Interference and Diffraction

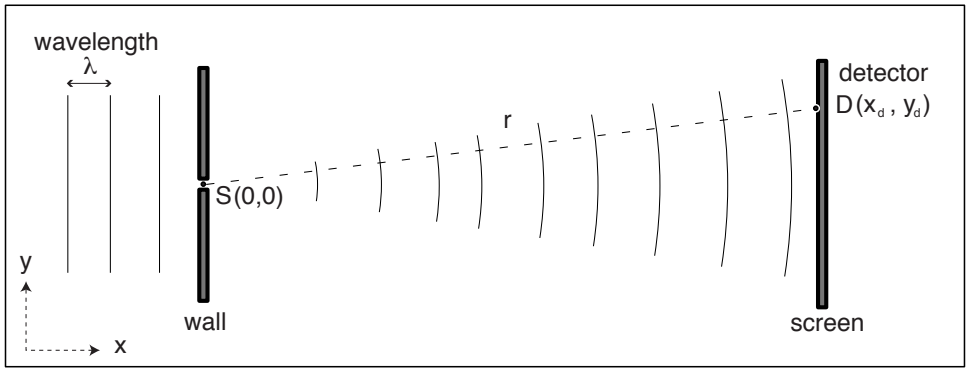
In this session we investigate methods of calculation diffraction patterns from slits using Huygens' Principle.

This set of exercises should take you two weeks to complete.

1.1 Pre-lab

When light travels past an object or through an opening, it spreads out or **diffracts**. You can calculate the pattern of light intensity this produces by using **Huygens' Principle**. This states that every point on a wavefront can be considered as a source of secondary wavelets which spread out in all directions. To find the resultant displacement at any point, you combine all the individual displacements produced by these secondary waves, using the superposition principle and taking into account their amplitudes and relative phases.

Consider a plane wave impinging on an opaque wall which has a tiny hole in it. The hole acts as a source, **S**, of spherical waves, which we can detect when they hit a screen some distance away. Now imagine that the the



electric field variable close to the point S, which we write E_s, is varying sinusoidally with time, with an angular frequency ω and an amplitude E_{max},

$$E_s(t) = E_{max}\cos(\omega t). \tag{1.1}$$

Then the field at the point **D**, a distance r away, will also be varying sinusoidally. The intensity of the light at that point falls off as $1/r^2$, because it spreads out uniformly in all directions. Therefore its amplitude can be written as A/r, where A is some constant. Its phase will depend on how many wavelengths fit into the distance r, i.e.

$$E_d(t) = \frac{A}{r}\cos(\omega t - 2\pi r/\lambda), \qquad \text{where} \quad r = \sqrt{x_d^2 + y_d^2}. \tag{1.2}$$

From arguments given in lectures and in the textbook (see: Young and Freedman, discussion leading to Eq. 35-10) you know that the intensity of the light at **D** is equal to the time average value of E_d^2, multiplied by $\epsilon_0 c$. In other words

$$\text{Intensity at D} \;=\; \epsilon_0 c \,\langle\, E_d(t)^2 \,\rangle \;=\; \frac{\epsilon_0 c}{2}\frac{A^2}{r^2}. \tag{1.3}$$

Note, if you didn't know the amplitude of the oscillation, A, you can (in principle) measure the field at time $t = 0$ and measure it again a quarter of a period later ($t = \tau/4 = \pi/2\omega$). Then you know that,

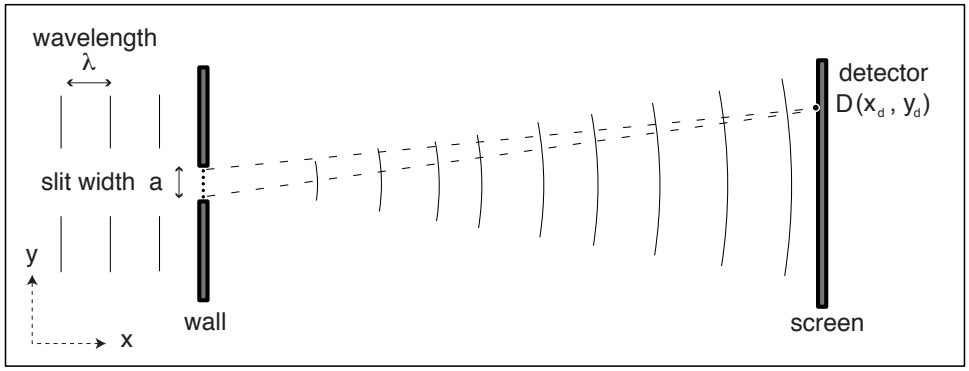
$$E_d(0) = \frac{A}{r}\cos(-2\pi r/\lambda) = \frac{A}{r}\cos(2\pi r/\lambda) \tag{1.4}$$

$$E_d(\tau/4) = \frac{A}{r}\cos(\pi/2 - 2\pi r/\lambda) = \frac{A}{r}\sin(2\pi r/\lambda) \tag{1.5}$$

So if you square and add the two measurements, the fluctuations due to the sinusoidal variation disappear (because $\sin^2 + \cos^2 = 1$), and the intensity at **D** will be given by

$$\text{Intensity} = \frac{\epsilon_0 c}{2}\left(E_d(0)^2 + E_d(\tau/4)^2\right). \tag{1.6}$$

Now consider that the source consists of a number, M, of similar point sources.



If the sources are located at points (x_{s_j}, y_{s_j}), each with the same initial amplitude (A), the total field at the point **D**, E_{tot}, is given by

$$E_{\text{tot}}(0) = \sum_{j=1}^{M}\frac{A}{r_j}\cos(2\pi r_j/\lambda) \quad \text{and} \quad E_{\text{tot}}(\tau/4) = \sum_{j=1}^{M}\frac{A}{r_j}\sin(2\pi r_j/\lambda) \tag{1.7}$$

where:

$$r_j = \sqrt{(x_d - x_{s_j})^2 + (y_d - y_{s_j})^2}\,. \tag{1.8}$$

And the total intensity is got by squaring and adding,

$$I_{\text{tot}}(x_d, y_d) = \frac{\epsilon_0 c}{2}\left(E_{\text{tot}}(0)^2 + E_{\text{tot}}(\tau/4)^2\right). \tag{1.9}$$

It would simplify the arithmetic if we could forget about the constant quantity $\epsilon_0 c$. In all the computations we will do, the amplitude of the wave at the source is arbitrary. We will never calculate absolute values of intensity. So we are at liberty to choose the way intensity is measured so that this constant is effectively equal to 1. In such a measurement system,

$$I_{\text{tot}}(x_d, y_d) = \frac{1}{2}\left(E_{\text{tot}}(0)^2 + E_{\text{tot}}(\tau/4)^2\right). \tag{1.10}$$

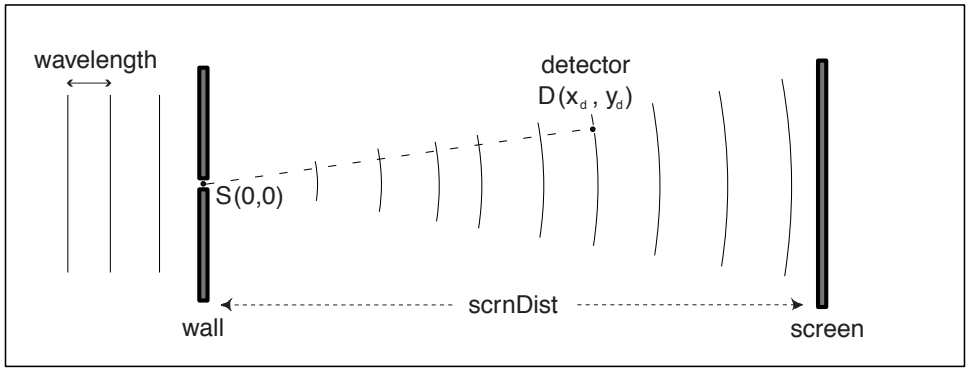
Note, we keep the factor $\frac{1}{2}$ because it has a relevant meaning. It states that the intensity is related to the *time average* of the amplitude squared. (See Young and Freedman, pages 1215–1216.)

As you will see, the intensity pattern produced by light coming through an opening (or openings) is quite complicated in the immediate vicinity of the opening (the **Fresnel regime**) but much simpler if you go far away (the **Fraunhofer regime**). In these exercises you will investigate where the transition between these two regimes is located, and how its location depends on the size of the opening and on the wavelength of the light.

1.2 Lab Exercises

Q1. Visualization of wave fields

We start off by investigating how the field from a point source varies as we move the detector to all points throughout the space between the wall and the screen. Your first task is to write a script `ex1_single_point.m` to calculate the field E at the points **D**, due to a single point source at **S**, using equation 1.2.



1. Start by defining the variables that represent each of the parameters in this scenario. Some suitable variable names and initial values (all lengths are in metres) are shown below:

```
lambda = 4.0e-3;      % wavelength (4.0 mm)
scrnDist = 5.0e-2;    % distance to the screen (50 mm)
scrnWdth = 2.4e-2;    % width of the screen (+/- 12 mm)
xs = 0;               % x-coord of the source
ys = 0;               % y-coord of the source
A = 1;                % amplitude of the source
```

   Notice that we have chosen an unusually large value for the wavelength. In the succeeding exercises we will be dealing with waves in the *optical* range, that is, wavelengths between 400–800 nm. However right now we want to see the oscillatory behaviour of the waves, and so we will pretend we are doing these experiments with microwaves.

2. Next we want to calculate E throughout the space between the source and the screen. Define a vector of x-coordinates that vary between close to 0 and **`scrnDst`** and a vector of y-coordinates that vary between **`-scrnWdth/2`** and **`+scrnWdth/2`**. Each vector should have N values, where N is a suitable number, say five to start with. Use **`meshgrid`** to generate the arrays of detection points in preparation for plotting. Check the size of the resulting arrays make sense.

3. You can change the size of the array of detection points by setting the variable **`N = 500`** (rather than **`5`**) and rerunning the script.

   Investigate the values of the field variable E between the source and the screen by plotting it as a 3D plot using **`mesh()`** with **`xd`** and **`yd`** along the x- and y-axes, and **`E0`** along the z-axis.

   Sketch the resulting plot below.

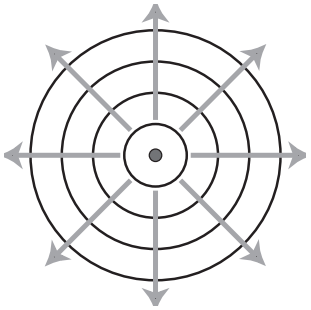
4. Plot E for values of wavelength = 8 mm and wavelength = 2 mm.
   Describe the changes in the figure. Are they what you expected?

5. Another useful way to represent the field is to do a **pseudo-colour contour plot**. We have written a MATLAB function `PseudoColor()` that does this type of plotting. You can download it from the course website and run it:

```
>> PseudoColor(xd,yd,E0);
```

   The code in this function is quite simple. If you open up the function in MATLAB you should be able to understand what is going on. Produce a contour plot for three different wavelengths: 8 mm, 4 mm and 2 mm. **Is everything still as you expected?**

6. In many optics textbooks, the way harmonic waves radiate out from a point source is usually represented by a diagram like this (see for example Young and Freedman, Fig 34-2(a)).
   Make sure that you are happy with the correspondence between the two diagrams. In particular you should be clear that both diagrams represent a picture taken at some particular instant in time, and we have to imagine that the circles on this diagram and the contour plot move outwards as time goes on.

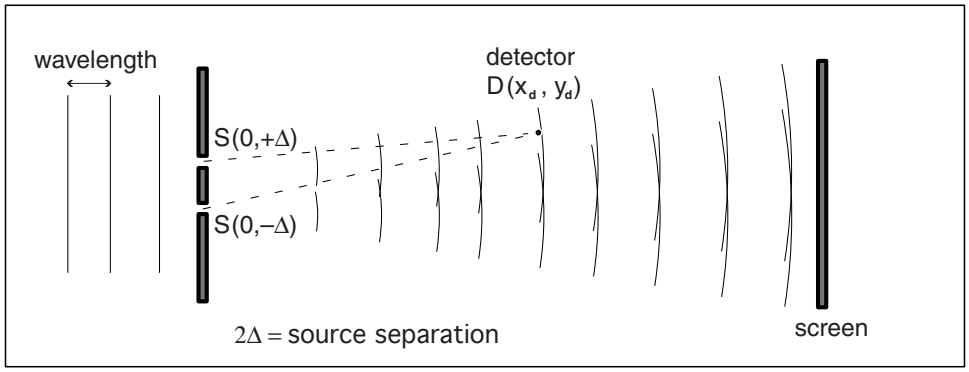


   If these two diagrams do indeed convey the same information, what features on your contour plot do the circles on the conventional diagram correspond to?

Checkpoint 1:

Q2. Point source interference

The next exercise is to calculate the field due to *two* point sources separated by a small distance. As you know, this is a situation which gives rise to interference.



You need to extend the script used in problem 1, to get it to handle two sources instead of one. Make a copy of your script by saving it with a different name `ex2_double_point.m`. You now have a second copy that you can make changes to.

1. Firstly you need to add another parameter to the ones that are set at the top of your script—the separation between the two sources.

```
srcSepn = 1.2e-2;     % separation of sources (1.2 cm)
```

   Write the appropriate statement into the script directly under the other assignments. *Remember that MATLAB is case sensitive when it comes to names of variables.*

2. There are two sources now, so we will make the x- and y-coordinates of the source a (1×2) row vector. The appropriate statements are:

```
xs = [0 , 0];
ys = [-srcSepn/2 , srcSepn/2];
```

   Replace the statements that originally set **`xs`** and **`ys`** to zero with these two statements.

