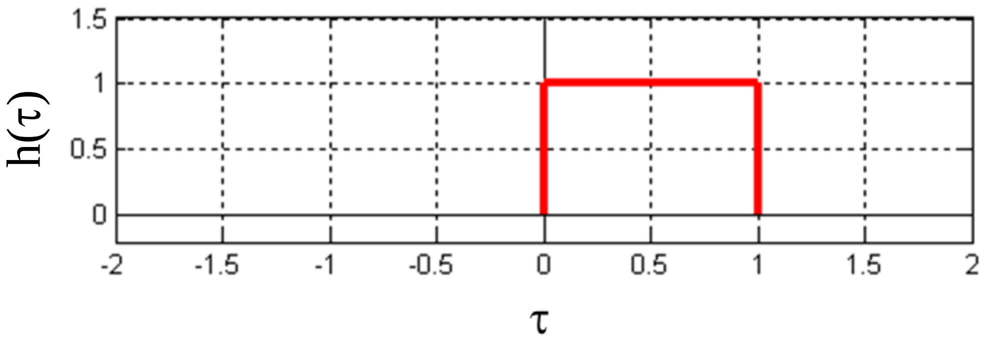


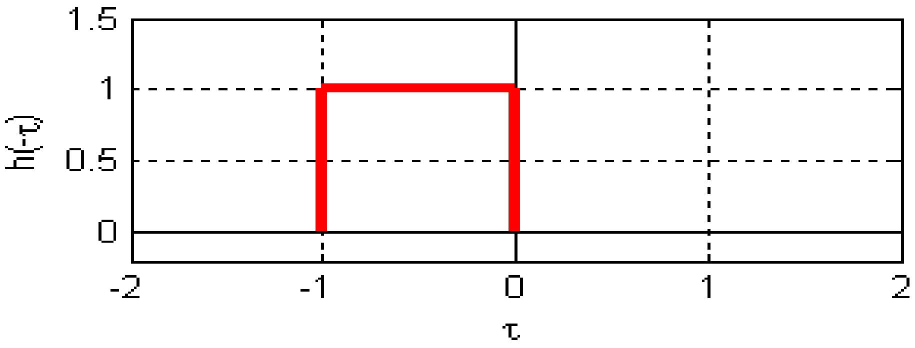
- Recall the definition of the convolution (variables are changed appropriately)

$$y(t) = x(\tau) \otimes h(\tau) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$$

https://cnx.org/contents/-eLl_Dug@19.20:h0BzlV-O@7/Convolution-Complete-example

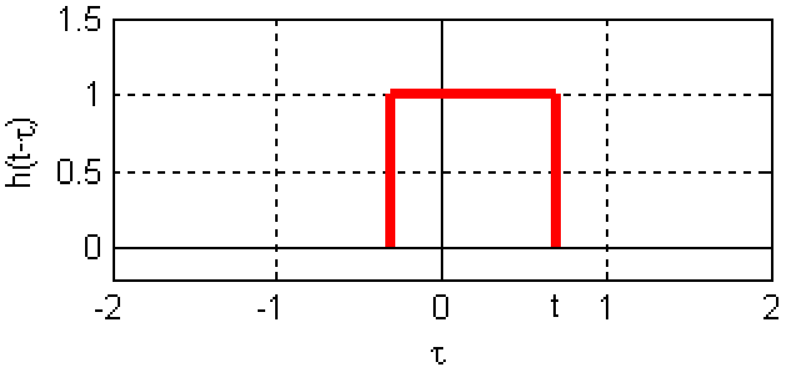
- **For the 1st step, we take one of the functions ;i.e., *h*(τ), and reflect it around the y-axis by using the method described in the next slide.**
- The function then becomes *h*(-τ). In fact, the full form of the function is *h*(*t* - τ) and at this initial location *t* = 0.
  Note that before reflection : $0 < \tau < 1$; after reflection : $-1 < \tau < 0$.





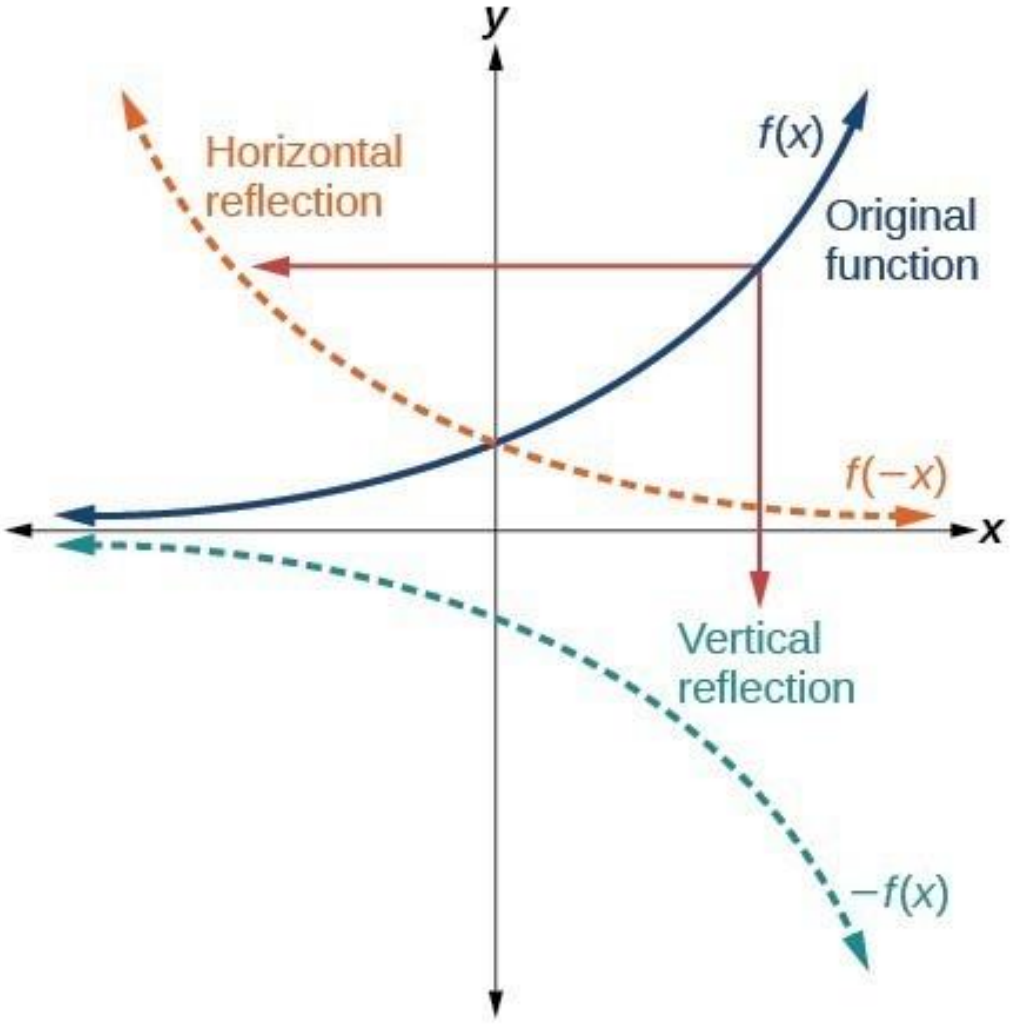
Reflected square wave

- To shift the function *h*(*t* - τ) to the right, *t* has to be appropriately chosen.
-



Reflected and shifted square wave

HOW TO: GIVEN A FUNCTION, REFLECT THE GRAPH BOTH VERTICALLY AND HORIZONTALLY.



1. Multiply all outputs by –1 for a vertical reflection. The new graph is a reflection of the original graph about the *x*-axis.

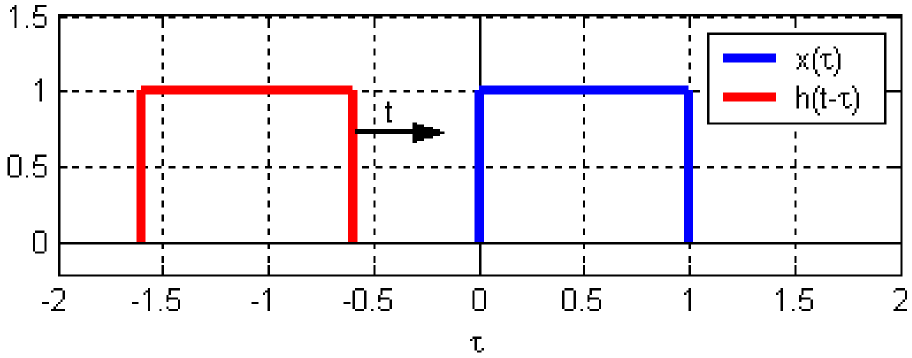
2. **Multiply all inputs by –1 for a horizontal reflection. The new graph is a reflection of the original graph about the *y*-axis.**

https://courses.lumenlearning.com/epcc-atdcoursereview-collegealgebra-1-2/chapter/graph-functions-using-reflections-about-the-x-axis-and-the-y-axis/

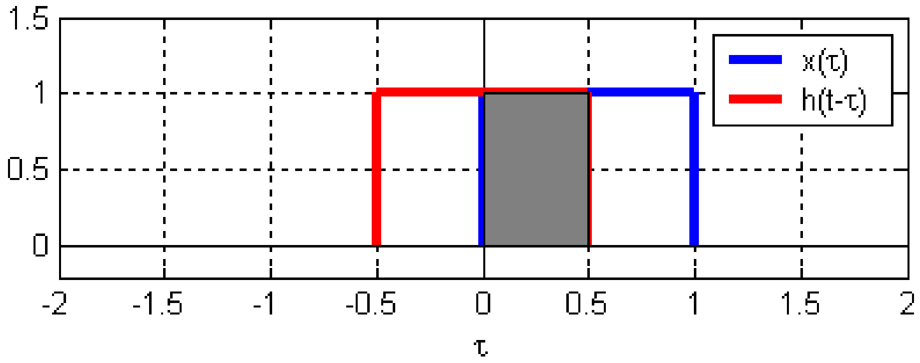
- Since convolution is **commutative** ;i.e. $x \otimes h = h \otimes x$, it will never be matter which function is reflected and shifted.
- However, as the functions become more complicated reflecting and shifting the "right one" will often make the problem much easier.

- **Now, the second step : shifting the reflected function onto the other function.**
- By shifting with an appropriate value of t, an amount of overlapping between $h(t - \tau)$ and $x(\tau)$ is found. This in fact follows the convolution integral : $y(t) = \int_{-\infty}^{\infty} x(\tau)h(\mathrm{t} - \tau)d\tau$.

- For example: if t is chosen to be less than 0 such as -0.6, $h(t - \tau)$ becomes non zero over a range of $-1.6 < t\text{-}\tau < -0.6$.
- Over this range, no overlapping between the two functions is found. This clearly give $y(t) = 0$.



- **The 3rd step : determine the convolution y(t)** by choosing an appropriate range of t used to shift $h(t - \tau)$ is given as $0 \le t \le 1$.
- For example if t is chosen to be 0.5, the shifting function becomes $h(0.5 - \tau)$.
- The function $h(0.5 - \tau)$ is now partly overlapping the $x(\tau)$ as shown in the figure below.



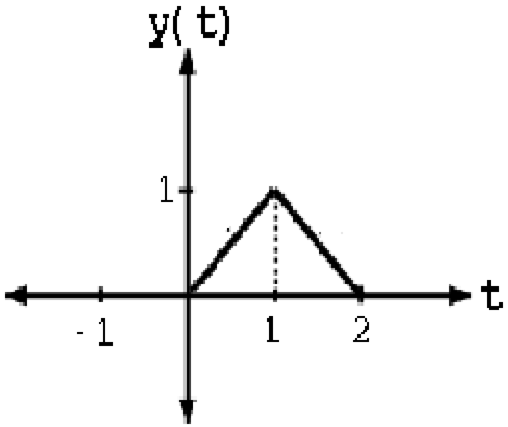
- According to the **convolution** $y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau$, the value of $y(t)$ at a specific $t = 0.5$ can be (informally) determined from the combination of the product function $x(\tau)h(t-\tau)$ for all τ. In other words, the value of $y(t)$ can formally be determined from the overlapping area within the range of $0 \le \tau \le 0.5$. This is because over the range the product function is non zero.
- $y(t = 0.5) = \int_{-\infty}^{\infty} x(\tau)h(0.5-\tau)d\tau = 0.5$.
- By implementing the same idea, $y(t = 0.5) = \int_{-\infty}^{\infty} x(\tau)h(0.5-\tau)d\tau = 0.5$.

$y(t = 1.0) = \int_{-\infty}^{\infty} x(\tau)h(1.0-\tau)d\tau =$ ….. **1.0**

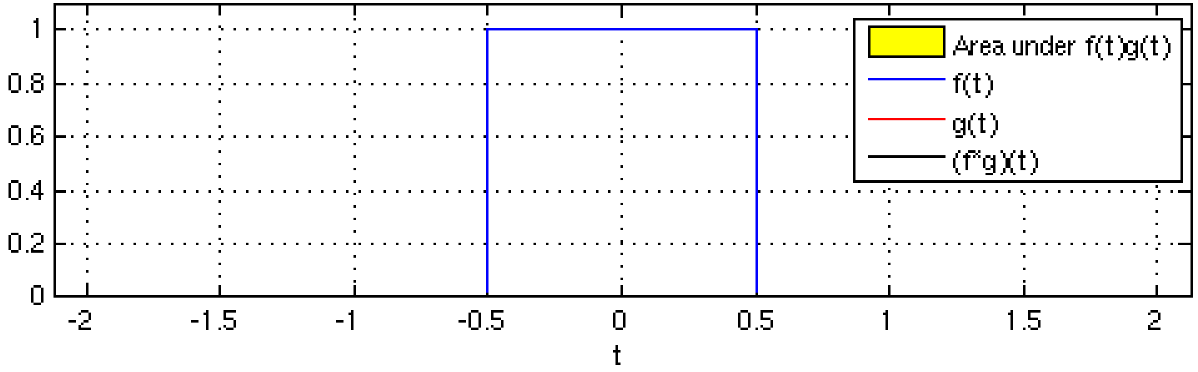
$y(t = 1.5) = \int_{-\infty}^{\infty} x(\tau)h(1.5-\tau)d\tau =$ ….. **0.5**

and $y(t = 2.0) = \int_{-\infty}^{\infty} x(\tau)h(2.0-\tau)d\tau =$ ….. **0.0**

- From specific steps of *t*, we can combine them together and the graph of the convolution function $y(\mathrm{t}) = x \otimes h$ is given as



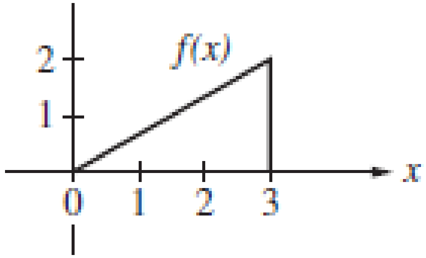
- In conclusion, when convolving two square pulses, the result will always be a **triangular pulse**.
- The width of the resulting convolution will always **equal the sum of the overall widths** of the two contributing functions.