3. Since there are now two sources, we must calculate two arrays of values of r between each source and all the detector points. You do this by replacing the single statement that calculated **`r`** previously with these two statements,

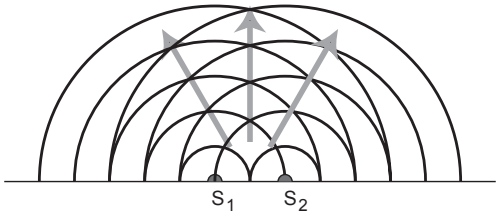
```
r1 = sqrt((xd-xs(1)).^2 + (yd-ys(1)).^2);
r2 = sqrt((xd-xs(2)).^2 + (yd-ys(2)).^2);
```

   Then the field variable at $t = 0$, at any of the detector points is the sum of two terms,

```
E0 = A*cos(2*pi*r1/lambda)./r1 + A*cos(2*pi*r2/lambda)./r2;}
```

   Make these changes in the script, save and run it, just to check that you haven't made any silly mistakes.

4. Examine the results of the calculation by again plotting $E(0)$ as a 3-D plot (**`mesh`**) and as a contour plot (**`PseudoColor`**). Describe briefly in this box the main characteristics of the pattern you observe when you do a contour plot.



Again, in many introductory optics textbooks, interference from a double slit is explained with the aid of a diagram like the following (see for example (See Young and Freedman, Figs 35-2(a) & 8.)).

Convince yourself that your contour plot is essentially the same as this diagram.

5. Now change the wavelength. Try three different wavelengths: 8 mm, 4 mm and 2 mm. Plot the results as a contour plot and describe in this box how the contour plot changes as the wavelength increases.

6. Set the wavelength back tp 4 mm and change the separation between the two sources to 24 mm. Describe in this box how the contour plot changes as the source separation increases.

Checkpoint 2:

Notes:

Q3. Interference of two point sources

This next exercise involves investigating the interference effects we have been observing by measuring the *intensity* of the wave field at points along the screen. That means that we consider that there is a row of detectors attached to the screen as in the diagram on page 1, rather than a rectangular array of detectors filling the area between sources and screen. Then these measurements can be plotted as a 1-dimensional graph of intensity against the y-coordinates of the detectors.

1. Firstly we need to calculate the intensity using equation 1.10. This involves calculating the wave field at time $t = \tau/4$ as in equation 1.7. Let us denote this quantity by **`E4`**.

   Make a copy of your script and save it as `ex3_interference.m`. Then you need to add these two statements to the end of your script.

```
E4 = A*sin(2*pi*r1/lambda)./r1 + A*sin(2*pi*r2/lambda)./r2;
Itot = (E0.^2 + E4.^2)/2;
```

   If you try to observe I_{tot} by plotting it using **`mesh`** or **`PseudoColor`** you won't be able to detect much useful information. The relative intensities are too small to register clearly on the graph.

2. Run the script from Part (1) with these new parameters (optical wavelengths now!),

wavelength =	600 nm	screen distance =	50 mm
screen width =	30 mm	source separation =	0.016 mm

   This time we plot the intensity only at points *on the screen*. The coordinates of the detector points on the screen are **`[xd(:,N),yd(:,N)]`**, and the field intensities at those points are **`Itot(:,N)`**. Remember, if you don't know what **`:,N`** means, ask. And to plot these values of I_{tot} against y on the screen you type, in the command window:

```
>> plot(yd(:,N),Itot(:,N));}
```

   (You might like to add this statement to the end of your script to save typing it over and over.) Draw in the box below the pattern you see on the computer.

3. Before investigating this pattern, save a copy of it for future reference. You will need it for future exercises. Make sure the figure is **active** by clicking inside it. Choose `Save as` from the `File` menu. Instruct it to be saved as a MATLAB *figure*, and give it a suitable filename, e.g. `op1fig1.fig`.

   When you want to recall it, you simply choose `Open` from the `File` menu and select it by name as usual.

4. We are now at the point where the physics of what is happening becomes important. Answer the following questions about this pattern.

   • Why is the total intensity zero at some points along the screen?

• Do the heights of the maximums change as you go away from the centre? Why?

• How does the pattern change if you increase the wavelength of the light? Why?

5. The theory of two-slit diffraction predicts that the maximums and minimums of these interference fringes should be arranged along the y-axis at distances from the centre of the screen given by,

$$y_m = m\,\lambda \times \frac{\texttt{scrnDist}}{\texttt{srcSepn}} \qquad \text{or} \qquad y_m = (m + \frac{1}{2})\,\lambda \times \frac{\texttt{scrnDist}}{\texttt{srcSepn}} \tag{1.11}$$

for a maximum (bright fringe) or a minimum (dark fringe) respectively, where m is an integer. (See Young and Freedman, Eq. 35-4/5.)

Measure the positions of several of these fringes, both bright and dark. Make sure your detector points along the screen are close enough together — `N` should be at least 1000. *Note: you may need to use* `ginput` *to measure these positions.*

Calculate the equivalent value of m for each (using 1.11) and enter them in these boxes. You should be able to achieve no more than 2 significant figures. Right?

m :	**bright fringes**	**dark fringes**

Are these numbers significantly different from integers?

6. Lastly, there is another way of representing the intensity at points along the screen, and that is by using a **grey scale** representation. Download from the module's web site the file which does this, and open it by typing in the command window,

```
GrayScale(yd(:,N),Itot(:,N));
```

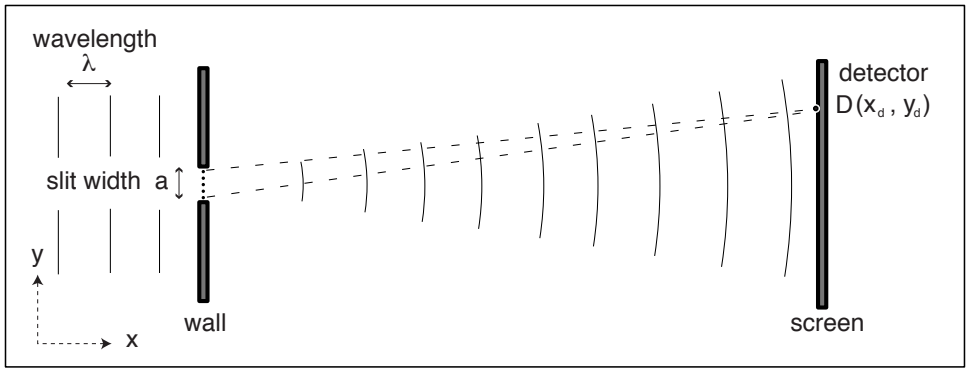
This way of representing the data assumes that the sources in the figures on pages 1 and 13 are not small round holes, but long narrow **slits**, parallel to the z-direction, perpendicular to the plane of the paper. This mimics what you normally find in an experimental laboratory. What you have just drawn on the computer face is very similar to what you would actually see on a physical screen in the laboratory — a series of bright **fringes**, each one being parallel to the z-axis. Confirm that this is what you see.

Checkpoint 3:

Notes:

Q4. Slit of finite width

The next exercise is to calculate the intensity due to a slit of finite width (by which we mean that the width is normally small but non-zero). As you know, this is a situation which gives rise to diffraction of a different character.



The computing task is quite straightforward. The slit through which light passes can be considered as equivalent to a *row* of point sources. The light reaching the screen from the slit is therefore the superposition of the light coming from all of the individual equivalent point sources. In problems OP1.1.2/1.2 we calculated the superposition of light from two sources.

To start, download the script `HuygensConstruction.m` from the course website. This script contains a function which wraps up all of the code we have developed in the last few exercises. By wrapping up the code as a function we have made it easier to run different experiments simply by calling the function.

Have a look through the script to make sure you understand what it does.

1. From now on we want to investigate how *light* diffracts, which means we have to work with wavelengths in the range 400–800 nm. In order to keep the diffraction patterns looking much the same as before we will also need to make the width of the slit correspondingly smaller — of the order of 10^{-2} mm.

   To change our experimental parameters we just need to define the new setup and then run our function. The function can access the variables that we set in the command window because they are declared as **`global`** inside the function. We also need to declare them as **`global`** in the command window:

```
>> global lambda scrnDist scrnWdth slitWdth;
>> lambda   = 600e-9;
>> scrnDist = 50e-3;
>> scrnWdth = 30e-3;
>> slitWdth = 4.0e-6;
>> [yd Itot] = HuygensConstruction;
```

   (Remember all lengths are in metres). If you call the function without the final "`;`" and MATLAB will print out two (5×1) column vectors for **`yd`** and **`Itot`**. You can plot them if you like.

   You will notice inside the **`HuygensConstruction`** funcion that *the numbers of sources and detectors have been set to 7 and 5 respectively*. This is for present convenience. We'll increase them later.

2. Look at the rest of the file, and see how the calculation is done. The coordinates of the sources, **`xs`** and **`ys`**, are calculated as they were before, except that now they are row vectors and not just pairs of numbers. Likewise the coordinates of the detectors, **`xd`** and **`yd`**, are calculated similarly, as column vectors.

   The next step is the tricky one. There are M sources and N detectors, therefore there are $M \times N$ different distances between any one of the sources and any one of the detectors. The program calculates all of these in one $M \times N$ matrix. To do this the program replaces each of **`xd-xs`** and **`yd-ys`** by matrices of the same size, the sources with all their columns the same and the detectors with all their rows the same.

3. Once the $M \times N$ differences have been calculated, it is trivial to work out the $M \times N$ values of r_j (see Eq.(1.8)). Then the program sums over the M different sources to get the final values of $E(0)$ and $E(\tau/4)$ at each of the detectors (note the MATLAB function **`sum`**). Squaring and adding these gives you the intensities at the N detectors.

When you are sure you understand what is going on, change the numbers M and N to something more reasonable, say 51 sources and 201 detectors, and call and run the function. Plot **`Itot`** against **`yd`** and draw the final intensity distribution here.

4. You will need to refer back to this graph later so save a copy of it. Name the new saved file **`op1fig2.fig`**. Again it would be sensible to re-open it immediately to make sure you saved it correctly.

5. At this point, we want to investigate a lot more about this pattern, and the way it depends on the various physical parameters. With **`HuygensConstruction`** this is easy enough to do, but it will get tedious if you have to type a lot of stuff each time you want to change one of the numbers. What you need is a standard **front end** which will streamline the business of inputting data.

   One of the strengths of MATLAB is that it allows the writing of GUI (**Graphical User Interface**) objects. On the course web site there is a file called `op1driver.m`. Download this file and call and run it as usual,

```
>> op1driver;
```

   and you will see that it opens a panel in which five physical variables are displayed and can be changed.

   It also contains a button labelled **`Recalculate`**. When it is clicked the driver passes temporary control to **`HuygensConstruction`**, which takes the values which have been entered and recalculates and plots the new intensity distribution. Click it and you will see that it opens a window in which the intensity is plotted, as a simple graph and as a grayscale representation (both colour-coded to match the wavelength).

   You will use **`op1driver`** for the rest of this set of exercises.

   Also, it will be assumed that the numbers of sources and detectors remain at the values specified in part (3), namely $M = 51$ and $N = 201$.

Checkpoint 4:

Notes:

Q5. Fresnel and Fraunhofer diffraction (single slit)

The next exercise is to investigate, in considerable detail, the diffraction pattern from a single slit—that is, the situation for which you have just prepared a function to do the calculation in Exercise 1.2. Start off with these physical parameters:

wavelength	=	600 nm	screen distance	=	50 mm
slit width	=	0.005 mm	screen width	=	24 mm

Don't forget. We are not using **`slitSepn`** *yet. Make sure it is set to zero.*

1. Firstly observe the details of the intensity pattern resulting from this data. You should recognize immediately that it matches the diagram shown in the textbook. (See Young and Freedman, Fig. 36-9.). Furthermore the discussion in the textbook outlines a method for deriving approximately some features of this pattern, and contains explicit formulas for calculating the following quantities.

   - The distance from the centre of the screen to the mth intensity minimum, Eq. 36-8,

     $$y_m = x(m\lambda/a)$$

     where x is the distance source–screen, a is the slit width and $m \geq 1$.

   - The relative magnitude of the intensity of the mth side maximum, Eq. 36-10.

     $$I_m \approx I_0 \left((m + \frac{1}{2})^2 \pi^2\right)^{-1}$$

   Measure these two quantities for $m = 1$ from the pattern on your computer screen (use the MATLAB commands **`zoom`** and **`grid`** to read numbers accurately). Then compare them with the theoretical values calculated from the formulas in the textbook.

   `yd`(first min): **measured values** **theoretical values**

   $\frac{\texttt{Itot(first max)}}{\texttt{Itot(centre)}}$: **measured values** **theoretical values**

2. Firstly you are asked to observe how the intensity pattern changes with distance from the screen. Vary the distance to the screen over a wide range, from very close to the slit (of order of the width of the slit) to, say, tens of cm.

   Caution: As you change the screen distance, the width of the diffraction pattern will change. You should vary the screen width as you go so that you see all of the significant parts of the pattern in sufficient detail. Also, if the graph cuts off the top of your plot when you make some change, close the graph and recalculate.

   You should notice quite different behaviour when you are near to the slit, from when you are far away. Describe in words the properties of the "far" pattern (the **Fraunhofer regime**), paying particular attention to how/if its shape changes with distance, and how/if its size changes with distance.

   Similarly, describe in words the properties of the "near" pattern (the **Fresnel regime**), paying particular attention to how/if its shape changes with distance, and how/if its size changes with distance.

3. It is not difficult to find a screen distance at which the pattern is unambiguously Fresnel, and one at which it is unambiguously Fraunhofer. What is not so easy to determine is the dividing line between the two. As you decrease the screen distance from large values, the first minimum of the Fraunhofer pattern seems to rise above the x-axis. Many workers (arbitrarily) choose this fact to signal the onset of Fresnel diffraction.

   A useful criterion would be to find the point at which the height of the first minimum $\approx 1\%$ of the height of the central maximum. Using this criterion, measure the screen distance at which, approximately, the "near" region stops and the "far" region begins.

   Hint: You should start at the default distance to the screen, and systematically reduce it until you see the shape of the diffraction pattern change dramatically. However, as you get closer to the source you will find you need to decrease the width of the screen, so that you can see details of the diffraction pattern. It would be most convenient to start decreasing both **`scrnDist`** *and* **`scrnWdth`** *by a factor of 10 each time.*

4. **Advanced questions:** Choose two different values of the slit width, 0.0025 mm and 0.010 mm and estimate the transition point for both of these, as you did in part (3).

   for slit width 0.0025 mm **for slit width 0.010 mm**

5. Set the slit width back to its default value, and choose two different values of the wavelength, 400 nm and 500 nm. Repeat the observations you made in part (4). Estimate the transition point for both of these.

   for wavelength 400 nm **for wavelength 500 nm**

6. Even though the definition of where the transition occurs is a subjective judgement, it can be seen to depend in a predictable way on the physical parameters wavelength and slit width. In parts (3) and (4) you have three "measurements" with the same wavelength but different slit widths. In parts (3) and (5) you have three "measurements" with the same slit width but different wavelengths. Use these to find an empirical formula from which you might calculate where the transition point is, in the form

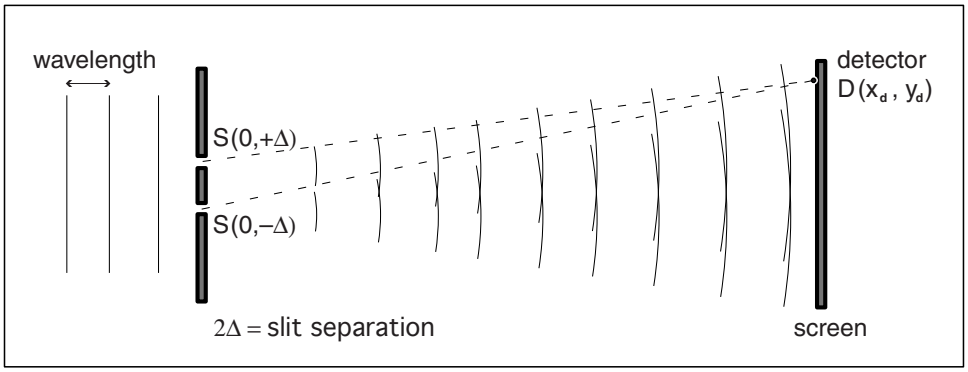
$$distance = constant \times (slitwidth)^a \times (wavelength)^b. \tag{1.12}$$

Checkpoint 5:

Notes:

Q6. Fresnel and Fraunhofer diffraction (double slit)

The next exercise is to calculate the intensity due to two slits of finite width, separated by a very small distance.



Start off with these values for the physical parameters,

wavelength =	600 nm	screen distance =	50 mm	slit width =	0.005 mm
		screen width =	24 mm	slit separation =	0.010 mm

Note: the "slit separation" is taken to be the distance between the centres of the two slits.

1. The first job will be to change the function **`HuygensConstruction`** so that it will handle two slits. We don't plan to make the function flexible enough to handle an arbitrary number of slits, so we take the simplest and crudest way of making the change we want.

   Change the calculation of the **`xs`** and **`ys`**, making them $2M \times 1$ row vectors.

```
xs  = zeros(1,2*M);
ys1 = [-slitWdth/2: slitWdth/(M-1): slitWdth/2] - slitSepn/2;
ys2 = [-slitWdth/2: slitWdth/(M-1): slitWdth/2] + slitSepn/2;
ys  = [ys1, ys2];
```

   Then change **`M`** to **`2*M`** whenever it occurs subsequently. Note that, if you set **`slitSepn`** equal to zero, you will have effectively one slit, and you can do all the exercises in Exercise 1.1.2.

2. When you have made the changes and saved the new version, re-run **`op1driver`**. If you haven't made any errors, the new calculation should automatically appear.

   Do some preliminary observations of your own on the diffraction pattern you get in this situation. Keep the slit width and separation the same and vary the distance to the screen (and the screen width as necessary). By this exercise, convince yourself that you still can distinguish a Fraunhofer regime and a Fresnel regime.

3. Draw the shape of the diffraction pattern you observe in the Fraunhofer (far) regime.

   We will investigate this shape in greater detail in the next exercise. For now, verify for yourself that it behaves as the single slit Fraunhofer pattern did—that is, its size changes with distance from the screen, but its shape stays the same.

4. Similarly, observe the diffraction pattern in the Fresnel (near) regime and verify for yourself that it behaves as the single slit Fresnel pattern did—that is, its size stays more or less the same but its shape changes a great deal with distance from the screen.

   There is however one feature of the shape of the Fresnel pattern which is fairly constant. As an illustration, draw here the intensity distribution at a distance of 0.05 mm (use a screen width of 0.05 mm).

   You should find two strong maximums, at `yd` $\approx \pm 0.005$ mm, and these should persist as you change the screen distance. Why do these occur?

5. **Advanced question:** What has happened to the near/far transition point?

Checkpoint 6:

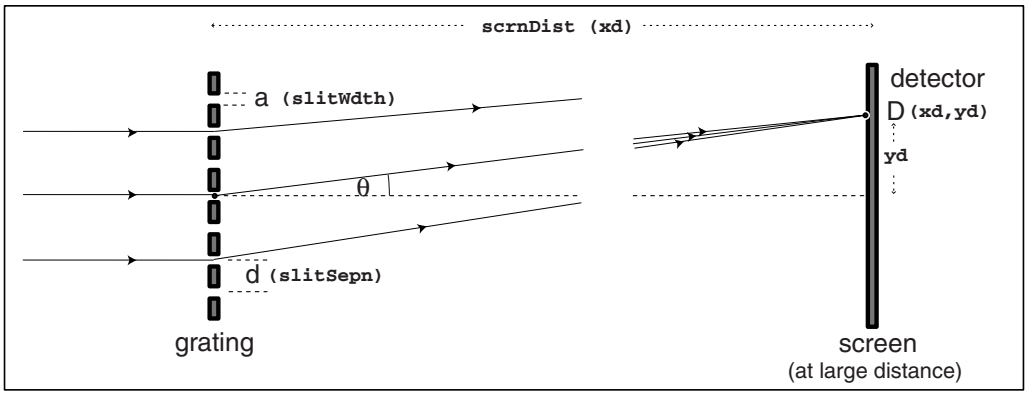
Diffraction Gratings

In this set you will investigate Fraunhofer patterns produced by diffraction gratings — that is, by a regular array of many slits or reflecting strips. The physics involved in these exercises is covered in Young and Freedman, §36–4,5.

This set of exercises should take you two weeks to complete.

3.1 Pre-lab

A regular array of slits, or grooves etched into a surface, is known as a **transmission grating**. Previously you investigated the Fraunhofer diffraction patterns produced by light travelling through two slits. If we do the same calculation with more slits, the maximums of the diffraction pattern remain in the same positions, but become progressively narrower. With hundreds of slits, these maximums become so narrow that we can measure their angular position with very high precision. We already know that, provided we are in the Fraunhofer regime, the position of these maximums depend linearly (approximately) on the wavelength of the light used. Hence a grating can be used as a **spectrometer**, an extremely sensitive device for measuring wavelengths and spectral intensities.



The grating equation

Consider a plane electromagnetic wave of wavelength λ incident on a grating consisting of N slits, as shown. We imagine that the screen is so far away that the rays converging at D are essentially parallel, all making the same angle θ with the normal to the grating. Then theoretical analysis predicts that the intensity of light hitting the screen, as a function of θ, is given by

$$I_N(\theta) \;=\; I_N(0)\,\left(\frac{\sin(\beta/2)}{\beta/2}\right)^2\,\left(\frac{\sin(N\phi/2)}{N\sin(\phi/2)}\right)^2 \tag{3.1}$$

where

$$\tan\theta \;=\; \frac{\texttt{yd}}{\texttt{scrnDist}}\;, \qquad \beta \;=\; \frac{2\pi a}{\lambda}\sin\theta \quad\text{and}\quad \phi \;=\; \frac{2\pi d}{\lambda}\sin\theta. \tag{3.2}$$

You may care to show that, in the special case $N = 2$, these equations match exactly the corresponding equation in the textbook, Young and Freedman, Eq.(36-12).

The term involving β varies slowly with angle, and represents the diffraction envelope pattern for a single slit. The term involving ϕ varies more rapidly, especially when N is large. The successive sharp peaks to either side of the centre are referred to as the first, second, third, *etc.*, **orders** of diffraction. These peaks are quite narrow and each is separated from the next order by smaller subsidiary peaks.

It can be shown from this expression that the maximum intensities in the diffraction pattern, the bright fringes, occur at angles θ given by

$$d \sin\theta = m\lambda \tag{3.3}$$

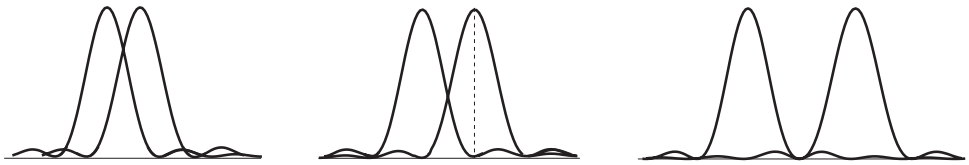
where m is the order of the fringe. This is known as the **grating equation** (see Young and Freedman, Eq.(36-13)). Note especially that we have always assumed *normal incidence*, that is that the incident light is perpendicular to the plane of the grating. If not, Equation 3.3 will look a bit different.

Resolving power

In a spectroscopic laboratory, the usefulness of a grating is specified by its **resolving power**. This is a measure of the minimum wavelength difference, $\Delta\lambda$, that can be distinguished by the grating. Its actual definition is the ratio of the average of two just distinguishable wavelengths divided by their difference. See Young and Freedman, Eq.(36-14).

$$R = \lambda/\Delta\lambda. \tag{3.4}$$

Whether or not two wavelengths can be distinguished often involves a semi-subjective judgement, usually made on the following criterion. So long as the intensity peak for one of the wavelengths lies outside the first minimum of the other, we agree to say that we can tell them apart. The two wavelengths are said to be **resolved**. If the two intensity peaks are closer together than this, then we agree to say that we cannot tell them apart. The two wavelengths are **unresolved**.



This recipe for making the decision is known as the **Rayleigh criterion**.

The term involving ϕ in Eq.(3.1) determines how close together are the neighbouring maximums and minimums. You can use this, and Eq.(3.3), to find what change in wavelength ($\Delta\lambda$) will shift the intensity peak by this amount. Whence you can get this expression for the resolving power:

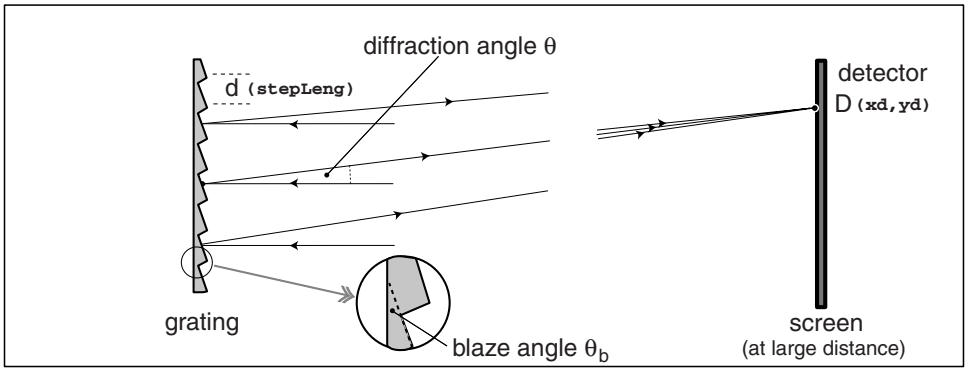
$$R = \lambda/\Delta\lambda = mN \tag{3.5}$$

where m is the (integer) order of the fringe in question. See Young and Freedman, Eq.(36-15).)

Useful gratings have large values of N. Usually this is specified in terms of the number of grooves per millimetre and the grating width. Typical densities of grooves are 150 per mm up to 1200 per mm, and typical gratings are 50 mm or 100 mm wide.

Advanced topic

There is another kind of grating, consisting of a regular array of very thin reflecting grooves — a **reflection grating**. Clearly such arrays will produce the same kind of diffraction pattern as a transmission grating, if the reflected light is allowed to fall on a screen.



However they offer one special advantage. In a transmission grating, most of the light goes straight through without being diffracted sideways, and ends up in a bright band at the centre of the screen (called the **zeroth order fringe**). Very little energy reaches the screen where the diffracted fringes are found.

But in some reflection gratings the reflecting grooves are cut at an angle to the plane of the grating. This means that most of the incident light (assumed perpendicular to the plane of the grating, remember) does not reflect straight back. Instead, following the laws of specular (mirror) reflection — angle of reflection equals angle of incidence — it goes at an angle different from the zeroth order diffraction angle. Such gratings are said to be **blazed**.

The trick is to construct the grating in such a way that the direction in which most of the light goes corresponds to the direction in which one of the fringes occurs (often the first order fringe). If we restrict ourselves to the case where the light is incident normally to the plane of the grating, specular reflection occurs at an angle $\theta = 2\theta_b$, where θ_b is the **blaze angle** — the angle of the reflecting grooves to the plane of the grating. Putting this angle into the grating equation (3.3) gives the **blaze wavelength**, λ_b, for which the peak intensity will occur,

$$m\,\lambda_b \;=\; d\,\sin 2\theta_b \qquad (\text{ most often with } m = 1\,.) \tag{3.6}$$

3.2 Lab Exercises

The first two exercises of the set employ the same Huygens' Principle calculation we used in Set 1, to study diffraction from a small number of slits. Our aim is to establish that the Fraunhofer patterns we observe are identical with those described by Eq.(3.1). Once this has been established, you will use this equation in the remaining exercises to study the properties of diffraction gratings.

To help with this, several pre-written MATLAB scripts are on the course web site:

- `HuygensConstruction2` — a slightly altered version of the function used in Set 1, which can handle different numbers of slits, within a small range.
- `op2driverA/B/C` — three GUI Data Input objects with facilities for inputting and changing various variables, to be used with different exercises in this set;
- Three functions which will calculate the Fraunhofer pattern under different circumstances —

  `ManySlitCalculation`: for a small number of slits;

  `GratingCalculation`: for a grating of up to 50 slits; and

  `BlazingCalculation`: for a blazed reflection grating.

Before starting these exercises, download these files to your local directory.

Q1. Double slit diffraction

Start from where you were at the end of the two slit experiment from the previous lab, examining the Fraunhofer pattern from *two slits of finite (that is, not infinitesimally small) width*. You are asked to keep the screen distance constant (at the default value of 50 mm) and to observe what happens to the pattern as you change the width and separation of the slits and the wavelength of the light.