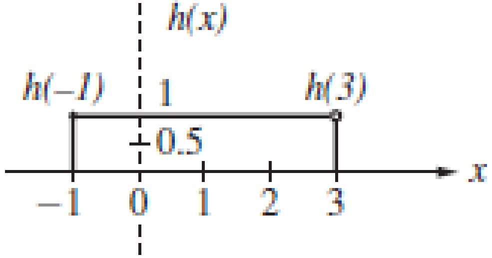


https://upload.wikimedia.org/wikipedia/commons/6/6e/Convolution_of_box_signal_with_itself.gif

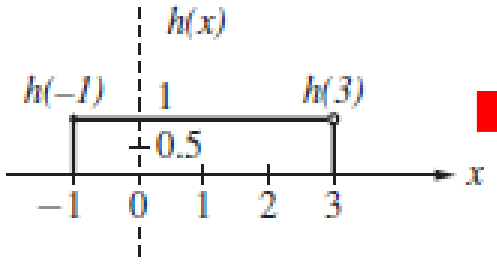
Example 2

- Now consider the two asymmetrical spatial function *f*(*x*) and *h*(*x*).
- Determine the 1-D convolution $g(X) = \int_{-\infty}^{\infty} f(x)h(X-x)dx$

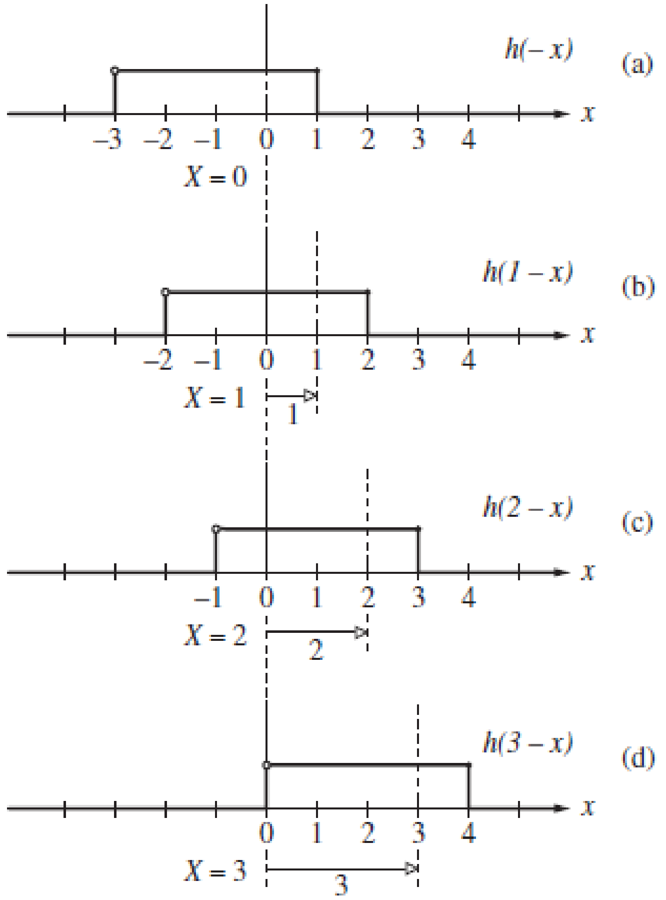




1^{st} step : reflection



- $h(x)$ is chosen to sweep across $f(x)$.
- To find the reflection of $h(x)$ about y axis, multiply all inputs with -1.
- Therefore, $h(x)$ is just flipped around the y axis and becomes $h(-x)$.
- Generally, the flipped function can be written as $h(X-x)$.
- With an appropriate X, the function may sweep rightward or leftward.

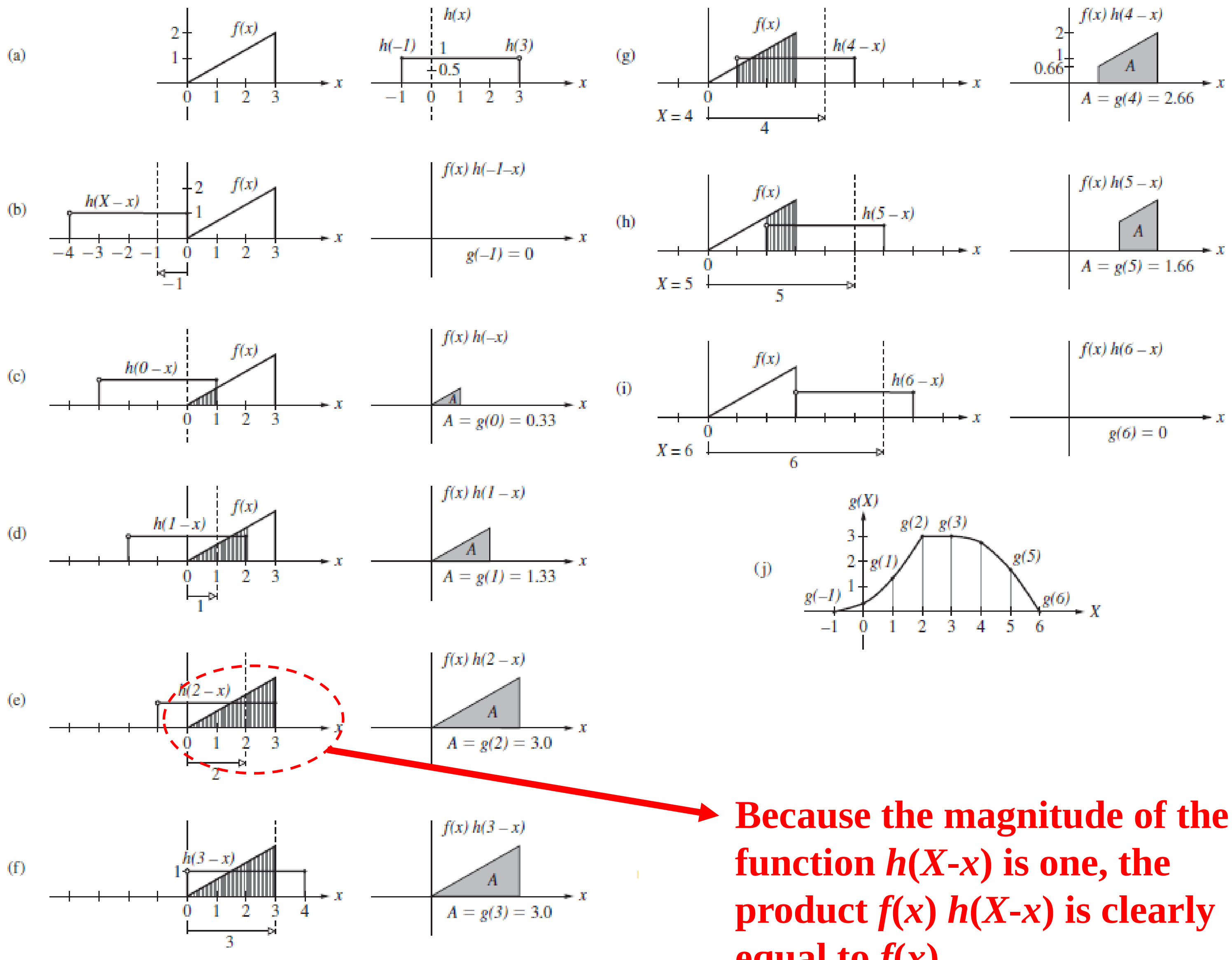


The function *h(X - x)* at *X* = 0, 1, 2, and 3

2nd step : shifting

- The *h(x)* is flipped or mirrored about the origin, becoming *h(-x)*.
- Note that *X* can be either positive or negative value depending on the shifting direction of the function *h*(*X*-*x*).
- Here, *X* = 0, 1, 2, and 3 and as a result, the rectangular pulse progresses to the right.

3rd step: Determination of $g(X)$

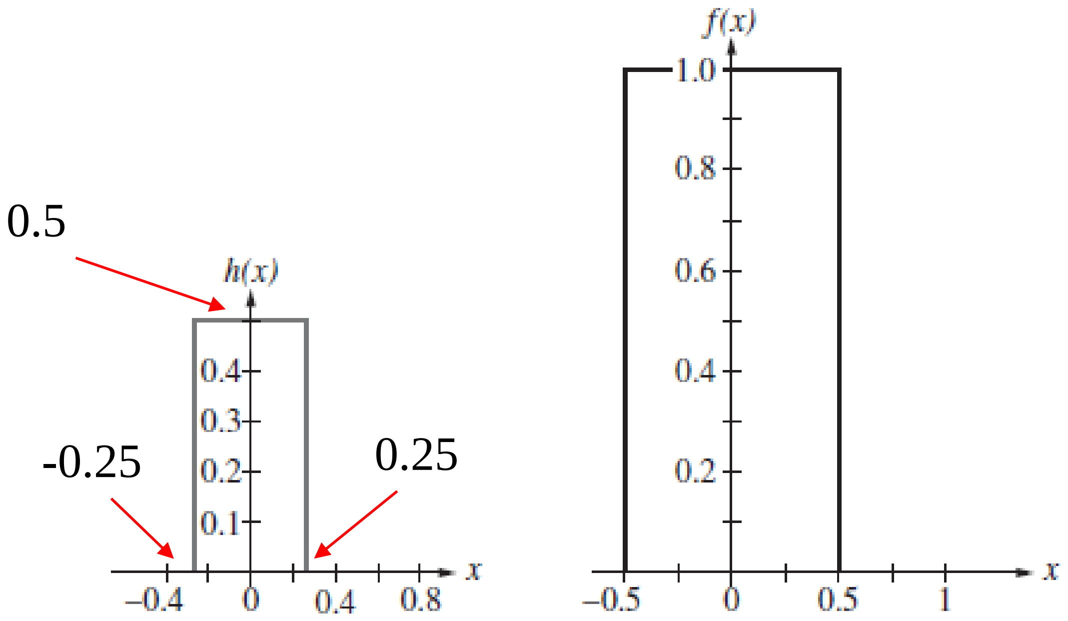


- From figs. (b) to (h), $g(X)$ becomes non zero because $h(X\text{-}x)$ sweeps over $f(x)$.
- $g(X)$ at particular values of X can be found from the integral over the product of the functions $f(x)\ h(X\text{-}x)$.
- The area, A, under the product curve is the value of $g(X)$ at that value of X.

Because the magnitude of the function $h(X\text{-}x)$ is one, the product $f(x)\ h(X\text{-}x)$ is clearly equal to $f(x)$.

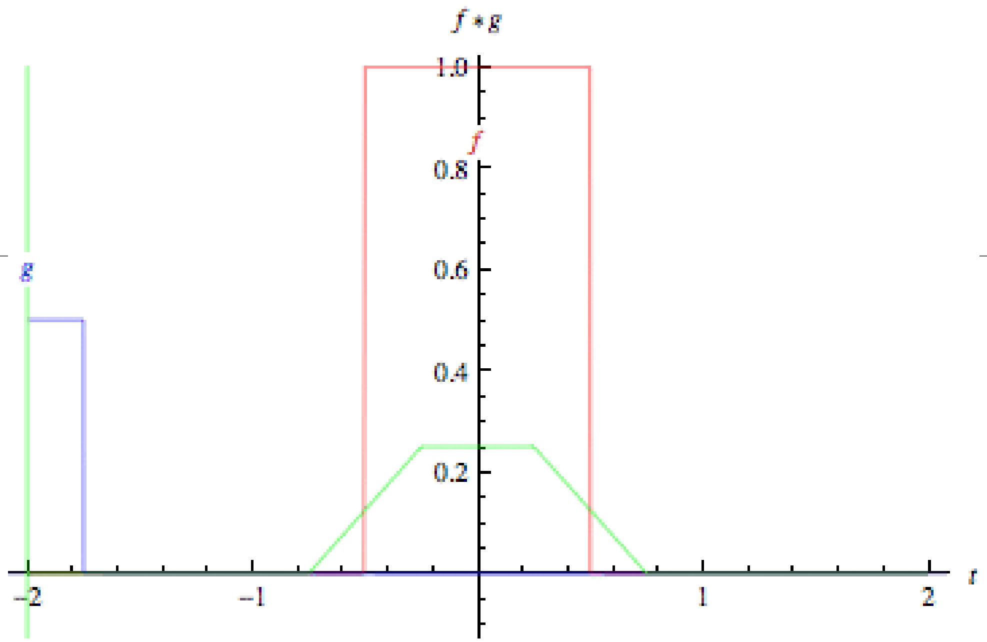
Example 3

- Consider the functions *f (x)* and *h(x)* depicted in the accompanying diagram.



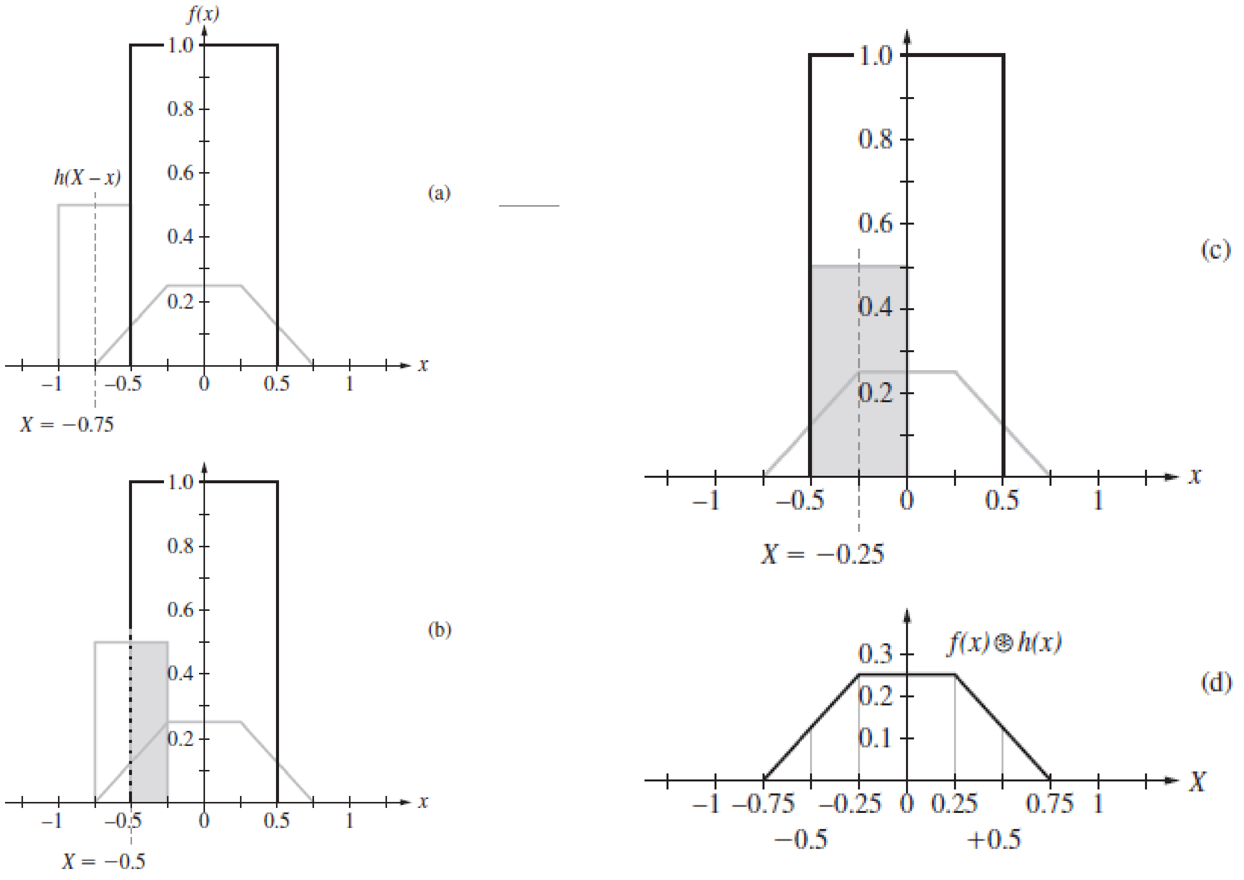
- Can you guess the convolution function $g(X)$.
- Graphically convolve those two functions, explaining each step of the process.