1. Download and run **`op2driverA`** and study the influence of the width of the slits on the diffraction pattern.

   Keep the slit separation fixed at 0.010 mm and the wavelength at 600 nm. Change the slit width successively to these values: 0.002 mm, 0.005 mm, and 0.008 mm.

   Draw the resultant intensity patterns in these boxes. Make sure your drawing shows clearly which features of the pattern change and which stay the same.

   Slit width

   0.002 mm

   0.005 mm

   0.008 mm

2. Study the influence of the separation of the slits on the diffraction pattern. Keep the slit width fixed at 0.005 mm, and the wavelength at 600 nm. Change the slit separation to these values: 0.010 mm, 0.030 mm and 0.050 mm.

   Note: the graph is no longer smooth, so you will need to use more points in the calculation. Inside **`HuygensConstruction2`** *change the number of detectors* **`N`** *from 151 to 701.*

   Draw the resultant intensity patterns in these boxes. Again, make sure your drawing shows clearly which features of the pattern change and which stay the same.

   Slit separation

0.010 mm	
0.030 mm	
0.050 mm	

3. Study the influence of the wavelength on the diffraction pattern. Keep the slit width fixed at 0.005 mm, and the slit separation fixed at 0.010 mm. Change the wavelength of the light to these values: 400 nm, 550 nm and 750 nm.

   Draw the resultant intensity patterns in these boxes. Again, make sure your drawing shows clearly which features of the pattern change and which stay the same.

   Wavelength

400 nm	
550 nm	
750 nm	

4. Lastly we would like to compare this with the intensity patterns that you calculated previously for two *point* sources, and for a *single* slit.

   Set up the graph of the intensity pattern for the parameters:

   wavelength = 600 nm
   screen width = 30 mm screen distance = 50 mm
   slit width = 0.004 mm slit separation = 0.016 mm

   In the previous lab you saved a copy of the diffraction pattern for *two point sources* a distance 0.016 mm apart, with the other parameters the same as you are using now. This file was named **`op1fig1.fig`**. Reopen this figure using **`Open`** from the **`File`** menu.

   Likewise, you saved a copy of the diffraction pattern for a single finite slit, of width 0.004 mm, also with the other parameters the same as you are using now — under the name **`op1fig2.fig`**. Reopen this figure, without closing the other two graphs. Arrange all three graphs on the screen so that the axes are all the same size. Study these graphs and answer this question: What is the relationship between the diffraction pattern of two finite slits, and the other two patterns?

   Convince yourself, without necessarily computing anything new, that the total intensity distribution for *two* slits of *finite width*, is

$$I_2(\theta) \;=\; I(0) \times \left(\frac{\sin(\beta/2)}{\beta/2}\right)^2 \times \cos^2(\phi/2) \tag{3.7}$$

Checkpoint 1:

Notes:

Q2. Multiple slit diffraction

At this point, if you haven't done so already, open the m-file **`HuygensConstruction2`** and observe what changes have been made in order to allow the number of slits to be changed. Make sure you understand how the calculation is done, and why it would be impractical to use this script for a very large number of slits.

1. Re-run **`op2driverA`** with the same parameters you used in Exercise OP2.3.2(4). Observe what happens to the diffraction pattern as you change **`numSlits`** through 1, 2, ... 5.

   Draw the resultant interference patterns for $N = 2, 4, 6$ in these boxes. Make sure your drawings show clearly which features of the patterns change, and which stay the same.

No of slits	
2	
4	
6	

2. Rerun the same computations, changing **`numSlits`** through, 1, 2, ... 5. Measure and record in the following table, to three significant figures only, the intensities at the middle of the screen. Calculate the *ratio* of the central intensity with N slits and the central intensity with a single slit (to two significant figures) and enter the values here.

N	1	2	3	4	5
$I_N(0)$					
$I_N(0)/I_1(0)$					

3. When we calculated these intensities, we modelled the light coming through each slit as coming from a fixed number, M, of discrete point sources making up each slit. In terms of this model, it should be obvious why the central intensity increases as the number of slits increases. The same amount of energy flows through each slit, so with N slits, the total energy flux should be N times that for a single slit. But this is *not* what your results show (or at least, should not be!) Why do the ratios $I_N(0)/I_1(0)$ have the values you found?

Checkpoint 2:

Q3. Mathematical formula for multiple slit diffraction

The point has already been made that it would be impractical to use the current script for a system with a large number of slits. In future we will use a MatLab function based on Eqn.3.1, rather than one based on Huygens' Principle directly. But first we need to make sure the two ways of doing the calculation give the same answers for a small number of slits.

1. Firstly, you need a MATLAB function which will calculate the right hand side of Eq.(3.1), for any positions along the screen. The m-file **`ManySlitCalculation`** contains, in skeleton form, such a function. The steps in the calculation are indicated, but initially each of the vectors you have to calculate are set equal to 1. The real statements are left for you to fill in, as follows.

   Firstly, write the three statements, using Eqs.(3.2), that calculate θ, β and ϕ for a vector **`yd`** which has been passed to the function. Note: you will need to use the MATLAB function **`atan`**.

   Next write the statement that calculates $(\sin(\beta/2)/(\beta/2))^2$, denoted by the name **`slitPatn`**. Do this in the most direct way, by adding the statement

```
slitPatn  = (sin(beta/2)./(beta/2)).^2;
```

   Test that this works properly, by setting up a vector in the command window for **`yd`** (redefining **`scrnwdth`** if necessary) and calling the function in its unfinished form. Then plot the result and you should see the usual finite slit diffraction pattern:

```
>> yd   = [-scrnWdth/2 : scrnWdth/300 : +scrnWdth/2]';
>> Irel = ManySlitCalculation(yd);
>> plot(yd,Irel);
```

   There is a slight problem here. Although the output looks OK, you get an error message, because the θ in the denominator vanishes at the centre of the screen. There are several ways you can get round this, since you know that the value of $\sin\theta/\theta$ is 1 when $\theta = 0$. However the simplest kludge is to make sure you never divide *quite* by zero.

   In MATLAB the smallest number you can add to 1.0, and make a difference, is called **`eps`**. It is a number of order 10^{-16}. Add this to your denominator, replacing the statement above by,

```
slitPatn  = (sin(beta/2+eps)./(beta/2+eps) ).^2;
```

   Run it again. You should get the same answer, but this time without the error message.

2. When you come to calculate the term involving ϕ in Eq.(3.1), which describes the contribution of the many sources, you have to use the same trick. Add this statement to the script,

```
srcePatn = (sin(numSlits*(phi/2+eps))./(numSlits*sin(phi/2+eps)) ).^2;
```

   Test that this part works properly as previously, by adding a temporary statement **`slitPatn = 1`** to override the calculation of **`slitPatn`** above. When you plot the result you should see the diffraction pattern due to a number of point sources. Draw in this box the pattern you see when you do this test with the number of slits equal to 4.

3. You are now ready to calculate the complete function, and can verify that the diffraction pattern from a number of finite slits is accurately described by Eq.(3.1), as follows:

   Take out the statement you added in OP3.3.2(2), which sets **`slitPatn`** equal to 1, so that it once again calculates **`slitPatn`** using β. Now compute the diffraction pattern again, using the Huygens Principle method of calculation, without using the driver (because you want to do your own plotting). Plot the result so that you can draw another graph on top of it.

```
>> [yd Itot] = HuygensConstruction2;
>> plot(yd, Itot);
>> hold on;
```

   Then calculate the complete theoretical diffraction pattern, given by Eq.(3.1), using the relative intensity distribution computed by the function **`ManySlitCalculation`** (with the correct slit width and spacing), *multiplied by* the central intensity that you recorded in exercise OP3.3.2(2).

   Plot this pattern on top of the graph you just computed.

```
>> Irel = ManySlitCalculation(yd);
>> plot(yd, Irel*..., 'red');
```

   (You fill in the multiplier.)

   Repeat with the number of slits equal to 2, 3, . . . 6. Estimate the discrepancy in the heights of the first order fringes in all cases.

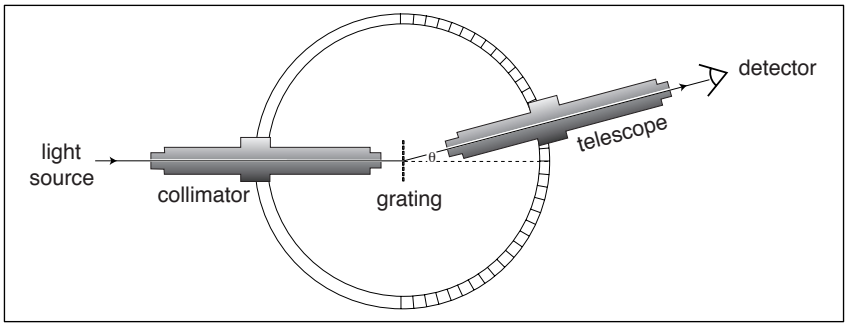
   Discrepancy:

Checkpoint 3:

Notes:

Q4. Transmission diffraction grating

We now consider a grating when used as a simple spectrometer. Schematically the experimental arrangement is something like this. (Compare with Young and Freedman, Fig.36-20.)



You will notice that the "detectors" are arranged around a semicircle, with the diffraction angle θ varying between $\pm\pi/2$. Although real diffraction gratings might consist of $\sim 10^5$ slits, we will confine ourselves to numbers less than 100. We will not attempt to use Huygens' Principle to calculate the intensity pattern at the detectors. Instead we will use Eq.(3.1).

1. Open the m-file **`GratingCalculation`** and examine its contents. It should match exactly the statements in **`ManySlitCalculation`** with these small changes:

   - The function requires the vector **`theta`** as input rather than the distances **`yd`** we have used till now. The variables **`scrnDist`** and **`scrnWdth`** are no longer needed.
   - The wavelength **`lambda`** is passed to the function, rather than being a global. You'll see why shortly.
   - The intensity at the middle of the screen is chosen so that, while arbitrary, it reflects the dependence on the number of slits in the grating and the width of each.

   Now run the new script **`op2driverB`**. The GUI input panel it opens up is similar to the previous one, except that the variables **`scrnDist`** and **`scrnWdth`** are no longer present, and there is facility to input *two* different wavelengths. Currently both these are set at the same figure. Check that everything runs without errors.

2. Start off with these parameters:

   wavelength1 = 600 nm slit width = 0.0003 mm number of slits = 20
   wavelength2 = 600 nm slit separation = 0.0015 mm

   Observe the pattern as you change the number of slits to 10, 20, 30, ... 100. Draw the resultant intensity patterns for values of 10, 50 and 100 in these boxes. Make sure your drawing shows clearly which features of the pattern change and which stay the same.

   No of slits

   10

   50

   100

3. With the same default values, and 50 slits, change the slit separation to these values: 0.001 mm, 0.003 mm and 0.005 mm. Draw the resultant interference patterns in these boxes. Again, make sure your drawing shows clearly which features of the pattern change and which stay the same.

Slit separation

0.001 mm	
0.003 mm	
0.005 mm	

4. With the same default values, and 50 slits, change the slit width to these values: 0.0002 mm, 0.0006 mm and 0.0010 mm. Draw the resultant interference patterns in these boxes. As usual, make sure your drawing shows clearly which features of the pattern change and which stay the same.

Slit width

0.0002 mm	
0.0006 mm	
0.0010 mm	

5. Lastly, with the same default values, and 50 slits, change both wavelengths together to these values: 400 nm, 550 nm and 750 nm. Draw the resultant interference patterns in these boxes. Again, make sure your drawing shows clearly which features of the pattern change and which stay the same.

Wavelength

400 nm

550 nm

750 nm

Checkpoint 4:

Q5. Diffraction grating spectrometer

Thus far we have explored diffraction patterns for a single wavelength.

1. *Test problem:* A grating was used in the third order with light from a helium discharge tube. The red line at 667.8 nm appeared at an angle 40.0°.

   Run **`op2driverB`** using the following starting parameters:

   wavelength1 = 667.8 nm slit width = 0.0002 mm number of slits = 40
   wavelength2 = 667.8 nm slit separation = 0.0030 mm

   and, by adjusting the slit separation in small steps, find how many slits per mm the grating must have had.

   Compare your answer (to four significant figures) with the theoretical value that you would get by using Eq.(3.3).

	measured value	**theoretical value**
Number of slits per mm:		

One of the most common uses of a grating spectrometer, however, is to investigate the light emitted by certain atomic or molecular species, which we know consists of several discrete, very pure frequencies — that is, **line spectrums**. The next two exercises involve exploring the (Fraunhofer) diffraction pattern when there are *two* frequencies present.

2. Run **`op2driverB`** using the following parameters:

   wavelength1 = 700 nm slit width = 0.0002 mm number of slits = 40
   wavelength2 = 500 nm slit separation = 0.0030 mm

   Observe the diffraction patterns you see. Identify the first, second, third, *etc.* order fringes associated with each wavelength.

   You should be able to appreciate a possible source of difficulty with diffraction gratings. The first order green fringe occurs at a smaller angle than the first order red fringe. Similarly the second order green comes before the second order red, and the third order green before the third order red. But the *fourth* order green also comes before the *third* order red. The two diffraction patterns are said to **overlap**. With a spectrum consisting of many wavelengths, this kind of overlapping can lead to confusion.

   Note: The word "overlap" here means that two whole spectrums sit on top of one another. Do not confuse this meaning with how it is used later in Exercise OP3.3.2 when you deal with two lines in one spectrum overlapping one another.

3. *Test problem:* By keeping the first wavelength fixed at 700 nm, and varying the second wavelength, find the largest difference in wavelength (to three significant figures) for which there is no overlap in the first, second and third orders. (*This means the third order lines must be just on the point of overlapping.*)

   Wavelength difference:

Checkpoint 5:

Notes:

Q6. Resolving power

1. We now wish to investigate the resolving power of a grating. We first do a preliminary exploration of which parameters determine the ability of a grating to distinguish two close frequencies. Start off with these values:

wavelength1	=	550 nm
wavelength2	=	575 nm
slit width	=	0.0002 mm
slit separation	=	0.0020 mm
number of slits	=	20

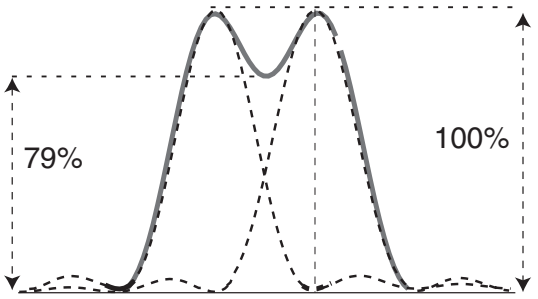
   If you look at the second or third order fringes, it should be perfectly obvious that there are two discrete frequencies present, even though they have very similar colours. However from the first order fringe it is not so obvious. If you zoom, you will see that the total intensity has a slight dip in the middle, but from the simulated spectrum it is hard to tell whether there are two narrow lines present, or a single wide line. The lines are **unresolved in first order**. However, if you change the number of slits to 40, the two lines are clearly **resolved in all orders**. As suggested by Eq.(3.5) the resolving power of the grating is increased by increasing the number of slits.

2. The **Rayleigh criterion** for judging when the two lines are just on the borderline of being resolved is that **the maximum of one line occurs at the same angle as the first minimum of the other**. Refer back to the discussion on page 2 of these notes.

   Change the number of slits to 23 and verify that the Rayleigh criterion is now just satisfied. *(Note: you need to be careful when zooming, so that the maximum does not disappear out the top of the window.)*

   In most real world situations, the Rayleigh criterion cannot be invoked in this form. Since all you can measure is the total intensity from both lines together, you can't say where the first minimum of either one occurs.

   For one dimensional sources, such as we are dealing with here, it can be shown theoretically that when the maximum of one line occurs at the same angle as the first minimum of the other (that is, when the Rayleigh criterion is just satsified) the total intensity halfway between the two is 79% of the maximum intensity of either(to two significant figures).

   Therefore for all one dimensional sources the criterion is taken to be *the two lines are considered just resolved when the ratio of the minimum intensity between the two lines to the maximum intensity is equal to 79%*.

   For the configuration you have set up, measure this ratio, together with an estimate of its uncertainty. Does it match the theoretical value?

   Rayleigh ratio, minimum/maximum:

3. *Test problem:* Using a grating with 50 slits, and a slit separation of 0.0030 mm, find the smallest difference of two wavelengths around 550 nm that can be resolved in third order. Estimate the accuracy to which you can make this measurement.

   Hence calculate the measured resolving power, R, using Eq.(3.4), and compare it with the theoretical value, calculated from Eq.(3.5).

wavelength difference $\Delta\lambda$	*R (measured)*	*R (theoretical)*

4. *Test problem:* Imagine that you need a grating which is powerful enough to separate two spectral lines of wavelengths 600 nm and 615 nm — that is, to be sure that there are in fact two distinct lines present. Using a slit separation of 0.0030 mm, find the total number of slits necessary to do this in first order. Repeat for second and third orders.

   As a check, use Eq.(3.5) to calculate the expected number of slits for each order.

	measured value	*expected value*
Number of slits, first order		
Number of slits, second order		
Number of slits, third order		

Checkpoint 6:

Notes:

The Mach-Zehnder Interferometer

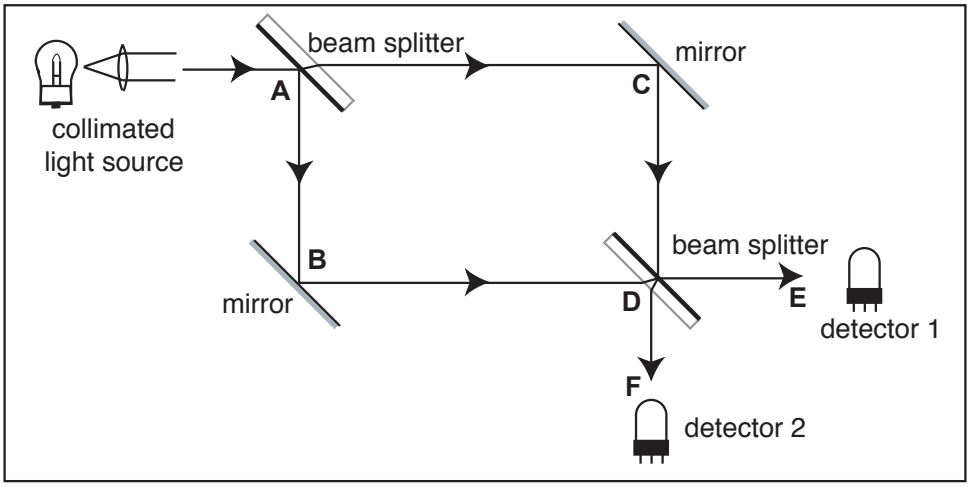
In this set you will investigate one example of an **interferometer***, and how it is used both in conventional optics and with optical fibres.*

This set of exercises should take you two weeks to complete.

5.1 Pre-lab

In optics, an **interferometer** is an instrument that uses interference patterns to make precision measurements, usually of wavelengths. In many such instruments, a beam of light is split into two or more beams before being brought together again. The re-combined beam then shines on a screen or some other detector in a characteristic interference pattern.

There are several interferometers in common use which share the same principle, but have different configurations, depending on paths taken by the split beams. One particular configuration is illustrated in the **Mach-Zehnder Interferometer** (see Physical Optics Lecture Notes, §2.3.1).



Plane wave light from the source strikes a **semi-reflective mirror** at A, often called a **beam splitter**, where it divides into two beams. The first beam *reflects* from the splitter towards B, where it reflects towards another beam splitter at D. From there half of its energy is *transmitted* towards E and half *reflected* towards F. The second beam is *transmitted* towards C and then reflected towards D, where half is *transmitted* towards F and half *reflected* towards E.

There are therefore *four* different beams to be considered, and in order to calculate the wave field at either of the detectors you need to know the way the amplitudes and phases of these four beams have changed since they were all a single beam at the point A. The following considerations apply.

- Because the beams are plane waves, their amplitudes do not decrease with distance. However at each reflection/transmission at a beam splitter, the energy of the beam divides into two. The ratio of the intensity of the reflected wave over the intensity of the incident wave is called the **reflectivity**, or **reflectance**, $\mathcal{R}$ of the semi-reflective surface. Note that the ratio of the amplitude of the reflected wave over the amplitude of

the incident wave is usually defined as the **coefficient of reflection** (r), and the reflectivity is $\mathcal{R} \equiv r^2$.

$$\mathcal{R} \equiv \left(\frac{\text{intensity of reflected wave}}{\text{intensity of incident wave}}\right) = \left(\frac{\text{amplitude of reflected wave}}{\text{amplitude of incident wave}}\right)^2 \equiv r^2 \tag{5.1}$$

It follows from conservation of energy that

$$\left(\frac{\text{intensity of transmitted wave}}{\text{intensity of incident wave}}\right) = \left(\frac{\text{amplitude of transmitted wave}}{\text{amplitude of incident wave}}\right)^2 = (1 - \mathcal{R}) \tag{5.2}$$

Ideally you would like the reflectivity of a good beam splitter to be 50%.

- The *phase* of an electromagnetic wave at any point in space, changes linearly with the length of the path along which the wave has travelled. If the wave has gone a distance L after leaving some reference point, the change in phase is proportional to how many wavelengths this distance is equivalent to. That is,

  $$\Delta\phi = 2\pi \frac{L}{\lambda} = 2\pi \frac{n\,L}{\lambda_0} \tag{5.3}$$

  where λ_0 is the wavelength *in vacuo* and n is the refractive index, since there is 2π phase for every wavelength of extra path. Refer to Physical Optics Lecture Notes, §1.3.4.

- There is one further point that must be borne in mind. For many dielectric materials, when a wave reflects from a surface of *greater* refractive index, it suffers a change of phase equal to π radians. But when it reflects from a surface of *smaller* refractive index, there is no change of phase. If you don't remember this result, read Physical Optics Lecture Notes, §1.3.6.

  Reflection from a mirror usually involves a change of phase of π. But at reflection from a semi-reflective mirror it is critical whether reflection takes place at the *front* or the *back* of the mirror, that is whether or not the beam travels *through* the glass before being reflected. In the present application, the reflection at A involves a π phase change, as does the reflection CDE. But for the reflection BDF there is *no phase change*.

You can now write down expressions for the amplitudes and phases of the two beams reaching detector 1, and those reaching detector 2. Assume the electric field of the initial beam at the point A, at time t, is:

$$E_{0,A}(t) = A_0 \cos(\omega t) \tag{5.4}$$

Then the electric fields of the two beams reaching detector 1 at the point E, the first which travels along the path ACDE, and the second along ABDE, are:

$$E_{1,E}(t) = A_0 \sqrt{(1-\mathcal{R})\mathcal{R}} \cos(\omega t - 2\pi n\, L_{ACDE}/\lambda_0 - 2\pi) \tag{5.5}$$

$$E_{2,E}(t) = A_0 \sqrt{\mathcal{R}(1-\mathcal{R})} \cos(\omega t - 2\pi n\, L_{ABDE}/\lambda_0 - 2\pi) \tag{5.6}$$

Likewise the electric fields of the two beams reaching detector 2 at the point F, travelling along ACDF and ABDF are:

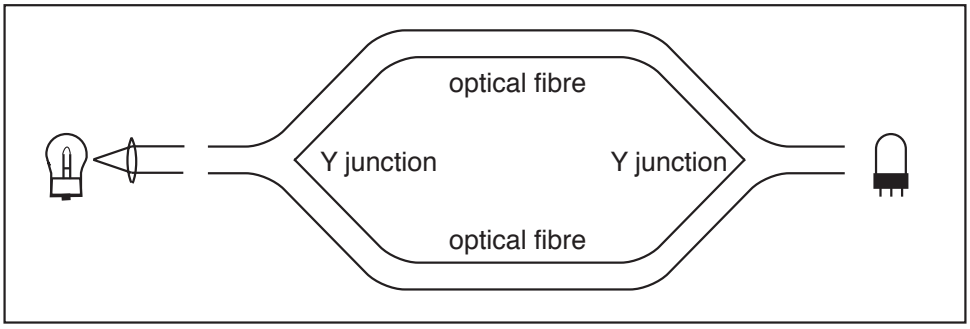
$$E_{3,F}(t) = A_0\,(1-\mathcal{R}) \cos(\omega t - 2\pi n\, L_{ACDF}/\lambda_0 - \pi) \tag{5.7}$$

$$E_{4,F}(t) = A_0\,\mathcal{R} \cos(\omega t - 2\pi n\, L_{ABDF}/\lambda_0 - 2\pi) \tag{5.8}$$

Note, in the ideal case where the reflectivity of each beam splitters is 50%, and the path lengths around both sides of the interferometer are equal, all the energy goes into one detector only.

You can see from Eqs.(5.5)–(5.8) that, if the refractive index is different in the two paths, whether there is destructive or constructive interference at the detectors will depend sensitively on that difference. It is not surprising therefore that the Mach-Zehnder interferometer finds its greatest application in measuring refractive indexes.

The principle of the Mach-Zehnder interferometer also finds application with optical fibres. The two optical paths are replaced by simple lengths of fibre or other light guides, and the two split beams are generated and recombined by Y-junctions, rather than semi-reflective mirrors.



Its usefulness lies in the fact that the refractive index of an optical fibre can easily be changed electrically. Such changes are usually very small, but the path lengths involved are a huge number of wavelengths, so any change in n (or L) produces a noticeable change in ϕ — that is, it is very sensitive, like all interferometric methods. Hence the signal received at the detector can be forced to vary between full constructive and full destructive interference, by varying the refractive index in one arm. This forms the basis of the **Mach-Zehnder modulator**. (See Physical Optics Lecture Notes, §2.3.0).

5.2 Lab Exercises

The central calculation in describing the working of a Mach-Zehnder interferometer is a simple Huygens' construction calculation. At one detector you add together values of the two wave fields which travel along the paths ABDE and ACDE. At the other you add together those travelling along ABDF and ACDF. The two calculations are essentially the same as were done in Lab 1, except for two new points.

The first new calculational element is the fact that the different beams travel along different paths, which may be of different lengths and may involve different refractive indexes. Hence, although the beams start off with the same **phase** at the point A, by the time they arrive at the detectors they will all have different phases.

The second new element is the fact that, at the one-way mirrors (the beam splitters), the **amplitude** of the beam which is reflected may not be the same as that which is transmitted, if the reflectivity of the beam splitter is not exactly 50%.

Therefore we will use very similar MATLAB code to model the Mach-Zehnder interferometer as we have been using so far. You are provided with two pre-written MATLAB scripts,

- **`eop3driver`**, to organize the calculations, as used in previous sets of exercises; and
- **`MachZehnderCalculation`**, (incomplete) to compute the light intensity at the two detectors.

Q1. Basic calculation

We tackle the problem of calculating the output from a Mach-Zehnder interferometer in two stages. First we write a MATLAB function to add together the beams arriving at the two detectors. When all that is working properly, we attach it to a 'driver'.

1. Download and open the m-file **`MachZehnderCalculation`**. Note that few of the real MATLAB statements have been entered, but the logic of the calculation is laid out in comments and dummy statements which set various quantities to zero. Note also that the script is written as a MATLAB **function** and that it requires one variable to be **passed** when it is called — a value of refractive index (which we will not use yet).

   For definiteness, define inside the script the following lengths (all in metres) which describe the physical instrument:

   lengths ACD and ABD: **`lACD`** = 0.100; **`lABD`** = 0.100;
   lengths DE and DF: **`lDE`** = 0.030; **`lDF`** = 0.030;

You will see shortly that the absolute values of these lengths are not important — only the difference in total lengths of the two paths. But for now, start off by taking a simple-minded approach and specifying the actual lengths.

Three other variables are needed — the wavelength of the light, the refractive index of air and the reflectivity of the beam splitters (assume they are both the same) — also defined inside the script. Choose

wavelength of the light:	`lambda`	=	500e-9;	(500 nm)
refractive index of air:	`n0`	=	1.00029;	
reflectivity of the splitters:	`R`	=	0.50;	(50%)

Even though it will not calculate anything yet, run this script to check the syntax is correct by typing in the command window:

```
>> Idets = MachZehnderCalculation
```

Note: Although the script requires one variable (`n`) to be passed when it is called, MATLAB allows you to proceed without the missing variable, provided that it is not referred to in the script (yet).