Example 3



- The animations above graphically illustrate the convolution of two boxcar functions.
- In the plot, the **green curve** shows the convolution of the **blue** and **red** curves as a function of *x*, the position indicated by the vertical green line.
- The gray region indicates the product $f(x)h(X-x)dx$ as a function of *X*, so its area as a function of *X* is precise the convolution.

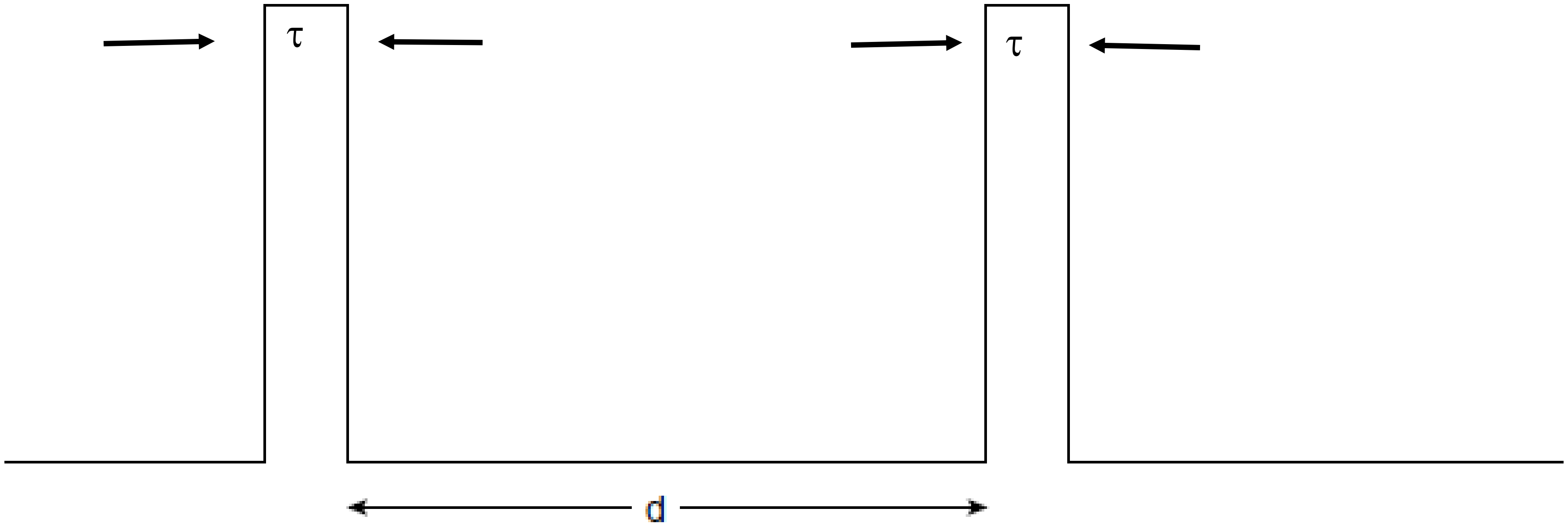
Example 3



- The symmetric $h(x)$ is chosen to sweep across $f(x)$.
- The full form of the flipped $h(x)$ is $h(X\text{-}x)$.
- As ever, $h(X\text{-}x)$ is shifted to a starting point where the functions contact. This can be achieved by choosing X = -0.75 (fig. (a)).
- At this specific X, $g(X) = f(x) \otimes h(x) = 0$.
- As $h(X\text{-}x)$ sweeps across $f(x)$, overlapping between the two functions occur and $g(X)$ can be worked out accordingly.
- For example, at X = -0.5 (fig. (b)), $f(x)h(-0.5 - x)$ over -0.5 ≤ x ≤ -0.25 is found to be 0.125 (= 0.5 × 0.25).
- At X = -0.25 (fig. (c)), $f(x)h(-0.25 - x)$ over -0.5 ≤ x ≤ 0 is found to be 0.25 (= 0.5 × 0.5).
- Eventually, the complete convolution is obtained as illustrated in fig. (d).

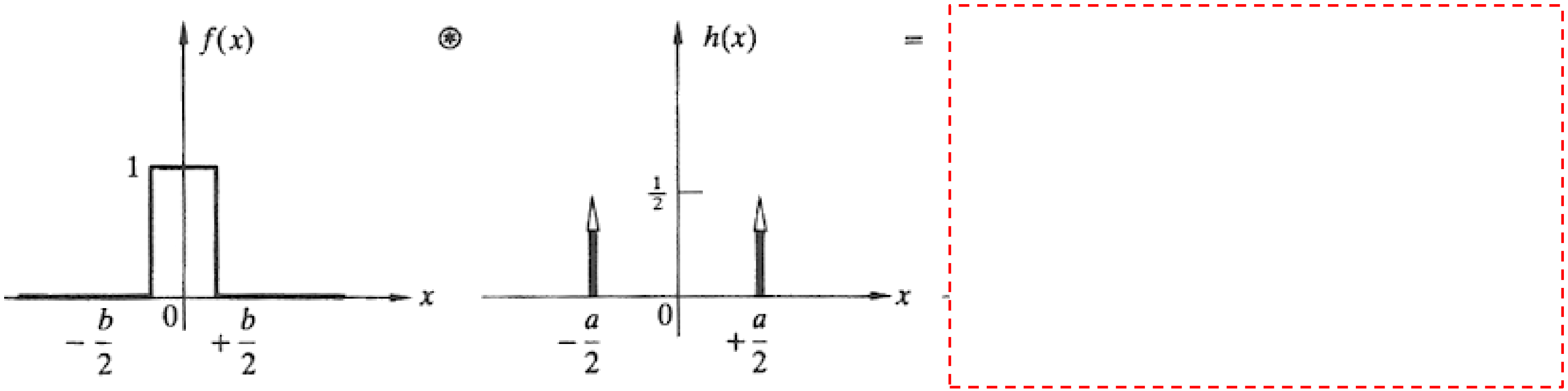
Example 4 : self convolution = autoconvolution

Sketch the self-convolution of the double slit function shown below



Example 5
Aperture function of a double slit

- Regarding the double slit, in practice, each aperture actually has some finite shape.
- A method of **convolution** can be used to create a proper **aperture function**, e.g., actual double slit.
- The **aperture function** $\boldsymbol{g(x)}$ is obtained by convolving the δ function spike $h(x)$, that locate each slit, with the rectangular pulse, $f(x)$ that corresponding to the particular opening.



Convolution theorem

- Suppose we have two function $f(x)$ and $h(x)$ with Fourier transform $F(f(x)) = F(k)$ and $F(h(x)) = H(k)$, respectively.
- The convolution theorem states that if $g = f \otimes h$

$$F\{g\} = F\{f \otimes g\} = F\{f\} \,.\, F\{h\}$$

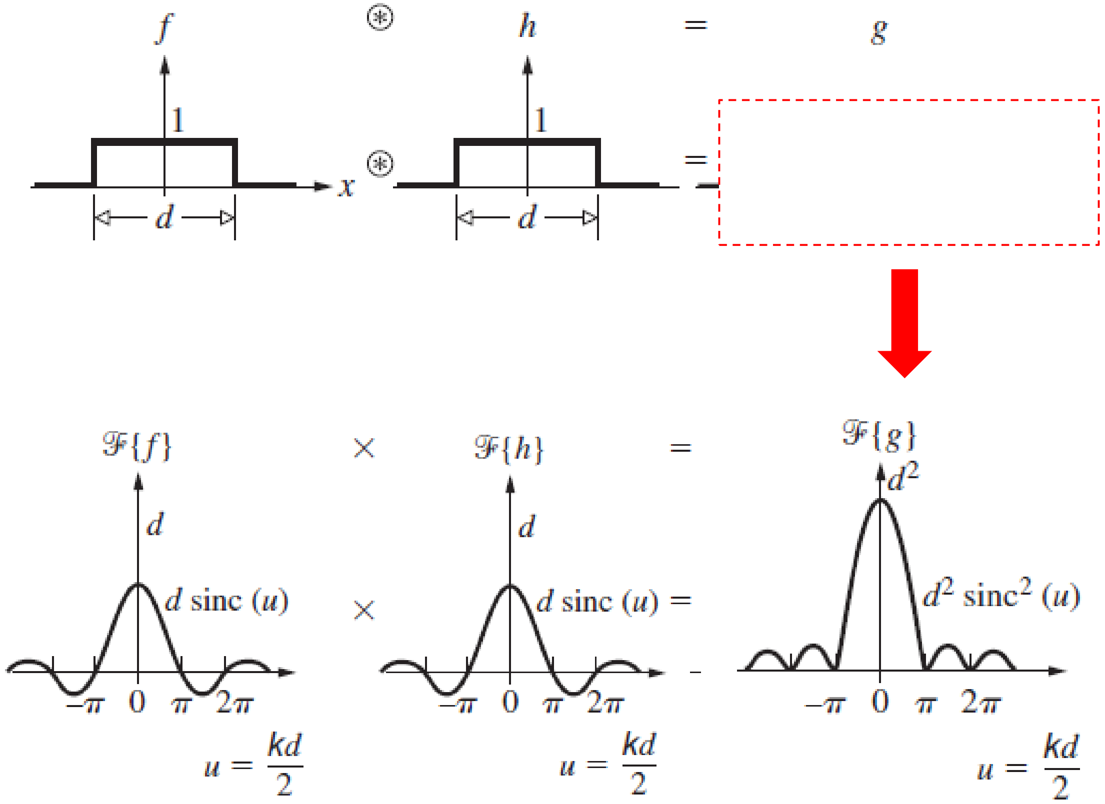
Or

$$G(k) = F(k)H(k)$$

Where $F\{g\} = G(k)$.

"The transform of the convolution of two functions is the product of their transforms."

Example 6
Proof of the convolution theorem



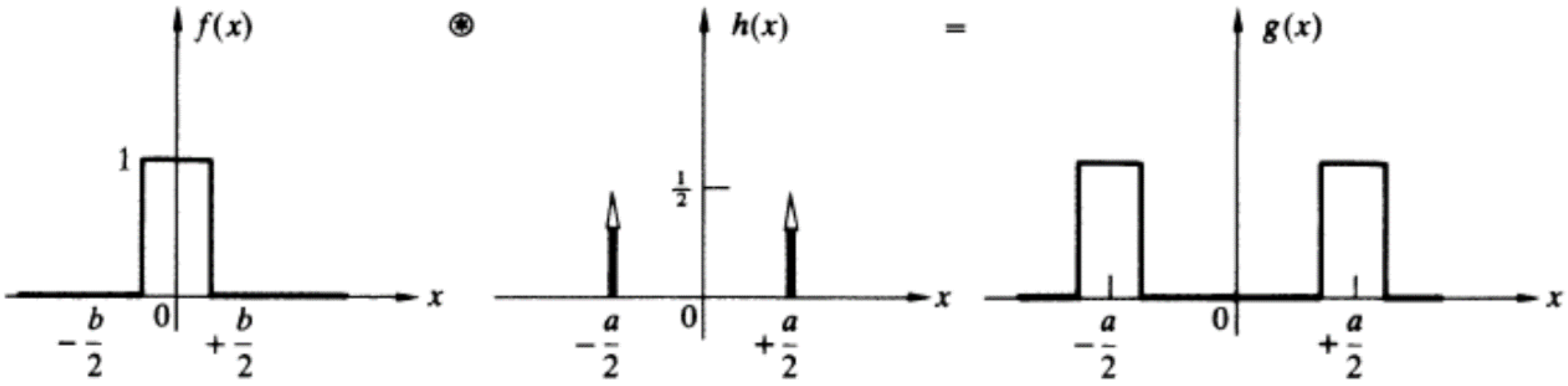
- Starting with the convolution between two square.
- This gives **a triangular pulse**.
- Now, find the Fourier transform of g (F(g)).
- Determine the Fourier transform of each square.
- Regarding to the convolution theorem:

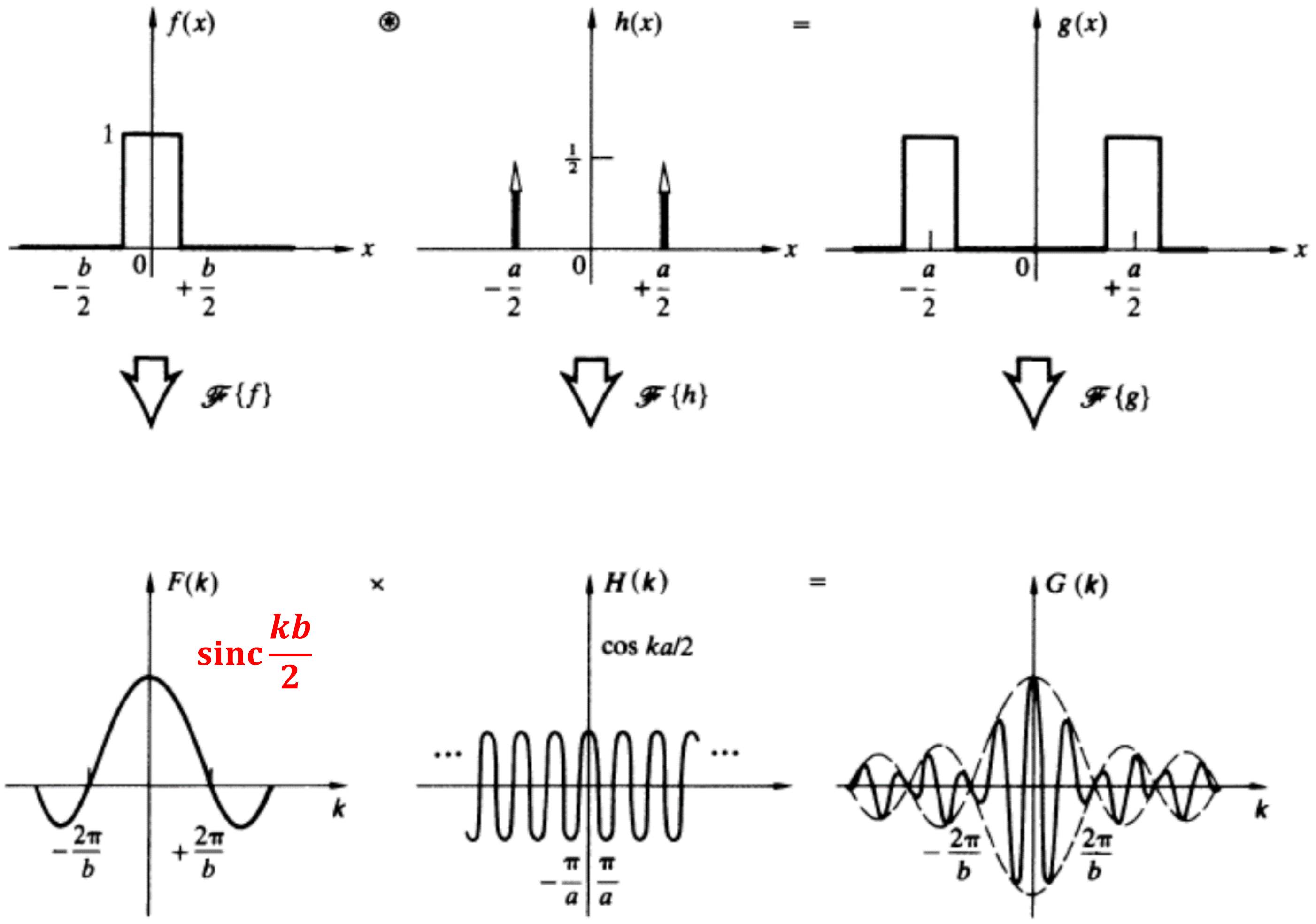
  $F\{g\} = F\{f \otimes g\} = F\{f\} \,.\, F\{h\}$

- This example clearly proves the theorem!