2. Now start filling in the MATLAB statements inside the script. Firstly, specify the amplitudes of the four beams, from Eqns.(5.5)–(5.8). It will prove most convenient to put them into a 2×2 matrix labelled `A`, in this order:

$$\begin{pmatrix} \text{path ACDE} & \text{path ABDE} \\ \text{path ACDF} & \text{path ABDF} \end{pmatrix}$$

so that the top row is relevant to detector 1, and the bottom row to detector 2.

Secondly, calculate the four optical path lengths, as another 2×2 matrix, labelled `PL`. (Don't forget to include the refractive indexes).

Thirdly, calculate the four phases, as yet another 2×2 matrix, labelled `phi`, using the path lengths you just calculated, and adding in the appropriate phase changes at reflection.

The rest of the computation proceeds just like the scripts you wrote in previous exercises. Calculate the electric fields at the detectors at times $t = 0$ and at $t = \tau/4$.

```
E0 = A .* cos(phi);
E4 = A .* sin(phi);
```

Then sum the fields to calculate the total electric field at the two detectors at times $t = 0$ and at $t = \tau/4$; and from that calculate the (time averaged) intensity at the 2 detectors.

3. Check that your script gives physically sensible results by running it with different values of $\mathcal{R}$. Choose $\mathcal{R} = 0.0, 0.5$ and 1.0, run the script each time and enter the computed intensities in this table.

$\mathcal{R}$	$I_{detector\ 1}$	$I_{detector\ 2}$
0.0		
0.5		
1.0		

The assumption $\mathcal{R} = 0$ implies that the semi-reflective mirrors have *zero* reflectivity — that is, they behave like simple sheets of transparent glass. Use this fact to explain why you got the result you did for the top line in this table.

Caution: Keep in mind that the amplitude of the light from the source was taken to be 1.0, and therefore its intensity is 0.5, in appropriate units.

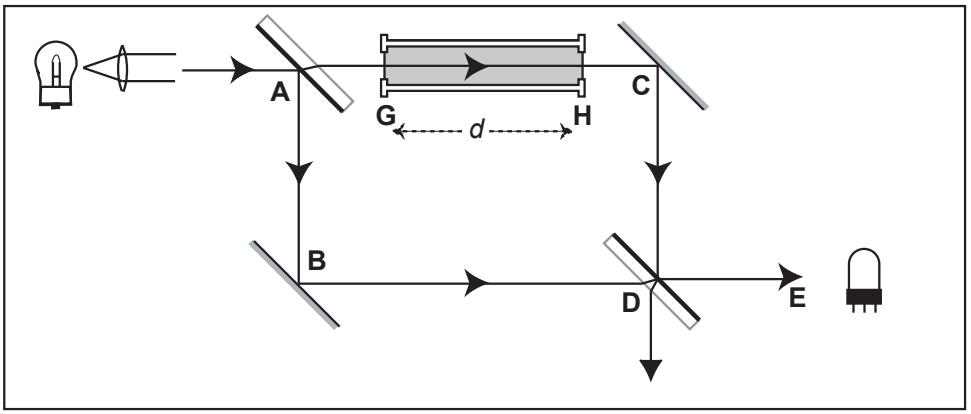
The assumption $\mathcal{R} = 1$ implies that the semi-reflective mirrors are perfectly reflecting — that is, they behave like simple mirrors. Use this fact to explain why you got the result you did for the bottom line in this table.

When $\mathcal{R} = 0.50$, you should have demonstrated that none of the incident energy went to detector 2, but all of it went to detector 1. For this result to be true, it is critically important that the optical path lengths ABD and ACD be exactly equal. Explain.

Checkpoint 1:

Q2. Varying the path length

It should be clear that the output of the detectors of a Mach-Zehnder interferometer are exquisitely sensitive to the optical path length its beams travel. If the length of either arm changes by just half a wavelength (a change less than 10^{-6} m in 10^{-1} m) the intensity can change from maximum to minimum. By the same token, it is also sensitive to changes in refractive index along the beam paths. That is why this instrument is often used to measure refractive indexes, particularly of gases.



Usually, a transparent cell of length d, filled with some gas, is introduced into one arm of the interferometer. The result is that the optical path lengths travelled by the various beams are different, and this difference shows itself as a change in the intensity measured by the detector. You are asked to make alterations to the script **`MachZehnderCalculation`** to compute this effect.

1. The only change you need to make is to add an extra optical path length for the two beams that travel along the arm ACD, given by,

   $$\Delta(\text{optical path length}) \quad = \quad (n - n_0) \times d \tag{5.9}$$

   where n is the refractive index of the gas in the cell, and n_0 is the refractive index of the medium in which the experiment is being done (probably air). Make this change by replacing the single MATLAB statement that calculated **`PL`** by:

```
PL  = n0 * [lACD+lDE, lABD+lDE ; lACD+lDF, lABD+lDF];
PL = PL + (n-n0)*d*[1, 0; 1, 0];
```

2. From now on you will be using this script to compute how the output of the detector changes as you alter the refractive index of the gas sample. So it is sensible to attach it to a 'driver' which will allow you to vary parameters easily and quickly.

   Firstly remove from inside your script, all the statements which set the values of the physical parameters R, n_0, λ and d.

   Note: there is no point in making the lengths of the four arms easily changeable, since their actual values are irrelevant, so long as **`lACD = lABD`** (exactly).

3. Download and run the script **`eop3driver.m`**, to check that the syntax is correct. Observe that, as the driver is written, you are expected to enter a range of refractive indexes, and it will plot the intensity at the detector as a function of n (actually $n - n_0$, because you typically need at least 5 figure accuracy to tell gases apart). It also prints out the value of the intensity at the maximum value of n.

   If you are not interested in a range of values of n, you can set the upper and lower limits equal, and the driver will display one intensity only.

   There is another variable, **`noise`**, which can be read from the input panel. You will not need this variable until later.

Q3. Measuring refractive index

1. The way this instrument is usually used in the laboratory is as follows. The sample cell is initially evacuated, and air (or other gas) is gradually allowed into the cell until it is at Standard atmospheric pressure, that is, the pressure has gone from zero to 101 kPa.

   Run the calculation using the following (default) parameters:

lambda	=	500e-9;	**R**	=	0.50;
n0	=	1.00029;	**d**	=	0.010;
nMin	=	1.00000;	**nMax**	=	1.00029;

   Assuming that the refractive index of a gas varies linearly with its pressure, draw in the box a graph showing how the intensity of the light at the detector varies as the pressure inside the cell has changed from zero to 101 kPa.

   You should have observed that the intensity at the detector fluctuates between zero and 0.5 several times as the gas cell is filled. In a real laboratory you have to keep watching the output of the detector and *count* the number of times it goes through the maximum–minimum–maximum cycle. In this present computation you can just read that number off the graph.

2. Repeat the same computation, this time varying the reflectivity of the splitters in steps of 0.1 between 0.0 and 1.0. Describe here how the output of the detector changes as $\mathcal{R}$ increases or decreases away from the ideal value of 0.50.

   On the basis of this observation would you say that operation of a Mach-Zehnder Interferometer is critically dependent on the reflectivity of the splitters being exactly 50%, or not?

3. *Test problem:* A Mach-Zehnder interferometer is used (open to the air) to measure the refractive index of a certain gas. The light source is a monochromatic beam of wavelength 632.8 nm. As the gas fills the 1.8 cm long gas cell, the output of the detector reaches a maximum 13 times. What is the refractive index of the gas (be careful to include an estimate of the error in your determination)? And can you identify what the gas is?

   Note: the refractive indexes for some common gases are:

Helium	1.000036	Hydrogen	1.000132
Air	1.000294	Nitrogen	1.000296
CO_2	1.000450	Xenon	1.000702

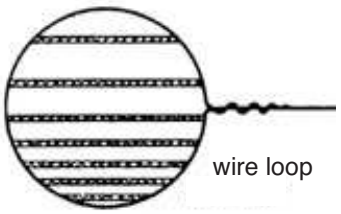
Measured refractive index: *Gas*:

4. *Test problem:* A Mach-Zehnder interferometer can also be used to measure the thickness of a transparent material. But it is only easy for very small thicknesses, because even 1 mm of glass would be equivalent to thousands of wavelengths. However it is ideal for measuring tiny *changes of length* of transparent crystals.

   Consider a glass slide, 5.0 mm thick. Its coefficient of thermal expansion may be taken to be 9×10^{-6} °C^{-1} and its refractive index to be 1.5 and constant (we choose to ignore any changes in refractive index resulting from temperature variations). If this slide were in one arm of a Mach-Zehnder interferometer with a HeNe laser as the source (λ = 632.8 nm) and if the temperature in the laboratory were to increase by 10 °C, by how much would the output intensity at the detector change?

   Change in intensity:

5. *Conceptual problem:* A thin film of soapy water on a wire loop is inserted into the test volume of a Mach-Zehnder interferometer so that the film is perpendicular to the HeNe laser probe beam. The detector is replaced by a camera lens, and the output beam is recorded as this photograph.



   Explain why you get this image.

   Hint: the soap film is thicker at the bottom than at the top because the water drains downwards.

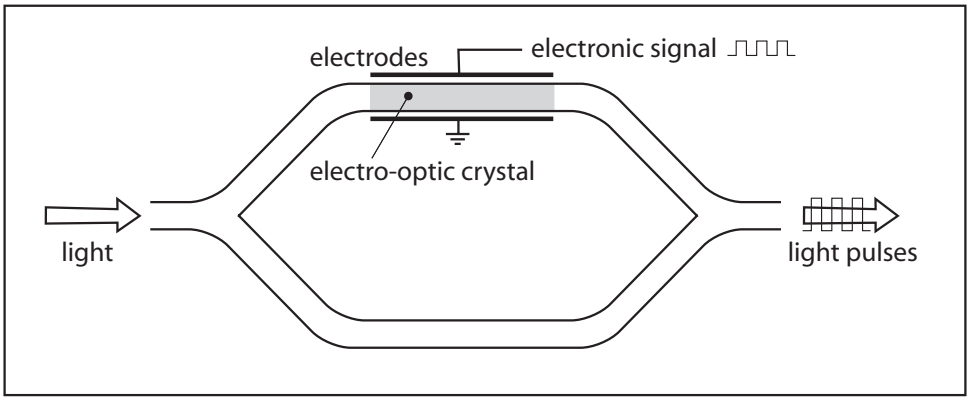
Checkpoint 2:

Notes:

Q4. The Mach-Zehnder modulator

As mentioned earlier (see page 3), a photonic device can be constructed using the Mach-Zehnder topology with various kinds of **light guides**, for example **optical fibres**, as the two arms. An initial beam of plane waves (amplitude 1.0 and intensity 0.5 units) is split into two equal beams (amplitude 0.71 and intensity 0.25) at a Y-junction. These travel down two different guides before being brought together again at a second Y-junction to form a beam whose amplitude and intensity vary, depending on the path lengths of the two 'arms'.

In some transparent materials, the so-called **electro-optic crystals**, you can change the refractive index by applying an external electric field. Hence by varying the electric field, you can control the intensity of the re-combined beam by causing it to interfere — constructively, destructively, or something in between. This is the basis of the **Mach-Zehnder modulator** as used in telecommunications. (See Physical Optics Lecture Notes, §2.3.0.)



Clearly, if the effect of the electric field is to increase the optical path length by a whole number of wavelengths, the output beam will have the same intensity as the input beam. But if it is equal to a half-integral number of wavelengths, no beam will emerge at the detector. By choosing proper values for the crystal you can switch the output from full on to full off. In other words, you have constructed an optical switch.

1. *Test problem:* The electro-optical property we are discussing is called the **Pockels effect**. It is expressed by this formula for the refractive index n as a functon of the electric field E,

$$n \approx n_1 + pE \tag{5.10}$$

One such material is **lithium niobate**, $LiNbO_3$. Its relevant properties are:

refractive index (zero field)	$n_1 = 2.20$
Pockels coefficient	$p = 3.7 \times 10^{-11}$ m/V

Consider the crystal in the diagram above to be a cylindrical sample of lithium niobate, approximately 10 cm in length and 1mm in diameter.

(a) Run **`eop3driver`** with a wavelength of 1550 nm, and define the material in the rest of the path to be glass, by setting $n_0 = 1.50$. Increase the length d slightly so that you get minimum intensity at the detector. *You can set* **`nMin = nMax = 2.20`** *for the crystal and get a single column bar graph for the intensity. Then increase d in increments of about 100 nm and watch the intensity go down to zero.*

Enter this length in the box below.

(b) Keeping that length, adjust the refractive index of the sample so that the intensity at the detector is maximum. *With* **`nMin`** *= 2.20 and* **`nMax`** *set at a slightly higher value you get a graph showing transition from zero intensity to maximum intensity with no other minimums in between.*

Enter the change in refractive index in the box below.

(c) Lastly calculate, using Eq. 5.10, the *potential difference* across the electrodes necessary to achieve this change in refractive index. *Note: this means across the thickness of the material, not its length. See the diagram above.*

length d	**change in** n	**potential difference**

2. *Test problem:* There are also **piezo-electric** materials, whose *length* changes slightly when an external field is applied. These too could be used as a phase shifter in an instrument like the Mach-Zehnder modulator.

   Again, this property is described mathematically by this formula for the **strain** — that is, the change in length divided by the original length — as a functon of the electric field which causes it,

$$\frac{\Delta d}{d} \quad \approx \quad K\,E \tag{5.11}$$

   Lead zirconate titanate (PZT) is a ceramic material that shows a marked piezoelectric effect. It has an extremely large dielectric constant and is used to make ultrasound transducers and other sensors and actuators, as well as high-value ceramic capacitors.

   Consider a thin film of PZT, inserted in the Mach-Zehnder modulator, approximately 1 cm in length and 10μm in thickness. Ignoring the fact that PZT is quite non-linear in its response to external stress, assume the following average electro/optical properties,

refractive index	$n_1 = 2.5$
piezoelectric coefficient	$K = 1.0 \times 10^{-10}$ m/V

   Repeat the computations you did in exercise 1, with the same wavelength and assuming the background material to be glass. Calculate an accurate length of the film which will maximize the output intensity, the increase in length necessary to minimize the output, and the potential difference necessary to cause this change in length.

length d	**change in** n	**potential difference**

Checkpoint 3:

Notes:

Q5. Noise

In any signal processing system, **noise** is a critical factor. Random fluctuations in any or all components of the system cause unpredictable changes in the output signal, and degrade the information content. (See Physical Optics Lecture Notes, §2.3.2).

In a Mach-Zehnder Interferometer, there are many possible sources of noise: fluctuations in air pressure, temperature and motion of the air; thermal expansion, twisting or warping of the glass; and so on. We will model the effect of noise by considering a simple instance. We will imagine that just one of the paths (ABD) is subjected to random variations in length, such that,

$$\Delta(lABD) \;=\; \text{noise} \times lABD \tag{5.12}$$

where the quantity 'noise' varies between 0 and a fraction of order 10^{-6}. (Remember from exercise 4 that the thermal expansion coefficient of glass is such that a 1° C change in temperature will change the length by a fraction of order 10^{-5}.)

In the MATLAB script, add a third line after the two MATLAB statements that calculated **`PL`**:

```
PL = PL + lABD*noise*rand*[0, 1; 0, 1];
```

The quantity **`noise`** can be read from the driver input panel, but make sure you add **`noise`** to the list of global variables in your script.

1. Run the Mach-Zehnder computation with the same parameters as in Exercise 5.2(1), but with **`nMin`** = 2.2 and **`nMax`** = 2.2 + 7.5e–6. You should observe that there is a smooth variation of the output intensity between maximum (0.5) when n = 2.2 and minimum (0.0) at n = 2.2000075.

   Now rerun the script with the quantity **`noise`** set at 1×10^{-6}, 2×10^{-6}, 3×10^{-6}, ... up to about 18×10^{-6}. You should observe increasing degradation of the output signal. Estimate the value of **`noise`** for which it would no longer be feasible to use the interferometer as an optical switch. What fluctuation in temperature would this degree of noise correspond to?

value of 'noise'	*temperature fluctuation*

2. It is often possible to design the system in such a way that any external disturbance affects both arms of the interferometer the same way. You might, for example, have the arms in close proximity and enclosed, so that they will experience the same changes in temperature (more or less).

   To simulate this, make one last change to the script to ensure the *there is exactly the same noise in both arms of the interferometer*. This is easily achieved by replacing the MATLAB statement above by,

   ```
   PL = PL + lABD*noise*rand*[1, 1; 1, 1];
   ```

   Repeat what you did in part (1), and describe in this box what you observe now.

   Note: whenever you make a change to the script **`MachZehnderCalculation`***, you need to restart the driver.*

   What you have just observed is known as **common mode rejection**.

   Remember that name.

Checkpoint 4:

Fibre Bragg Gratings

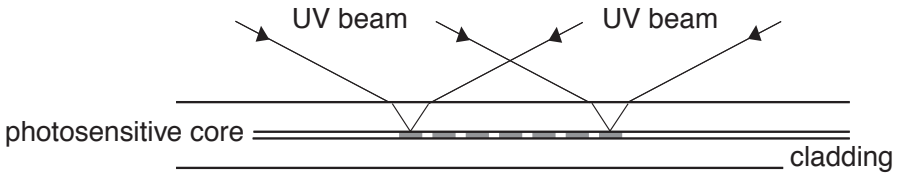
In this set you will investigate the passage of light through an optical fibre whose refractive index varies periodically along its length. It will be seen that such a fibre has some properties similar to those of an ordinary diffraction grating. This material is covered in Physical Optics Lecture Notes, §2.2.

This set of exercises should take you two weeks to complete.

7.1 Pre-lab

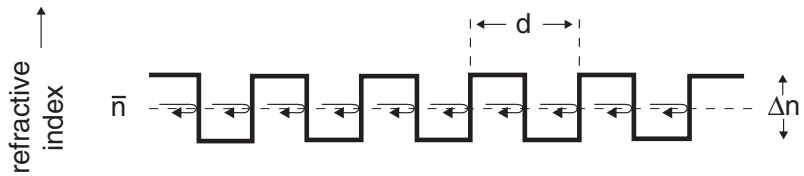
It was observed in 1978, during research into the properties of optical fibres, that exposure to ultra-violet light leads to a change in the refractive index of certain doped glasses (for example with germanium). For the first time it was realized that it was possible to imprint into an optical fibre a pattern of varying refractive index, which clearly would influence how light travelled through the fibre. In other words, it was possible to tailor-make optical fibres with an almost arbitrary range of optical properties.

One such is the **fibre Bragg grating**. It is usually fabricated by exposing the fibre, from the side, to an ultra-violet interference pattern. Since only the core is doped, a periodic variation is formed in the core but not in the cladding. Schematically the process might be represented by this diagram, showing that the refractive index in the core has been altered at the intensity maximums produced by two interfering ulta-violet beams.



Some features of what happens when light travels through this grating can be understood by the following very simple mathematical model. Let us consider that the variation in refractive index (1) is very small, and (2) changes up and down abruptly, like a square wave.

When light travels down the fibre, it travels essentially in one dimension, and every time it meets a discontinuous change a small amount of energy is reflected.



Compare this figure with Physical Optics Lecture Notes, Fig 84.

In general you might expect that you would never notice these reflections, especially if Δn is very small (and values of $\Delta n/\overline{n} \sim 10^{-3} - 10^{-4}$ are typical). Indeed, for an arbitrary wavelength, the weak reflections have no particular phase relation with each other, and therefore the *intensities* add together, giving only a small total reflection for the grating as a whole.

But now think what happens when the waves reflected off two successive (either both increasing or both decreasing) interfaces a distance d apart, are in phase. The (vacuum) wavelength of such waves is called the **Bragg wavelength**, see Young and Freedman, Eq. 2.2B,

$$\lambda_B = 2\overline{n}d. \tag{7.1}$$

Then all the reflections from all the like interfaces are in phase, and you must now add the *amplitudes* of the individual beams. The total reflectivity is a factor of M larger, where M is the number of grating periods. Since in a typical fibre grating $M \sim 10^4 - 10^5$, the grating's reflectivity can be substantial, even though the reflectivity of each interface is tiny. It is this behaviour, that a fibre grating can be highly reflective at certain specific wavelengths, but almost completely transparent for all other wavelengths, which we will explore in these exercises.

Calculating the total reflected wave, using this simplified model, should be meaningful for real fibre gratings provided that the assumptions going into the model remain valid — namely that the energy in the reflected wave is always a very small fraction of the energy in the incident wave.

However there is one theoretical point that must be borne in mind. When the wave reflects from a surface of *greater* refractive index, it suffers a change of phase equal to π radians. But when it reflects from a surface of *smaller* refractive index, there is no change of phase. If you don't remember this result, read Physical Optics Lecture Notes, §1.3.6.

So, consider the ray that goes back to the front end of the fibre after having reflected from the mth increase in refractive index. Its total phase change is equal to the phase change of the incident beam in going to the point of reflection *plus* the π during reflection *plus* the phase change involved in going back to the detector,

$$\phi_1 = 2\pi \frac{m\,\overline{n}\,d}{\lambda} + \pi + 2\pi \frac{m\,\overline{n}\,d}{\lambda} \tag{7.2}$$

Then there is the ray that reflects back to the front end of the fibre after having reflected from the next decrease in refractive index: the phase change of the incident beam in going to the new point of reflection *plus* nothing during reflection *plus* the phase change the reflected ray going back to the detector,

$$\phi_2 = 2 \times \left(2\pi \frac{m\,\overline{n}\,d + (\overline{n} + \Delta n/2)\,d/2}{\lambda} \right) \tag{7.3}$$

Therefore the total reflected wave field at the front end of the fibre (the detector) is calculated by adding together $2M$ fields, all with an amplitude a fixed factor of the incident field, and with the phases given by Eqs. 7.2 and 7.3.

The assumptions that went into our simple model break down badly whenever the intensity of the reflected wave becomes a substantial fraction of the incident beam (why?), which in turn becomes correspondingly less intense. We expect this to happen at wavelengths near the Bragg wavelength, Eq. 7.1. So, in the calculation just outlined, the energy reflected backwards at each interface is seriously overestimated. If you want to do a proper calculation, you will have to write down how the *two* waves, the incident and reflected waves, change from point to point along the fibre, as a pair of coupled differential equations.

We won't attempt to solve this problem directly, but merely quote the result. For a fibre grating with a uniform, sinusoidally varying refractive index, the ratio between the amplitude of the reflected wave at the end of the fibre to the incident amplitude — the total **reflection coefficient** of the fibre grating — is given by this (complex) formula:

$$r_{tot} = \frac{-q \sinh(\gamma L)}{\gamma \cosh(\gamma L) - i\,\delta \sinh(\gamma L)} \tag{7.4}$$

where the various quantities appearing in the equation are defined as follows:

$$\text{the wavelength detuning,} \quad \delta = 2\pi\,\overline{n}\,(1/\lambda - 1/\lambda_B) \tag{7.5}$$

$$\text{the coupling coefficient,} \quad q = 2\,\Delta n\,/\lambda_B \tag{7.6}$$

$$\text{the length of the fibre,} \quad L = M\,d \qquad \text{and} \tag{7.7}$$

$$\gamma^2 = |\,q\,|^2 - \delta^2. \tag{7.8}$$

The total **reflectivity**, the ratio of the two intensities, is the square of the magnitude of this quantity.

$$\mathcal{R}_{tot} \quad = \quad | \, r_{tot} \, |^2 \, . \tag{7.9}$$

Note: In Young and Freedman, §2.2.4, this result is presented as Eq. 2.2E, which is just our Eqn 7.4 in the special case when $\lambda = \lambda_B$. Note also that the coupling coefficient they use, κ_{ac}, (defined in Eq. 2.2D) is just our q multiplied by $\pi/2$.

7.2 In Lab

There are two independent calculations involved in our investigation of a fibre Bragg grating. The first is the same kind of Huygens' construction calculation we have been doing in these exercises up till now, which we expect to be valid only for wavelengths away from the Bragg wavelength. The second involves using Eq. 7.4, which should give trustworthy results at all wavelengths, and will hopefully match the first calculation at the wavelengths for which it is valid.

You are provided with three pre-written MATLAB scripts,

- **`BornApprox`** and **`FibreGrating`**, two uncompleted functions to perform these two calculations; and
- **`op4driver`**, to organize the computations in the later exercises.

Q1. The Born Approximation

The first exercise is to perform the calculation, described at the top of page 2, firstly for a single wavelength, and then for a range of wavelengths.

1. Download and open the file **`BornApprox`**. The logic of the calculation is laid out in comments but most of the MATLAB statements are dummies. You are required to enter the correct statements yourself in this exercise. It is written as a MATLAB *function* which requires one variable to be passed, the wavelength **`lambda`**. To start with, this quantity is a scalar. Set its value (in the command window) thus

```
>> lambda = 1.50e-6;
```

   The function returns the total reflectivity of the fibre, that is the intensity of the reflected wave at the detector (at the front end of the fibre) divided by the incident intensity.

```
>> Rtot = BornApprox(lambda);
```

   The other three important variables are the period of the grating, the mean and variation in refractive index. Choose

period of the grating:	**`d`**	=	5.0e-7	(500 nm)
mean refractive index:	**`nbar`**	=	1.5000	
variation in refractive index:	**`Dn`**	=	0.0004	

   Write the necessary statements into the script to set these variables. (These numbers mean that the Bragg wavelength is $\lambda_B = 1500$ nm. *Check this for yourself.*)

   You also need to specify the number of oscillations in refractive index, M. Set it, at least initially, to 50, and construct an $(M \times 1)$ column vector to number each of the oscillations, with these statements

```
M = 50; % total number of oscillations
m = [1:M]'; % numbering vector
```

   These values mean that the length of the grating is $L = 2.5 \times 10^{-5}$ m.

2. The next step is to calculate the phase of the ray that reflects from the mth increase in refractive index, using Eq. 7.2. Write the appropriate MATLAB statement to calculate this quantity, as an $(M \times 1)$ column vector. Call it **`phi1`**.

   Likewise write the appropriate MATLAB statement to calculate the phase of the ray that reflects from the next *decrease* in refractive index, using Eq. 7.3. This quantity is another $(M \times 1)$ column vector. Call it **`phi2`**.

3. We assume the amplitude of the original wave does not change significantly along the grating. Its value is arbitrary and may be taken as 1. The fraction of this amplitude which gets reflected is proportional to the ratio of Δn to $\overline{n}$.

   The proportionality constant has to come from theoretical arguments, which we won't go into here. The upshot is that this constant should be 1/2. So we have finally

   $$\text{amplitude of reflected waves} \;=\; \frac{\Delta n}{2\,\overline{n}} \tag{7.10}$$

   Write the MATLAB statement to calculate this quantity. Call it **`amplds`** (a scalar).

4. The rest of the calculation proceeds as in previous exercises. The two different reflected wave fields corresponding to each value of m are added, with appropriate phases. As discussed several times, you need to calculate the fields at times $t = 0$ and $t = \tau/4$,

```
E0 = amplds * (cos(phi1)+cos(phi2));
E4 = amplds * (sin(phi1)+sin(phi2));
```

   Each of these is an $(M \times 1)$ column vector, and you calculate the total field at the detector, by summing over all the M rays, giving two scalars,

```
E0tot = sum(E0,1);
E4tot = sum(E4,1);
```

   Lastly, square and add these amplitudes to get the total intensity at the detector.

```
Itot = (E0tot.^2 + E4tot.^2)/2;
```

   As a last twist, you want the total **reflectivity**. The amplitude of the incident wave was 1.0, so its intensity was 0.5. Hence

```
Rtot = 2 * Itot;
```

5. As soon as your script runs without errors, check that it does the arithmetic properly. Record the reflectivity calculated with the above parameters in the box below.