Example 7
Convolution theorem and double slit diffraction

- The aperture function can be derived from the convolution between a rectangular pulse and two symmetric delta functions.





Solution

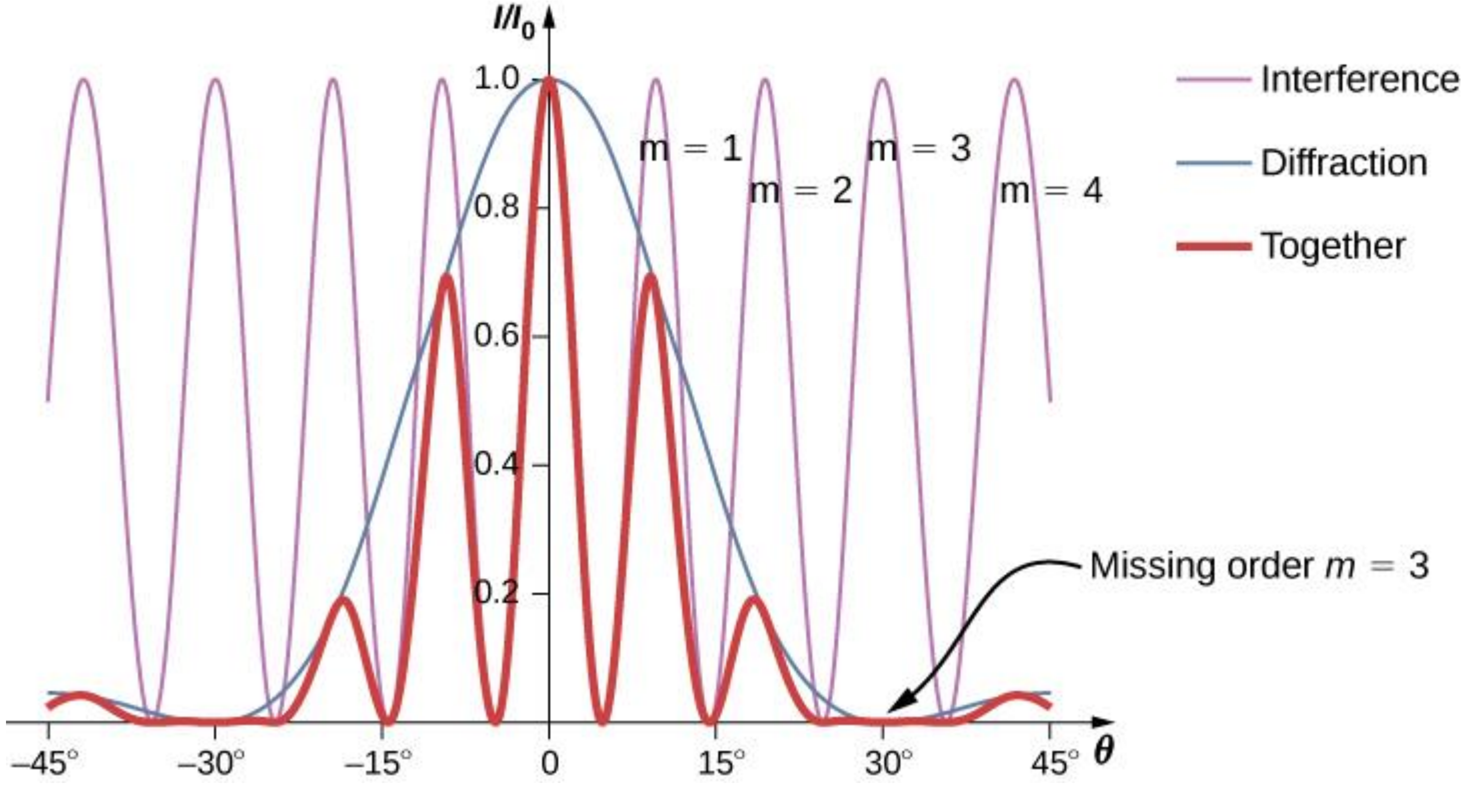
- Recall the intensity of the double slit diffraction : $I(\theta) = 4I_0\left(\frac{sin^2\beta}{\beta^2}\right)cos^2\alpha$.

- The convolution theorem:

- F{g} = F{f⊗g} = F{f} . F{h}

show that the Fourier transform of the aperture function, F(g(x)) = G(k), representing the field distribution of the diffraction can be found from the products of F(k) and H(k).

Intensity from double slit diffraction



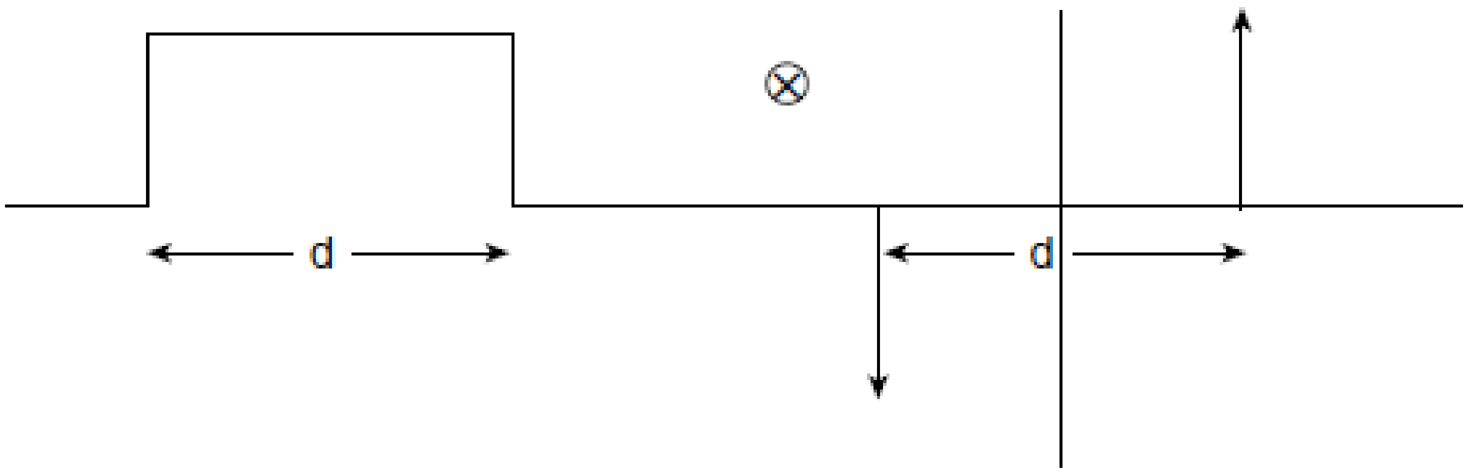
$: I(\theta) = 4I_0\left(\frac{sin^2\beta}{\beta^2}\right)cos^2\alpha$

$\alpha = \frac{ka}{2}, \beta = \frac{kb}{2}$

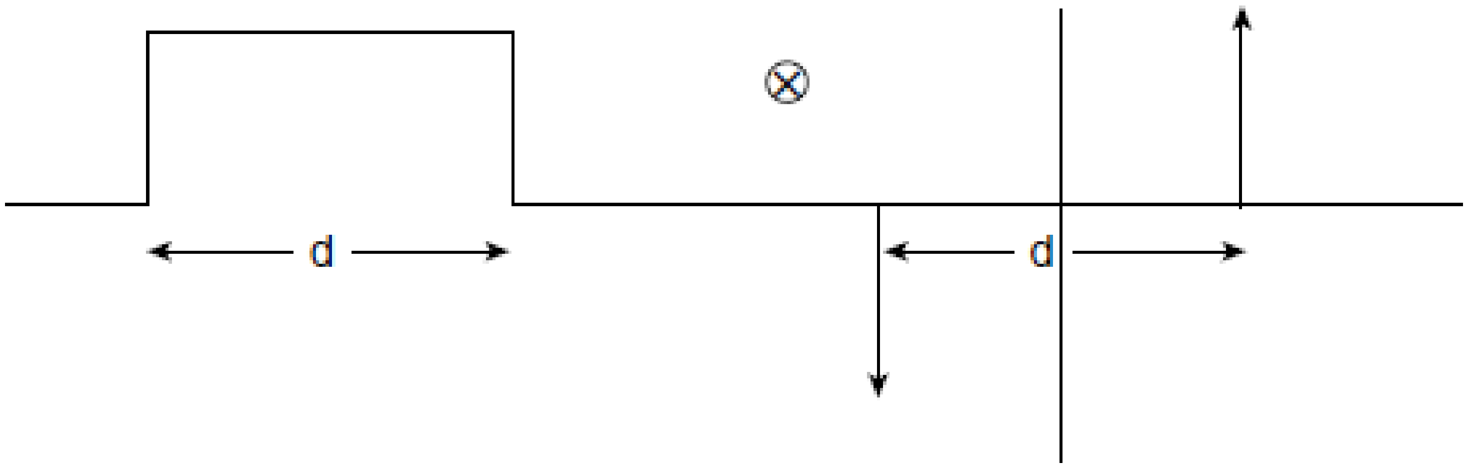
a = slit space
b = slit width

Example 8

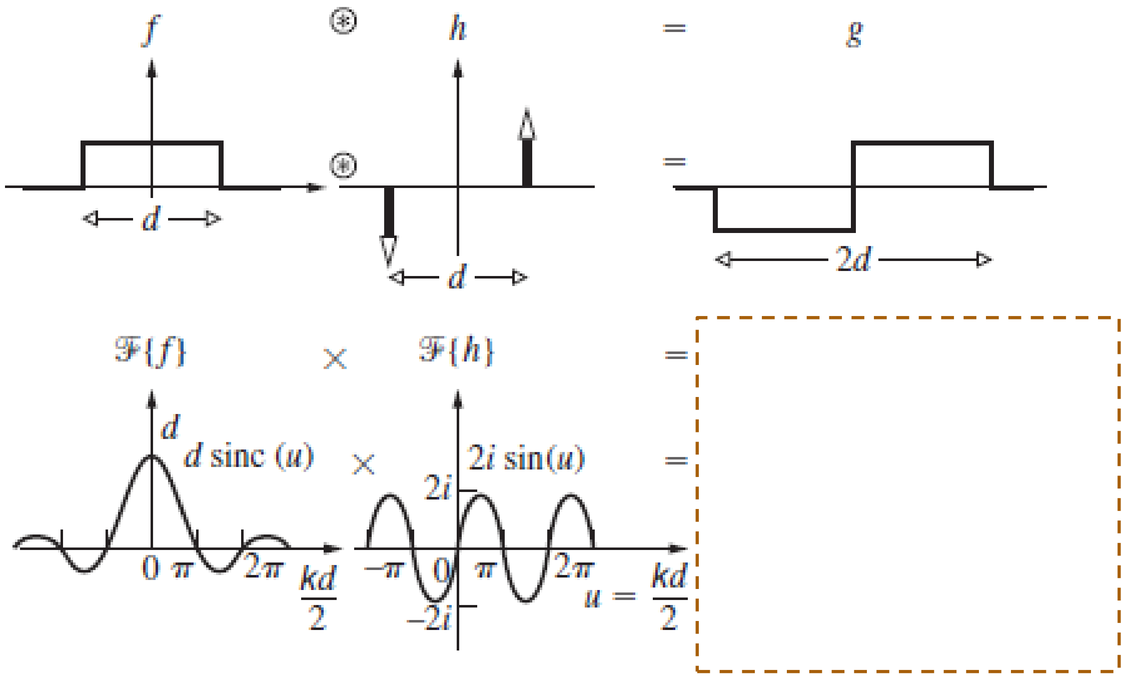
- **Sketch the convolution of the two functions in the figure below.**
- **Find the Fourier transform of the convolution by using the convolution theorem.**



Solution



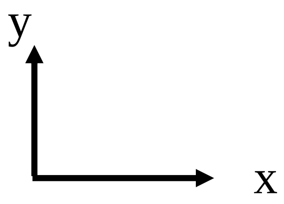
Solution (cont.)

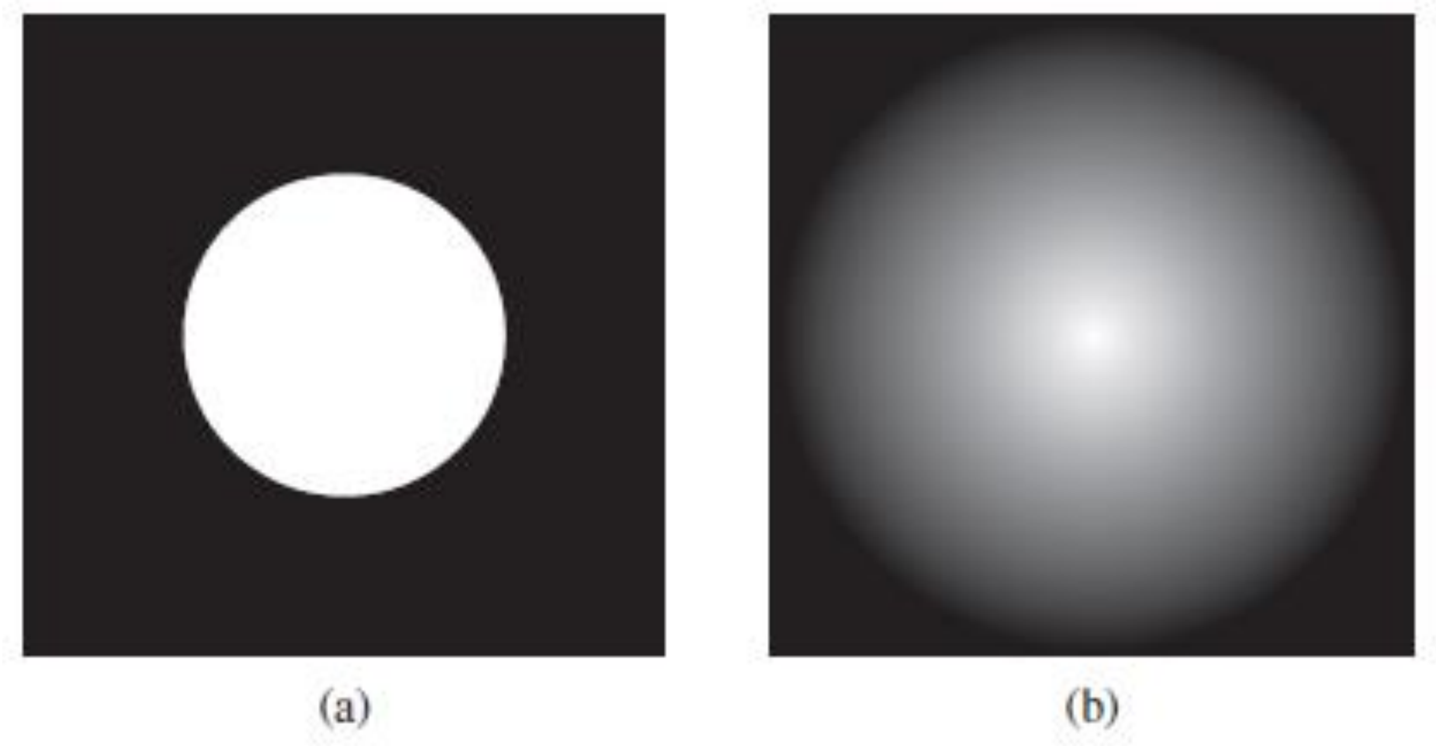


- The Fourier transform of the convolution can be found by applying the convolution theorem.
- This means that the Fourier transform of the rectangular pulse and two asymmetric delta functions have to be determined.
- Then the product of the Fourier transform is the Fourier transform of the convolution according to the convolution theorem.
- The procedure mentioned can be roughly illustrated by the figure on the left.

Example 9 Autoconvolution of a uniformly illuminated circular hole

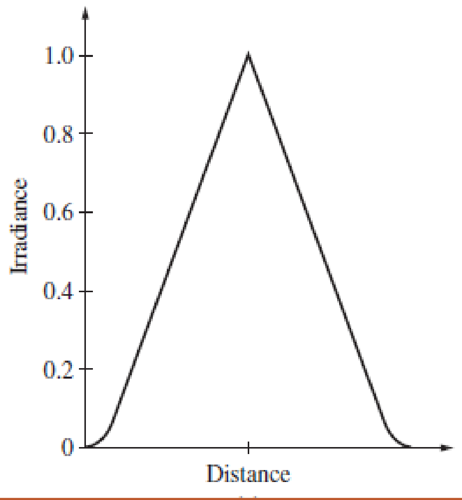
- The circular, uniformly illuminated hole in an opaque screen has to be described by an aperture function $f(x,y)$.
- Therefore, to find the autoconvolution, we just have to sweep one circle over the other and record the product area at each displacement.
- The irradiance of the convolution can be described by this graph.



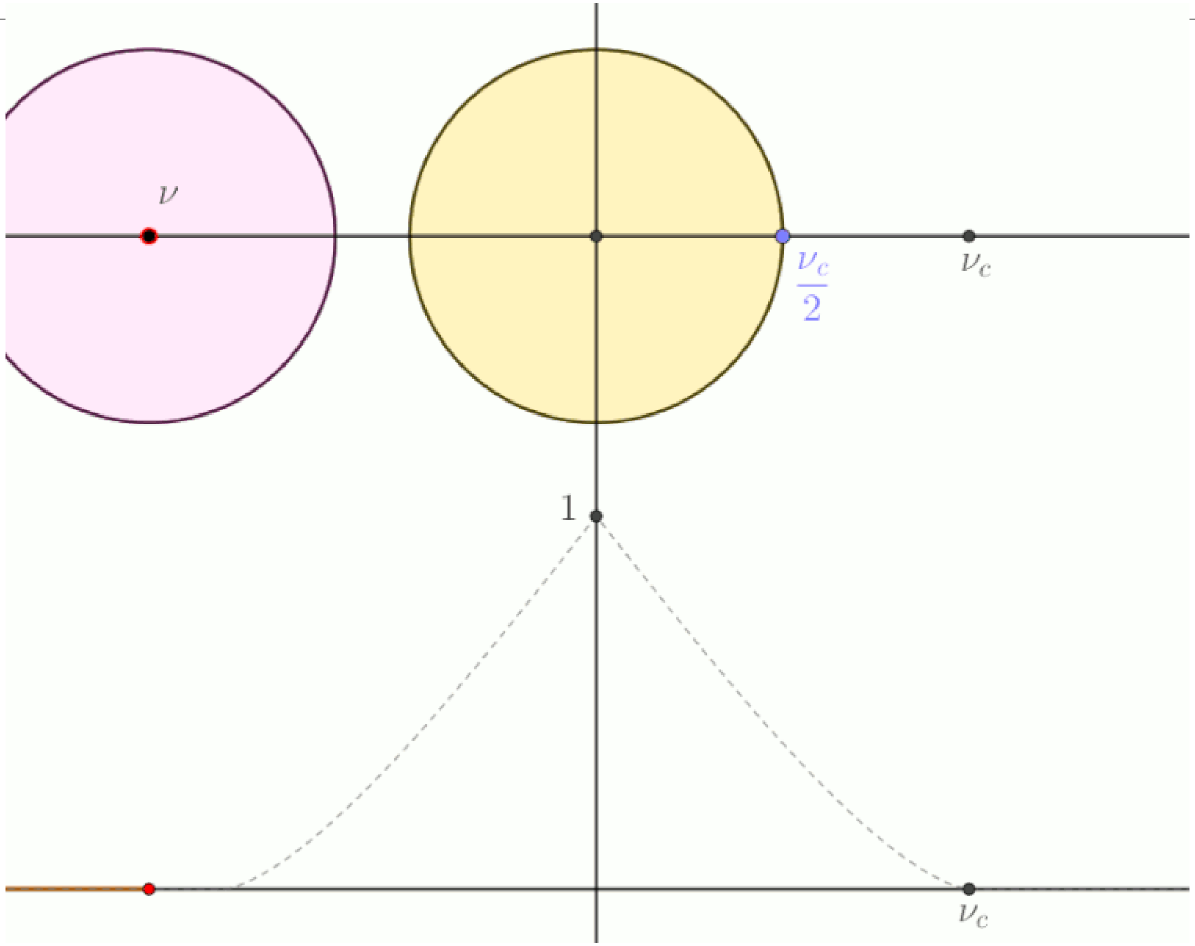


Uniformly illuminated circular hole

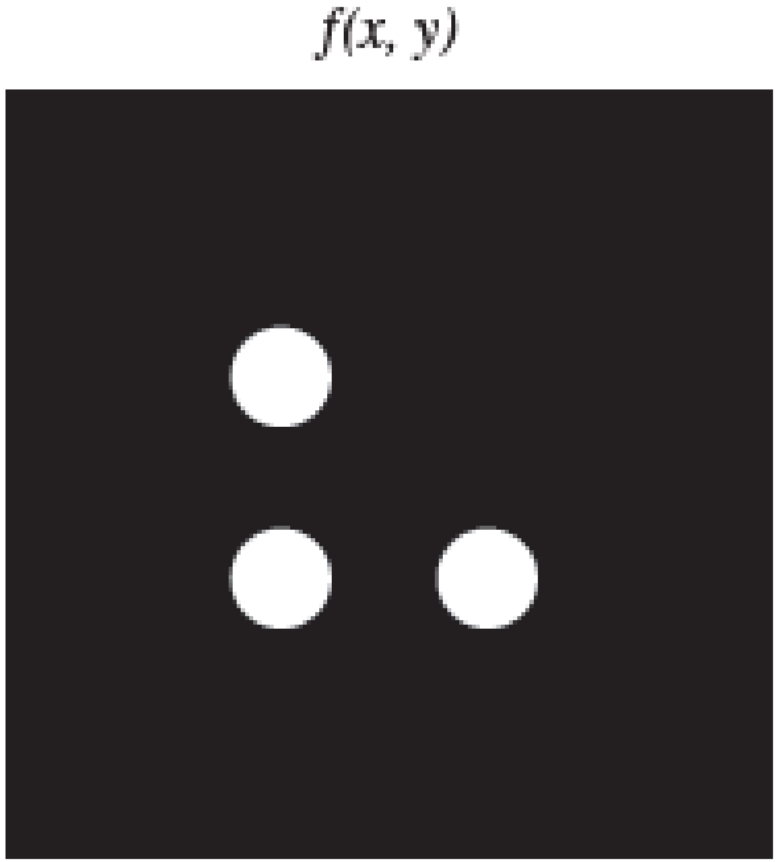
autoconvolution



The auto convolution of uniformly illuminated circular hole



https://commons.wikimedia.org/wiki/File:Autoconvolution_disk.gif



Example 10
2-D Autoconvolution

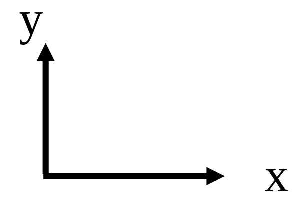
- The figure given here is the two-dimensional signal.
- Find the autoconvolution of the signal by using a graphical method.

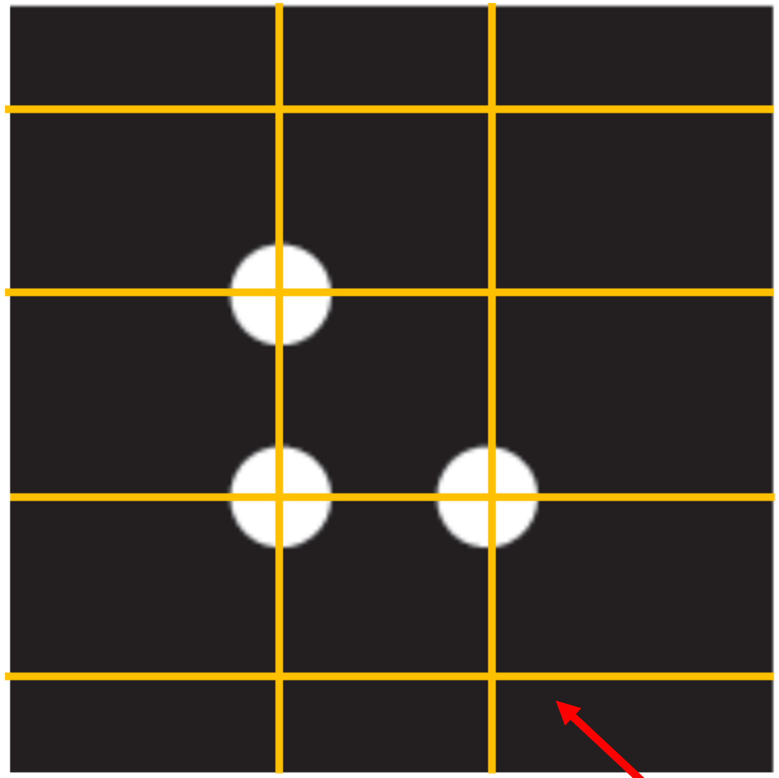
Recall 3 steps to find the convolution

1. **Reflection : in this case, the reflection must be done about vertical and horizontal axes due to 2-D operation,**
2. **Shifting horizontally and vertically, and**
3. **Record the product at each displacement.**

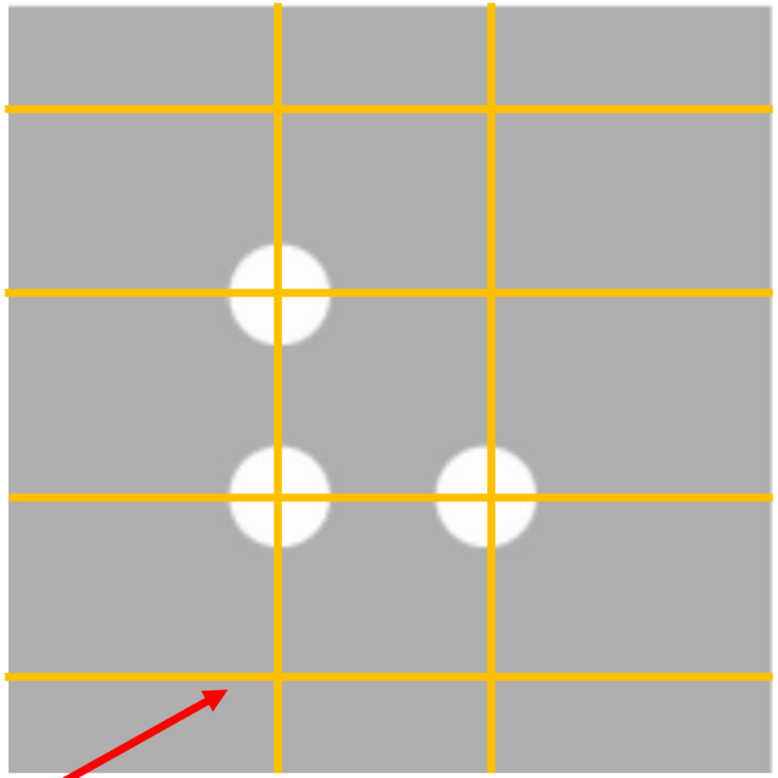
From this point on the procedure for finding 2-D convolution by a graphical method will be described.

(1) First of all let's define what we need to perform the 2-D convolution graphically. Since what we are going to do is the autoconvolution, the starting functions are identical. One function composed of three uniformly illuminated circular dots arranged in L shape is designated as *f*(x,y) while the other identical function is assigned as *h*(x,y). The convolution g(X,Y) is prepared on the right.

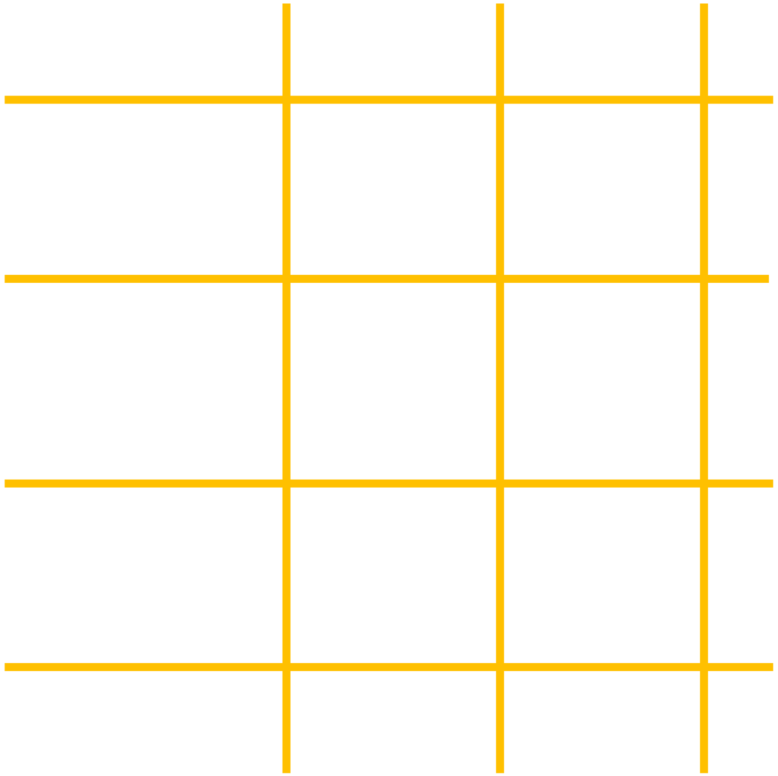




$f(x,y)$



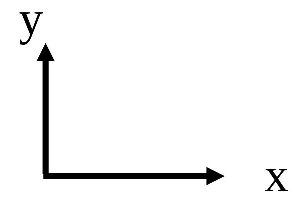
$h(x,y)$

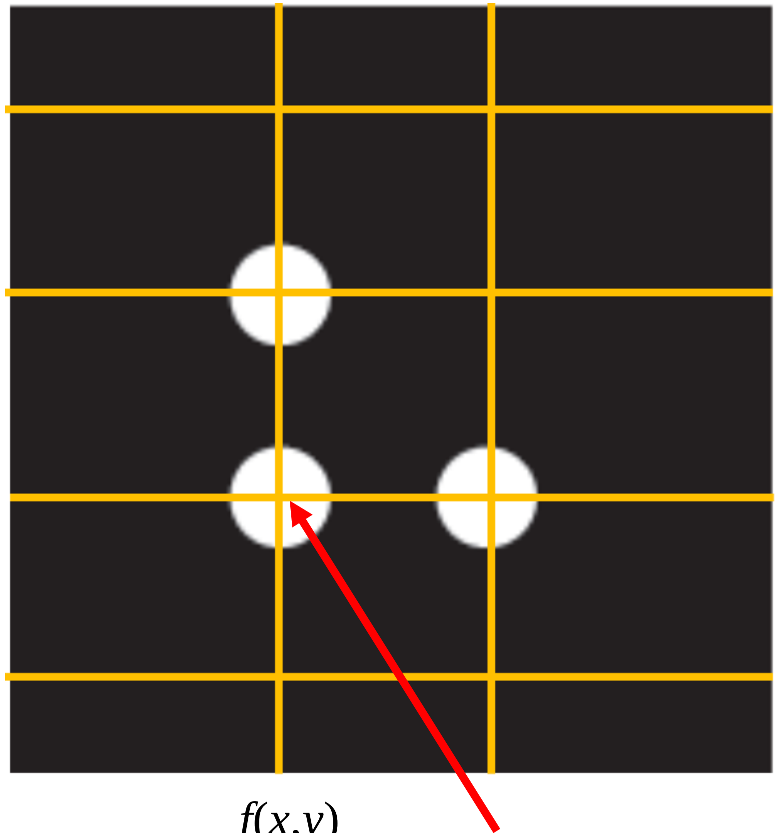


$g(X,Y)$ **2-D autoconvolution**

Grid is introduced to make the shifting easy to follow.

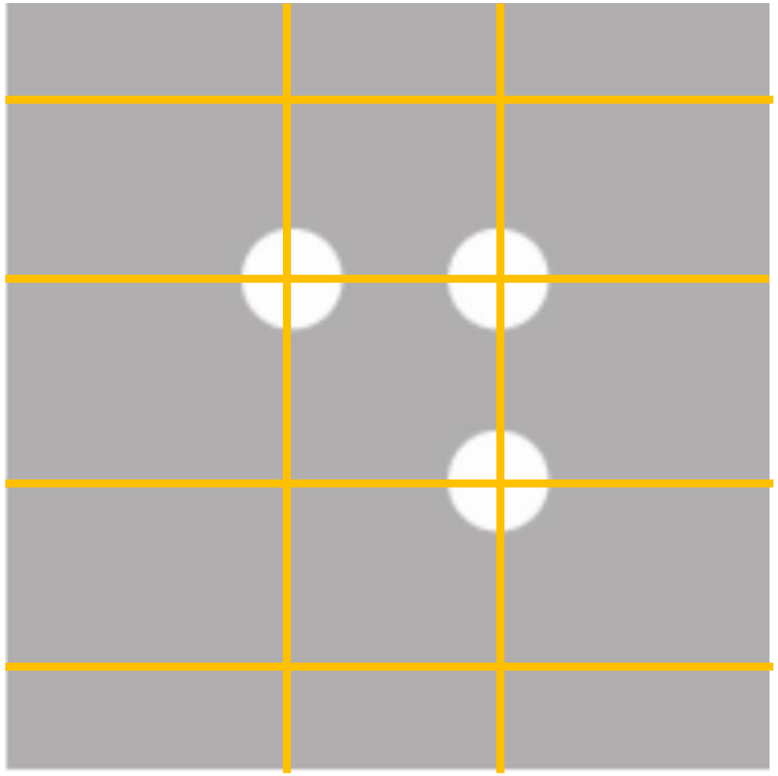
(2) Find the reflection of $h(x,y)$ twice. One is the reflection about y axis (= horizontal flip) and the other is the reflection about x axis (vertical flip). By doing the horizontal flip $h(x,y)$ becomes $h(-x, y)$. Likewise, by doing the vertical flip $h(-x,y)$ becomes $h(-x,-y)$. Generally, this reflected function can be written as $h(X\text{-}x, Y\text{-}y)$. The function $h(X\text{-}x, Y\text{-}y)$ represented by L faces to the left. This is equivalent to rotate the $f(x,y)$ by 180^0 around the dot at the origin.



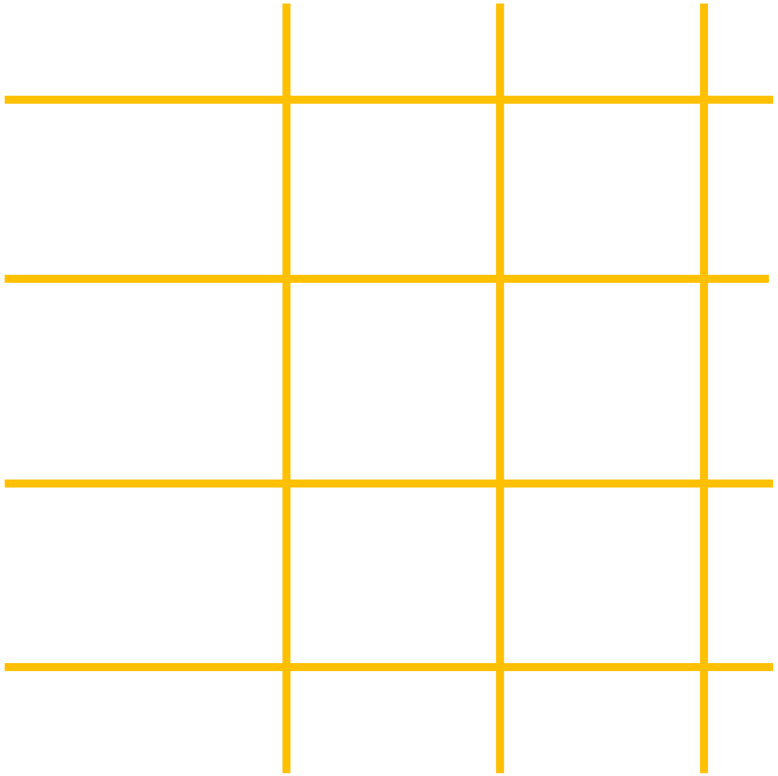


$f(x,y)$

This dot is assumed to locate at the origin.

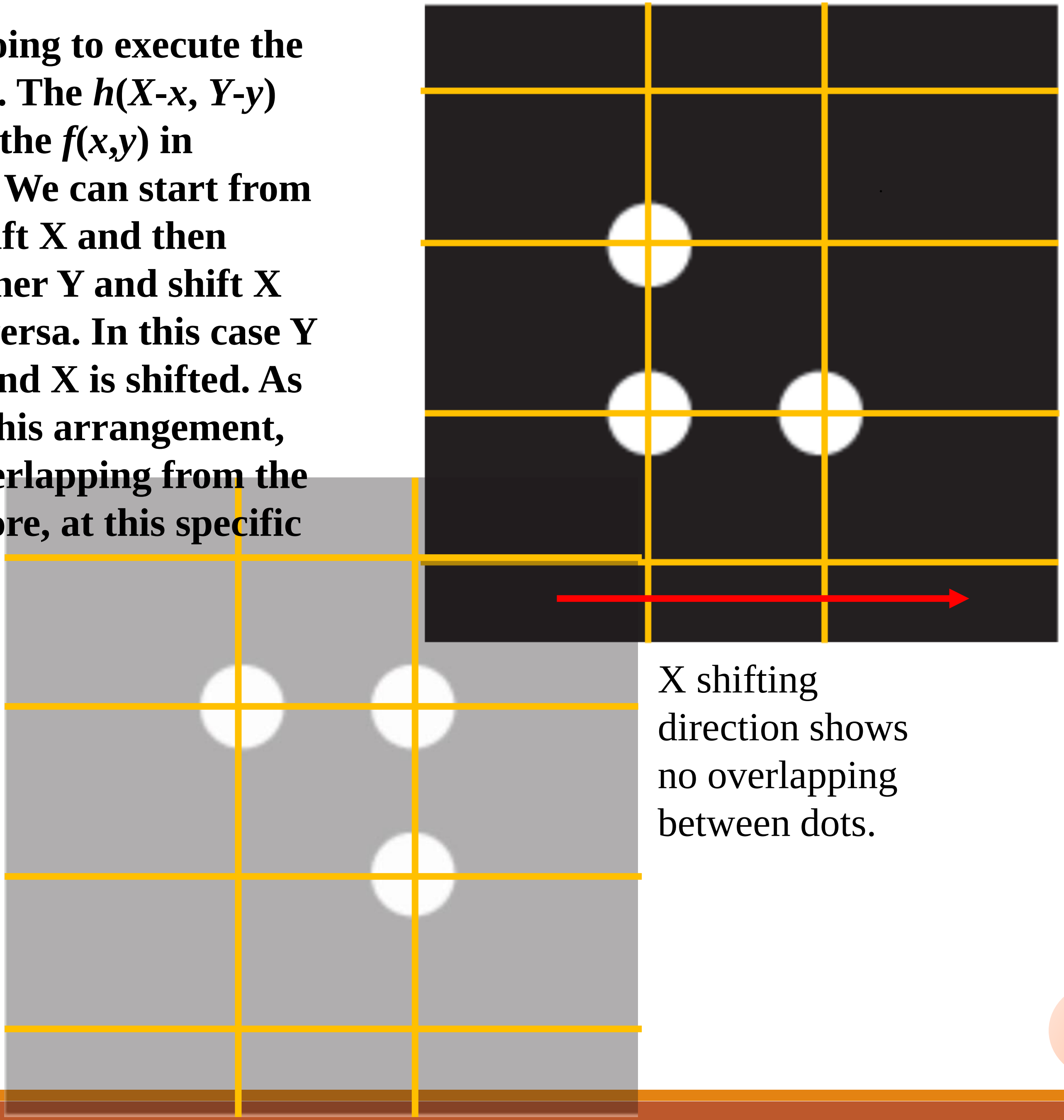


$h(x,y) \rightarrow h(X\text{-}x, Y\text{-}y)$

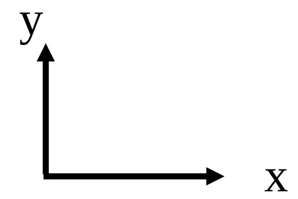


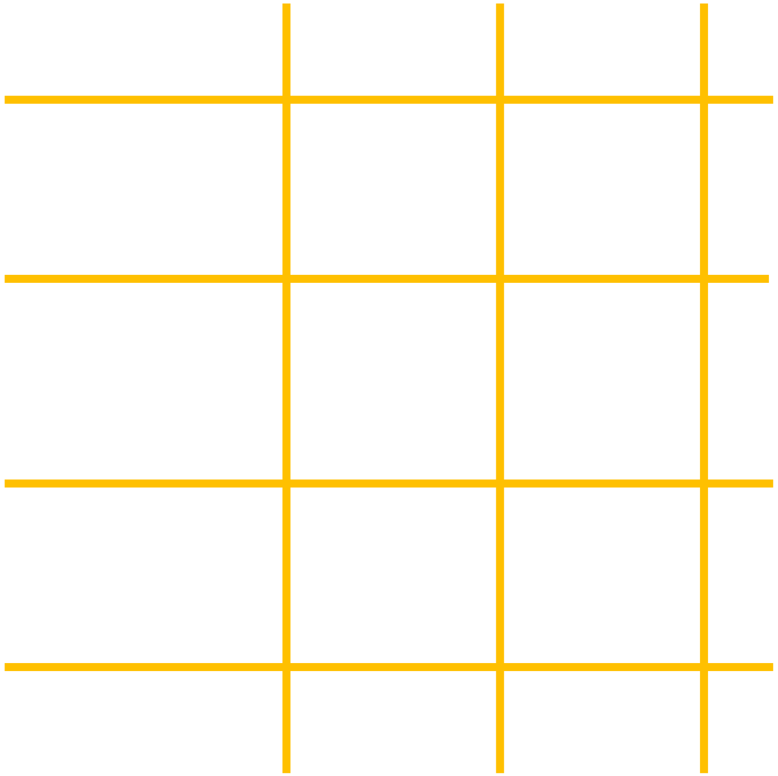
$g(X,Y)$ **2-D autoconvolution**

(3) Now, we are going to execute the shifting procedure. The *h*(*X*-*x*, *Y*-*y*) will be swept over the *f*(*x*,*y*) in directions x and y. We can start from choosing Y and shift X and then moves up for another Y and shift X and so on or vice versa. In this case Y is firstly selected and X is shifted. As can be seen from this arrangement, at a fixed Y, no overlapping from the X shifting. Therefore, at this specific *g*(X,Y) = 0.



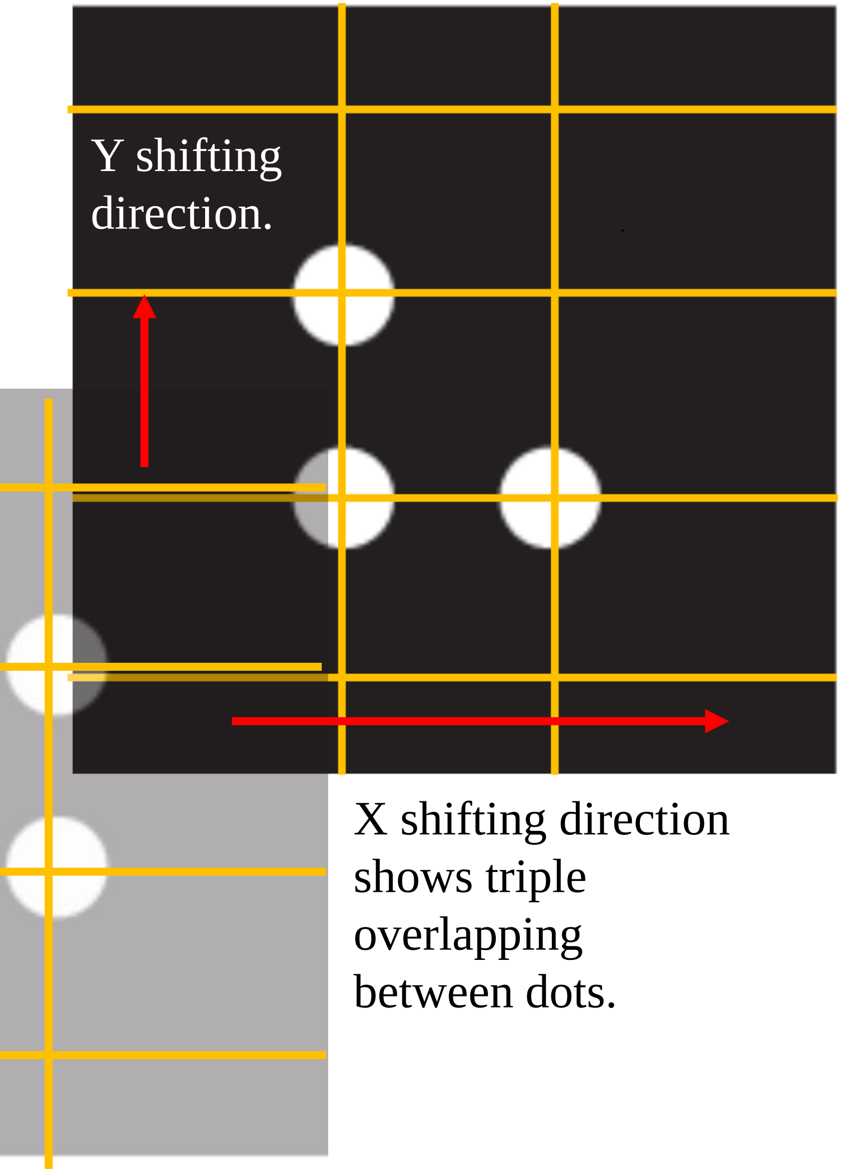
X shifting direction shows no overlapping between dots.



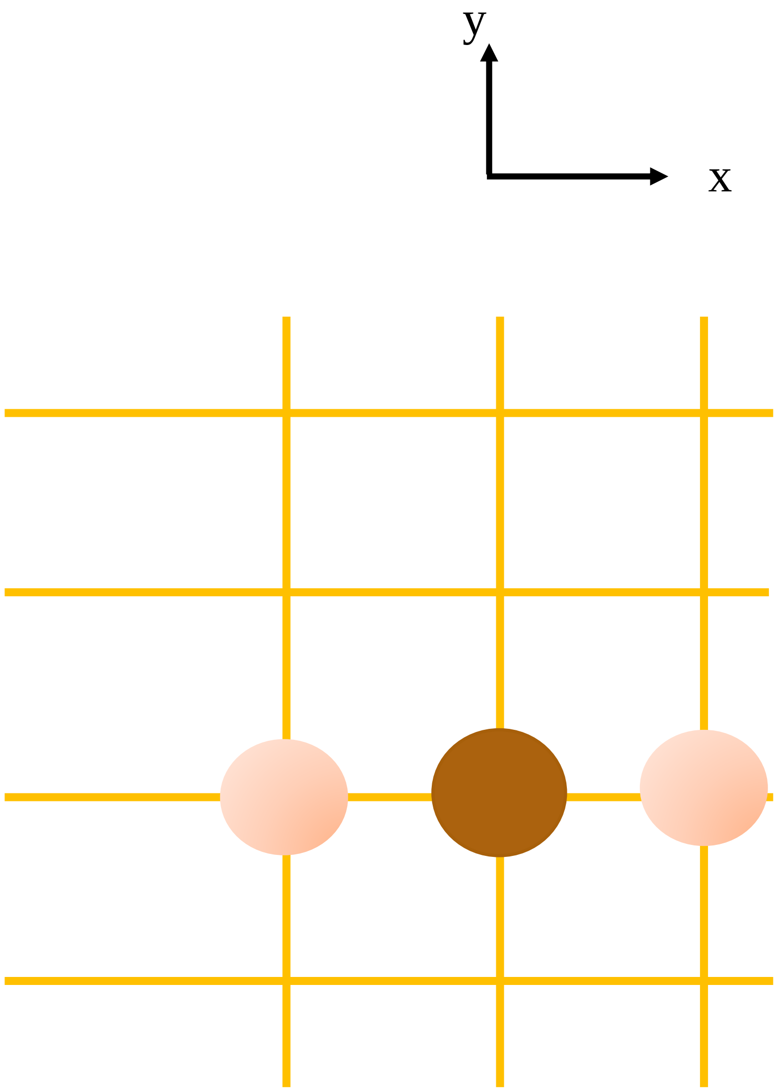


$g(X,Y)$ **2-D autoconvolution**

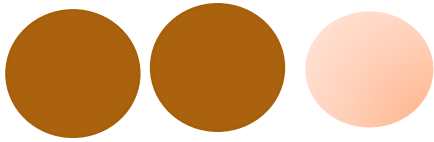
(4) **The shifting process is repeated. Now move the *h*(*X*-*x*, *Y*-*y*) up in the +y direction by one interval and then shift the function rightward. This step is equivalent to choose a Y (fixed Y) and then step X. One step is equal one interval. This is clear that by shifting to the right side, three overlapping occur. The first overlap is composed of one pair of dots, the second overlap is composed of two pairs of dots and the last overlap is composed of one pair of dots. At the first overlap located at specific coordinate (X,Y), the convolution g(X,Y) becomes non zero and is assigned as a pale circular dot due to one pair of overlapping dots. At the second overlap located at the next specific coordinate (X,Y), the convolution g(X,Y) again becomes non zero and is assigned as a dark circular dot due to 2 pairs of overlapping dots and the third overlap located at the next specific coordinate (X,Y), the convolution g(X,Y) again becomes non zero and is assigned as a pale circular dot due to 1 pair of overlapping dots.**



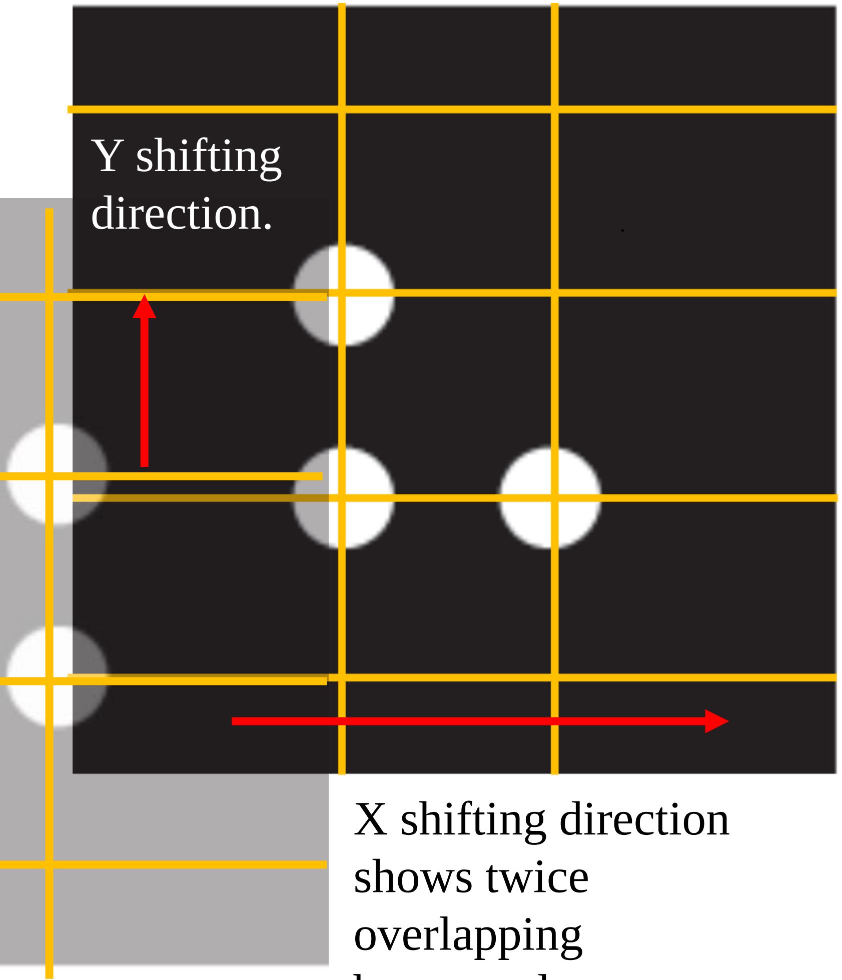
X shifting direction shows triple overlapping between dots.



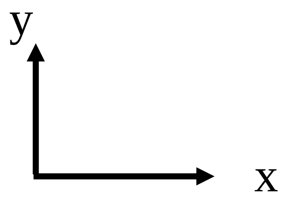
$g(X,Y)$ **2-D autoconvolution**

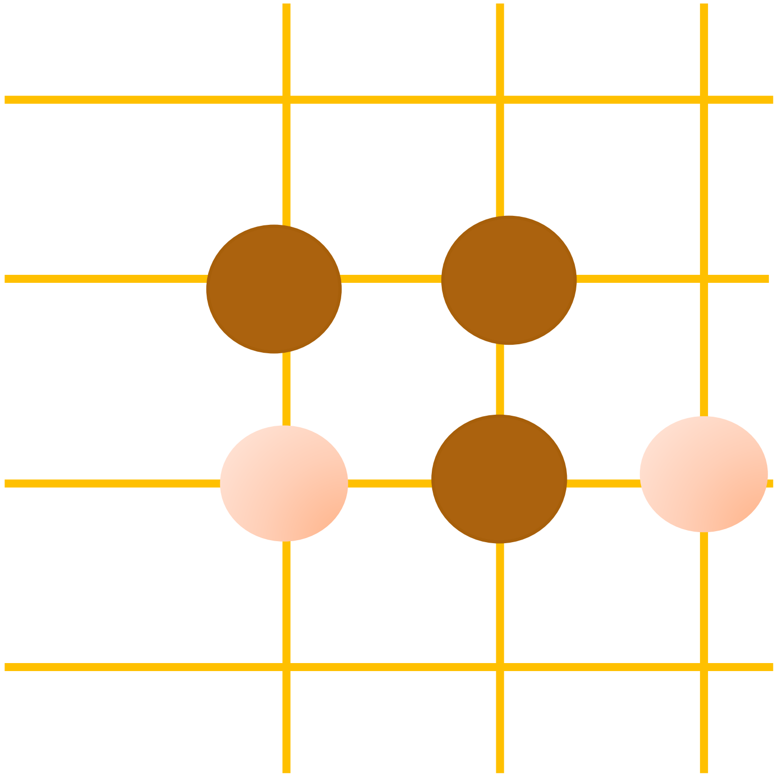


(5) The shifting process is repeated. Now move the $h(X\text{-}x, Y\text{-}y)$ up in the +y direction by one interval and then shift the function rightward. This step is equivalent to choose a Y (fixed Y) and then step X. One step is equal one interval. This is clear that by shifting to the right side, two overlapping occur. The first overlap is composed of two pair of dots and the second overlap is composed of two pairs of dots. At the first overlap located at specific coordinate (X,Y), the convolution g(X,Y) becomes non zero and is assigned as a dark circular dot due to two pair of overlapping dots. At the second overlap located at the next specific coordinate (X,Y), the convolution g(X,Y) again becomes non zero and is assigned as a dark circular dot due to 2 pairs of overlapping dots. And no more overlap occurs as the rightward shifting is made.

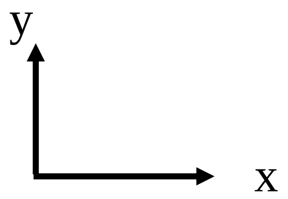


X shifting direction shows twice overlapping between dots.

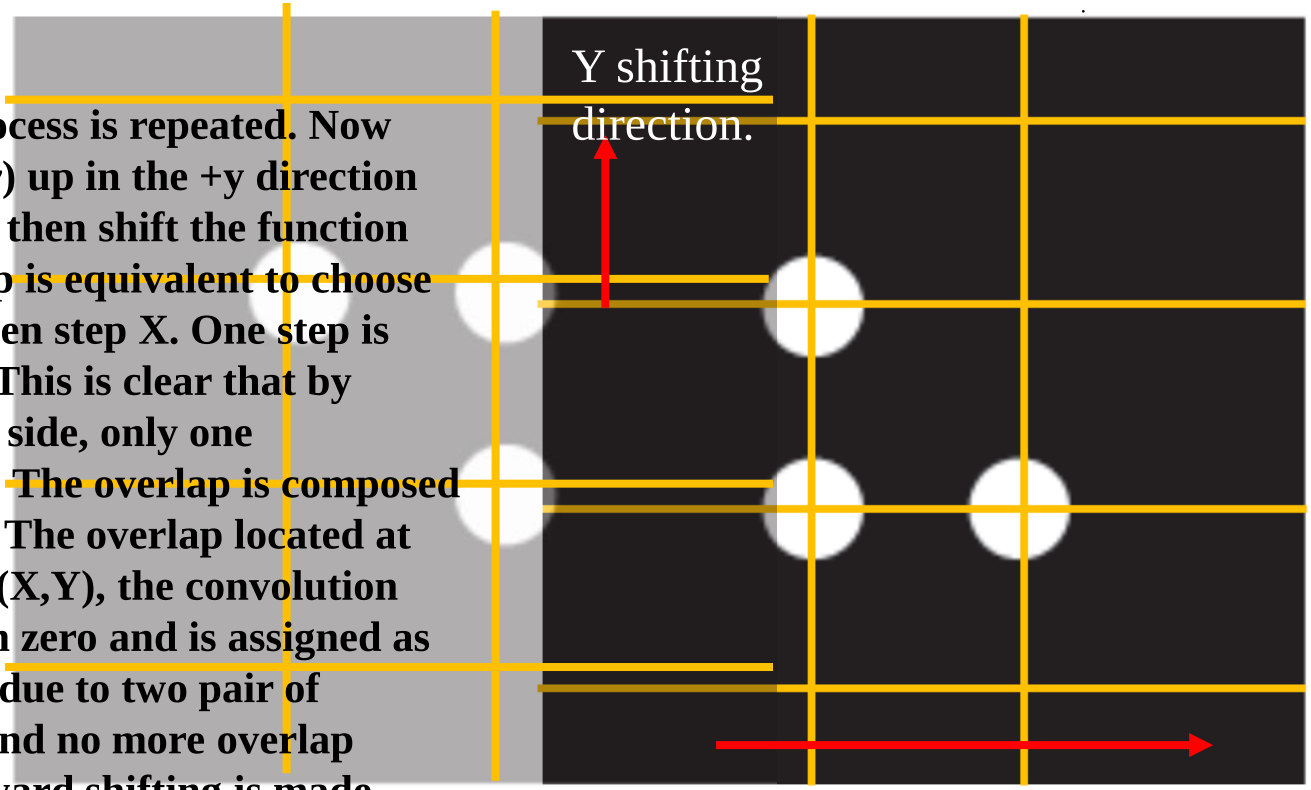




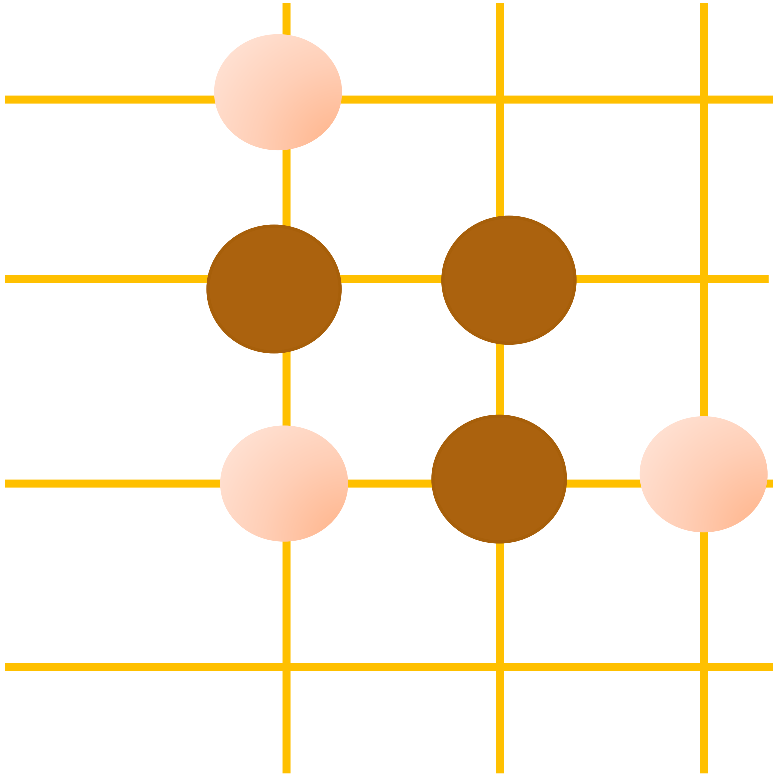
$g(X,Y)$ **2-D autoconvolution**



(6) The shifting process is repeated. Now move the $h(X\text{-}x, Y\text{-}y)$ up in the +y direction by one interval and then shift the function rightward. This step is equivalent to choose a Y (fixed Y) and then step X. One step is equal one interval. This is clear that by shifting to the right side, only one overlapping occurs. The overlap is composed of one pair of dots. The overlap located at specific coordinate (X,Y), the convolution g(X,Y) becomes non zero and is assigned as a dark circular dot due to two pair of overlapping dots. And no more overlap occurs as the rightward shifting is made.



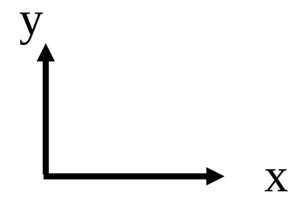
X shifting direction shows a single overlapping between dots.

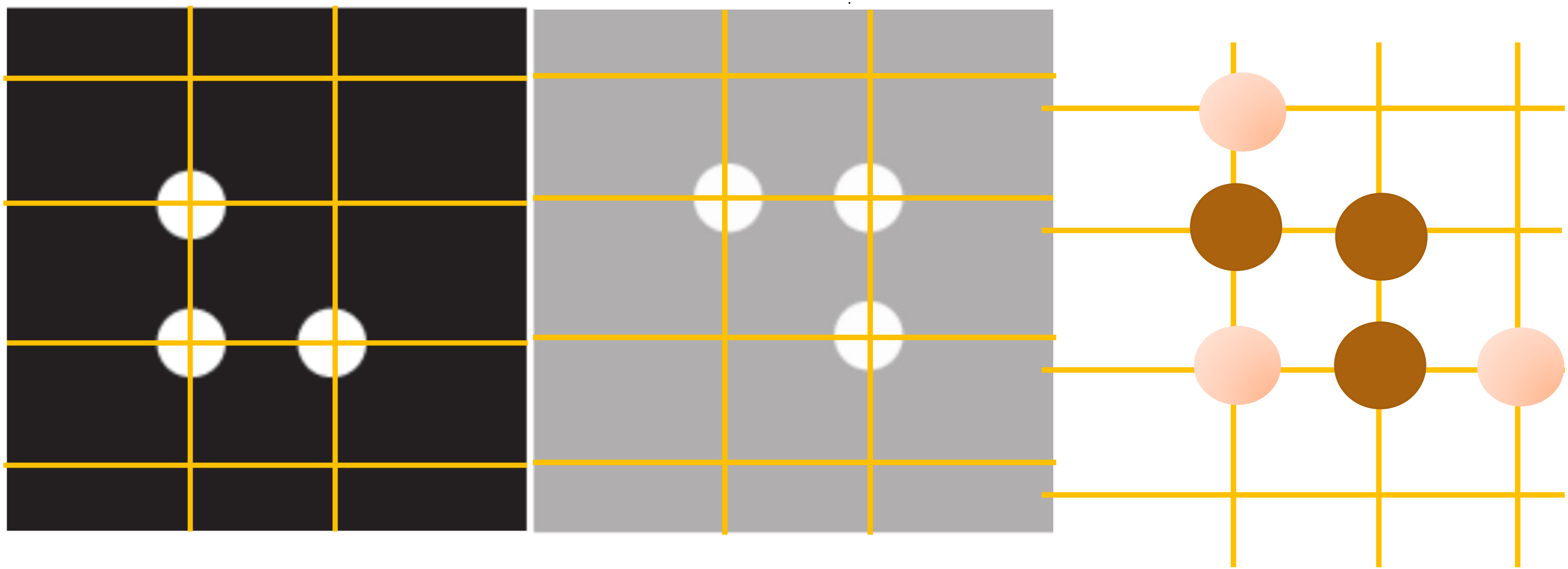


$g(X,Y)$ **2-D autoconvolution**

$f(x,y)$

$h(x,y) \rightarrow h(X\text{-}x, Y\text{-}y)$





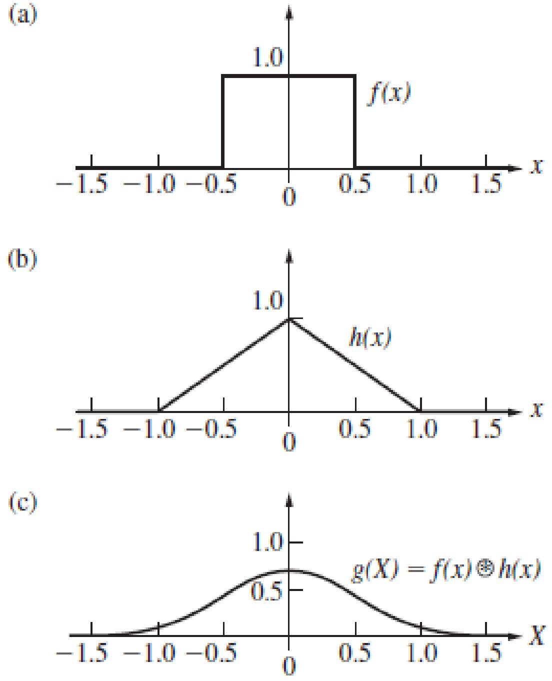
Description of all procedures mentioned here can be found in the "OPTICS" 5th ed by Hecht p. 564.

This is a complete 2-D autoconvolution.

Homework#13

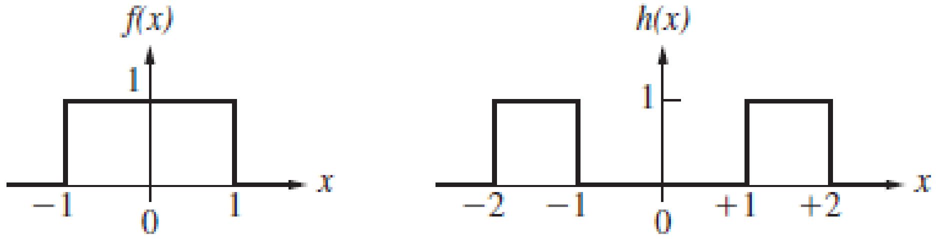
11.20* Examine the three graphs in Fig. P. 11.20 and explain what's being illustrated. Discuss how the shape of $g(X)$ arises. Why is $g(X)$ symmetrical about $X = 0$? What's the significance of the width of $g(x)$? Compute the peak value of $g(x)$.

Figure P.11.20



11.26* Graphically convolve the two functions $f(x)$ and $h(x)$ shown in Fig. P.11.26.

Figure P.11.26



How wide will the convolution be? Will it be symmetrical? Where will it start?

Final exam

Problem 10.19

The convolution of two identical circles of radius r is very important in the modern method of testing lenses against an ideal diffraction limited criterion.

In Figure Q 10.19 show that the area of overlap is

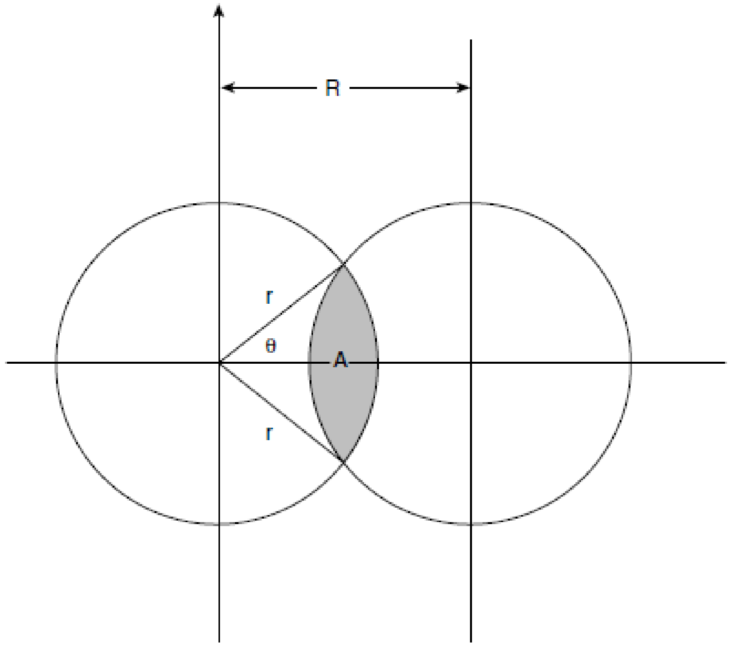


Figure Q.10.19

$$A = r^2(2\theta - 2\sin\theta\cos\theta)$$

and show for

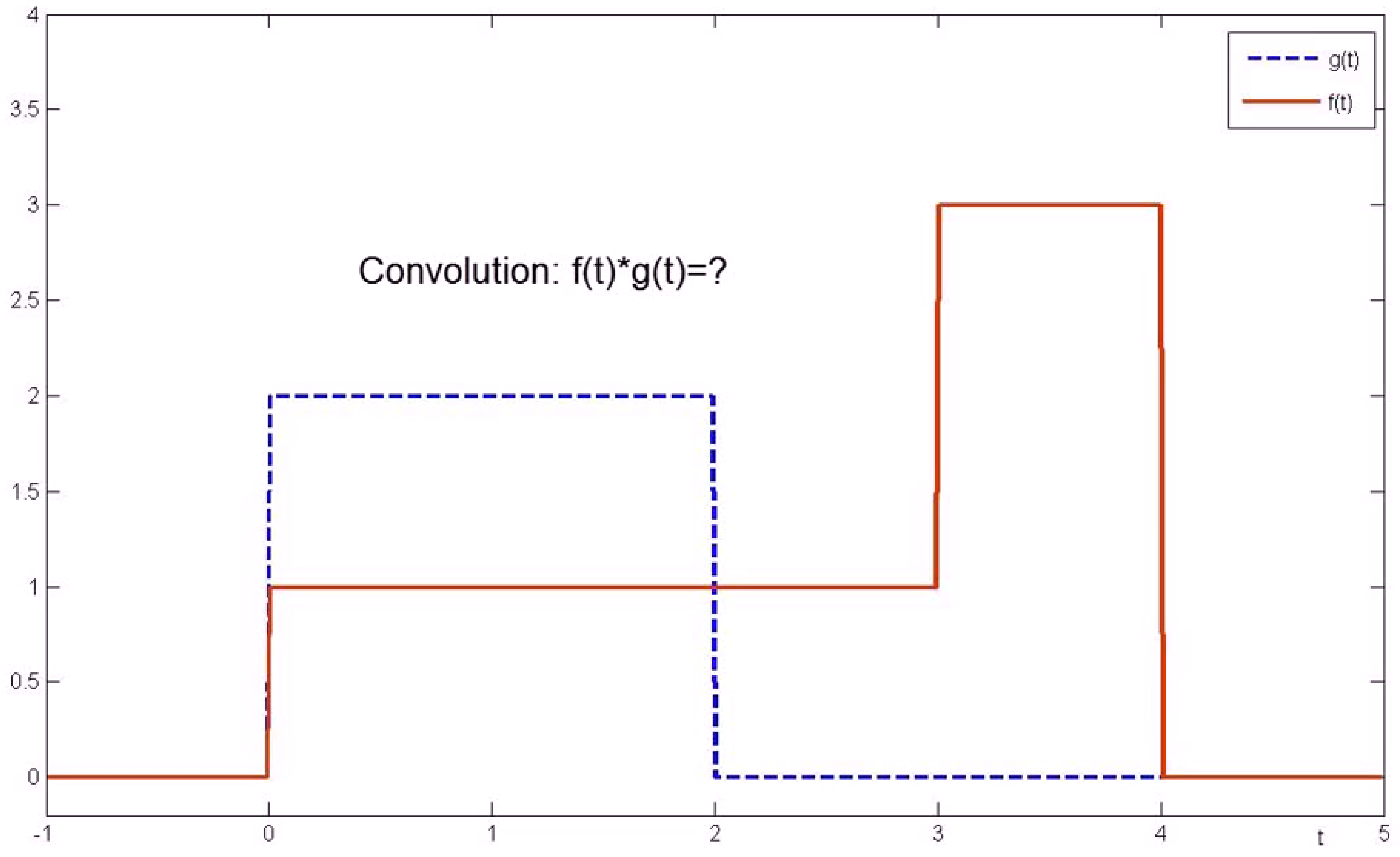
$$R \leq 2r$$

that the convolution

$$O(R) = r^2\left[2\cos^{-1}\frac{R}{2r} - 2\left(1 - \frac{R^2}{4r^2}\right)^{\frac{1}{2}}\frac{R}{2r}\right]$$

Sketch $O(R)$ for $O \leq R \leq 2r$

Apart from a constant the linear operator $\hat{O}$ is known as the modulation factor of the optical transfer function.



g(t)
f(t)
Convolution: f(t)*g(t)=?
4
3.5
3
2.5
2
1.5
1
0.5
0
-1
0
1
2
3
4
t
5