   There is an analytic value you can compare this with. Since the wavelength of the incident wave is equal to the Bragg wavelength, all the reflections should be in phase with one another. So the total reflected wave should have an amplitude equal to the sum of all the individual reflected waves. The total reflectivity should therefore be given by, (can you prove this?)

   $$R(\lambda_B) \quad = \quad (M\,\Delta n\,/\,\overline{n}\,)^2 \tag{7.11}$$

   Using this expression compare your computed value of reflectivity (to four significant figures) with the expected value.

computed reflectivity	*expected reflectivity*

Notes:

Q2. Variation with wavelength

1. You need now to repeat the calculation of reflectivity for a *range of wavelengths* rather than a single wavelength. This involves modifying the function you just wrote so that it treats **`lambda`** as a row vector, rather than as a scalar.

   In fact there is very little you need do, except for the following. In the statements that calculate **`phi1`** and **`phi2`** you replace the operation "**`/lambda`**" by "**`*(1./lambda)`**". You should be able to work out why this change is necessary when you realize that the quantities **`phi1`**, **`phi2`**, **`E0`** and **`E4`**, which were previously $(M \times 1)$ column vectors, are now rectangular matrices.

   When you have made these changes, run the function again, using these wavelengths,

```
>> lambda = linspace(1.4, 1.6, 201) * 1e-6;
>> R = BornApprox(lambda);
```

   The output **`R`** is now a row vector, the same size as **`lambda`**. Plot one against the other and draw in the box below the pattern you see on the computer when you do this.

2. Change the values of Δn and M and observe what happens to this pattern. Record in this box how the shape and size of the pattern changes as (1) Δn is increased; and (2) M is increased.

3. You will recall that this whole mathematical model relies on the assumption that the intensity of the incident wave does not change by too much as the wave travels down the fibre. A plausible rule of thumb might be to say that when the reflected wave gets to be ~ 0.25 of the peak incident intensity, that assumption must be starting to break down.

   At what value of the product $(M\,\Delta n\,/\,\overline{n})$ does this occur?

   At what value of the product qL (defined by Eqs. 7.6 and 7.7) does this occur?

   $(M\,\Delta n\,/\,\overline{n})_{max}$ $(qL)_{max}$

Checkpoint 1:

Q3. The Uniform grating model

The next exercise is to calculate the reflectivity for a real fibre grating for the same range of wavelengths, using Eqs. 7.4–7.9. The kind of grating these equations describe is one in which the physical characteristics of the fibre do not change along its length. We are therefore working with a **uniform grating model**.

1. Open the file **`FibreGrating`**. Again the logic of the calculation is laid out in comments but most of the MATLAB statements are dummies.

   The script has the same overall structure as **`BornApprox.m`**. So you can immediately copy across, from the old file, the beginning statements (under **`SET UP PHYSICAL VARIABLES`**) that are common to both. Make sure that, at least to start with, the physical parameters are the same as you used in Exercise OP7.7.2.

2. Replace the dummy statements under the heading **`CALCULATE THE THEORETICAL RE\-FLECTIVITY`** with the statements necessary to calculate $\mathcal{R}_{tot}$ from Eqs. 7.4–7.9. It is considered that these statements are straightforward enough that you do not need them to be spelled out here.

   Notice that Eq. 7.4 has terms in it which are explicitly complex. In physical terms this means that there is a change of phase between the incident and the reflected waves. This need not concern you. You are only interested in a ratio of *intensities*, and they do not involve the phase.

   So far as writing MATLAB statements is concerned you treat complex numbers the same as real numbers. The variable **`i`**, if you haven't redefined it elsewhere in your calculation, is treated by MATLAB as $\sqrt{-1}$.

   You need to keep in mind also that the quantity **`lambda`** which has to be passed to the function **`FibreGrating`**, is a row vector. In its turn it will make **`delta`** and **`gamma`** row vectors also. Hence you will have to be careful about element-by-element multiplications and divisions.

3. When you have got this function working, check that you have got everything correct by comparing it with the results you got in OP7.7.2. Do this by running

```
>> Rtheor = FibreGrating(lambda);
>> Rapprox = BornApprox(lambda);
```

   (with the same parameters).

   Then plot **`Rtheor`** against **`lambda`**, and on the same graph plot **`Rapprox`** against **`lambda`** also. (Don't forget to use **`hold on`**.)

   Describe here how well the two reflectivities compare with one another.

Checkpoint 2:

Q4. Strong uniform gratings

We now wish to explore the properties of strong uniform gratings, which are more commonly met in practice. That will involve much changing of the four physical variables. Hence it is sensible to work through another "driver".

1. Firstly, remove the statements inside the file **`FibreGrating.m`** which assign values to the four physical variables, d, $\overline{n}$, Δn and M, and replace them by

```
global d nbar Dn M;
```

   Then call **`op4driver`** and run it with the default values for the variables. Check that you get the same output as you did in OP7.7.2(3)

2. When everything works, start observing what effect each of the four physical variables has on the reflectivity diagram. Vary d within the range $(0.50 \pm 0.03) \times 10^{-6}$ m. Describe here the most obvious change to the reflectivity pattern. *(You will need to widen the wavelength range temporarily to at least 500 nm to 2500 nm.)*

   Vary $\overline{n}$ within the range 1.50 ± 0.10. Describe here the most obvious changes to the reflectivity pattern.

   You should have found that the wavelength of the central maximum — the **Bragg wavelength** — depends on the product of d and $\overline{n}$. Verify this by re-running the calculation with these pairs of values for $(d, \overline{n})$: $(0.60\times10^{-6}, 1.2500)$, $(0.50\times10^{-6}, 1.5000)$ and $(0.40\times10^{-6}, 1.8750)$. The only noticeable variation remaining should be that the height of the central maximum changes. *(Make sure you narrow the wavelength range again.)*

3. Now observe the effect of increasing Δn and L.

   Starting from the default values, run and re-run the calculation, multiplying **`Dn`** by 2 each time from 0.0004 up to 0.0128. Describe here what happens to the pattern.

   Starting from the default values, run and re-run the calculation, multiplying **`M`** by 2 each time from 50 up to 6400. *Note: the pattern gets very narrow quickly. You will find it necessary to change the range of wavelengths from $\lambda_B \pm 0.1$ nm to $\lambda_B \pm 0.01$ nm.*

   Describe here what happens to the pattern as M (and hence L) changes.

4. You should be able to see from Eq. 7.4 that the change in refractive index Δn only enters the formula for the reflectivity through the quantity q, and then only multiplied by L. The product qL is therefore an important quantity with which you describe the behaviour of a fibre grating. It is called the **strength** of the grating.

It is instructive to describe how the shape of the reflectivity graph changes as the situation changes from a weak grating ($qL \ll 1$) to a strong grating ($qL \gg 1$). Choose three different pairs of values for Δn and M which have the same (low) value for qL. Try for example, (0.0004, 1600), (0.0016, 400), and (0.0064, 100).

You should find that the width of the patterns varies a lot but the shape is the same for all three. Draw the common shape here. Calculate the value of the strength qL and write it on the graph.

Repeat the calculation for gratings of intermediate strength, using the values (0.0008, 3200), (0.0032, 800), and (0.0128, 200).

Repeat for strong gratings, using the values (0.0016, 6400), (0.0064, 1600), and (0.0256, 400).

Compare this last figure with Physical Optics Lecture Notes, Fig 86 (picture on the left).

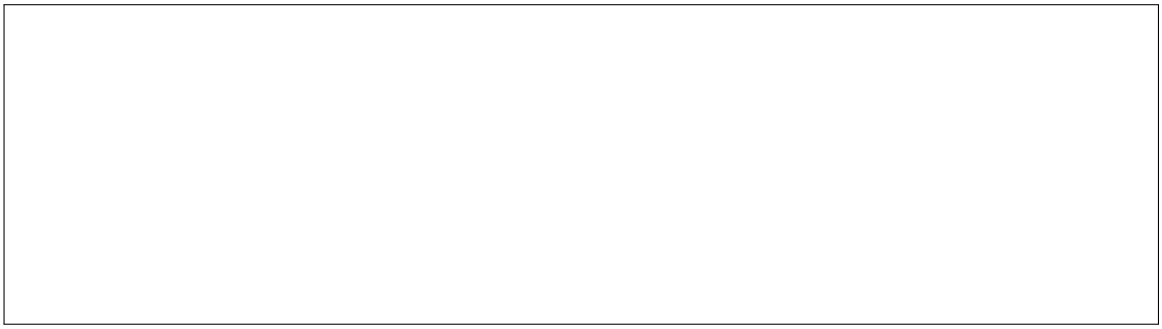
5. You should have noticed that for strong gratings, there is a narrow band around the Bragg wavelength where the reflectivity is almost constant and very close to unity. For wavelengths further out, the reflectivity fluctuates markedly but falls away quickly. This constant reflectivity band of wavelengths is known as the **photonic bandgap**. Clearly photonic bandgaps do not occur in weak or medium gratings.

   In engineering contexts it is more usual to talk about the **transmissivity** of a grating, rather than its reflectivity — that is, the ratio of the transmitted energy to the incident energy. The two quantities are related by: transmissivity = (1 − reflectivity). Furthermore, this ratio is usually measured on a *logarithmic* scale, in **decibels, dB**, which conventionally measures any intensity I in terms of a reference level I_0 thus,

$$\text{relative intensity (in dB)} \quad = \quad 10 \log_{10}\left(\frac{I}{I_0}\right) \tag{7.12}$$

   Repeat the last calculation, *without using the driver*. (Remember that all the physical variables are global.) Then plot the *transmissivity* in decibels as a function of λ.

```
< lambda = linspace(lambdaMin, lambdaMax, numSteps);
< Rtheor = FibreGrating(lambda);
< plot(lambda, 10*log10(1-Rtheor));
```



Notice particularly that, although the reflectivity curve looks completely flat within the bandgap, there is considerable variation evident in the transmissivity curve.

Compare this figure with Physical Optics Lecture Notes, Fig 86 (picture on the right).

Checkpoint 3:

Notes:

Q5. Photonic band gap

Clearly a photonic bandgap in a strong Bragg fibre grating is potentially of enormous practical importance. And the immediate question is: what determines how *wide* the bandgap is?

1. It is pretty clear what we mean by the width of the bandgap $\Delta\lambda$, but there is some uncertainty in its definition. Repeat a calculation with $qL \sim 8$ so that you have the figure in front of you and examine very carefully the edges of the bandgap. You will need to increase the fineness of the range of values of λ which you use in the calculation.

   One method of measuring $\Delta\lambda$ consistently is to measure the wavelengths at which the inner maximum falls to half its peak value — which in this case is 0.500 exactly. This is the method you should use here. In technical jargon, you are measuring the **full width at half maximum (FWHM)** of the photonic bandgap.

2. Choose the following six sets of values — two with the same grating strength, two with different strengths and two with different mean refractive indices. Measure $\Delta\lambda$ for each set and enter the value in the fourth column. Then fill in the last two columns.

   Note: You will find difficulty in reading numbers off the graph accurately enough. You should use **`ginput`** *and* **`format long`**.

$\overline{n}$	Δn	M	$\Delta\lambda$ (nm)	$\Delta\lambda/\lambda_B$	$\Delta n/\overline{n}$
1.50	0.0004	25 600			
1.50	0.0016	6 400			
1.50	0.0008	102 400			
1.50	0.0032	25 600			
1.60	0.0004	25 600			
1.40	0.0004	25 600			

   Hence write down a useful empirical formula which may be used to calculate the width of the photonic bandgap.

   Your result should be identical with Young and Freedman, Eq. 2.2C (except for the constant).

Checkpoint 4:

Notes:

Q6. Apodization

Although strong fibre gratings show a reasonably well defined photonic bandgap, there are a lot of sidebands which allow significant reflection of light outside this region. Theory says that these sidebands arise from the sudden mismatch in the properties of the fibre at the ends of the grating. Therefore it should be possible to reduce the sidebands by **tapering** the grating — that is, by making the amplitude of the refractive index modulation very small at the ends, and bigger at the centre. This process is called **apodization**. See Physical Optics Lecture Notes, §2.2.3.

In terms of our simple mathematical model, this means that Δn varies smoothly along the fibre length, high in the middle and zero at the ends. We could use any simple functional form, like a Gaussian or a $\sin^2$. Since we didn't actually do the calculation which led to Eq. 7.4, we cannot demonstrate the effect of apodization directly. However we can see the kinds of effects it produces by applying it to the calculation of OP7.7.2.

1. Open the file **`BornApprox`**, which you worked with before and save it under another name, e.g. **`ApodizedGrating`**. Previously, the quantity **`Dn`** was a scalar — it had the same value for all the fluctuations along the fibre (denoted by the counter **`m`**). Now you want it to have a different value for the different values of **`m`**. It must be replaced by a *vector* of the same size as **`m`**. We will give it the name **`Deltan`**.

   To start with, so you can check for errors, make all the elements of **`Deltan`** the same, and equal to the scalar **`Dn`**. Add to the setting-up section of the script

```
Deltan = Dn * ones(M,1);
```

   and replace **`Dn`** by **`Deltan`** in the statement which calculates **`phi2`**. The only other change you need to make is in the statement which calculates **`amplds`**. This was a scalar — all the reflected waves had a common amplitude. But now, since there are **`M`** reflected rays, and **`numSteps`** wavelengths, this quantity must be a **`M`**×**`numSteps`** matrix.

```
amplds = Deltan * ones(size(lambda)) / (2*nbar);
```

   You will also need to be careful about element-by-element multiplication in the calculation of **`E0`** and **`E4`**.

2. Run the script (remembering you must supply an array of values of **`lambda`**,

```
>> Rapod = ApodizedGrating(lambda);
```

   Plot the results, **`Rapod`** vs **`lambda`**, and check that it gives you exactly the same output as when you ran it previously.

   For future reference run it again using parameters for a *strong grating* — e.g. $\Delta n = 0.004$, $M = 1200$. When you plot it, truncate the y-axis at $R = 1$, to mimic the behaviour of a real grating, where the reflectivity cannot exceed 1. Use

```
>> set(gca,'Ylim',[0,1]);
```

   Draw the shape of your reflectivity vs wavelength graph here.

3. Now add a statement to the script to make **`Deltan`** vary from one end ($m = 1$) to the other ($m = M$). It could be

```
Deltan = Dn * sin(pi*(m-1)/(M-1)).^2;
```

Run the script with this statement included, and with exactly the same parameters.

Plot the result and draw the shape of your graph here.

Checkpoint 5:

Notes: