PHY1010:

An Introduction to

Physics through Experiments

Undergraduate Physics Laboratory

Spring Semester

Preface to the first edition

The contents of this book were written over a period of three years, from 2017 to 2020, when I was the Teaching Fellow for the introductory laboratory course at Ashoka University. This course, titled *An Introduction to Physics through Experiments*, was envisioned by the Department as a "gateway" course to give students a feel of what they could expect from doing physics at Ashoka. The aim of the course is to help students develop an intuitive understanding of experimental methods, and to develop a familiarity with real-world data.

The laboratory also has a strong focus on improving the student's report-writing skills. A typical iteration generally begins with three introductory sessions during which students are introduced to data collection, graphing, and uncertainty analysis. The sessions usually deal with a conceptually simple experiment that the students are already familiar with, like the Simple Pendulum. After each session, the students are asked to submit a draft report on the experiment by the next week. The course instructor or Teaching Fellow then provides each student with feedback about their drafts (preferably individually) and comments on how they may be improved. The process is repeated three times until the student produces a final report on the introductory experiment after three sessions. The results so far have been extremely promising: after just three sessions, aided by regular feedback, students are often able to write extremely engaging reports and – what is more important – are able to do so in their own unique voice.

I would like to express my gratitude to Dr. Sabyasachi Bhattacharya (then C. V. Raman Professor at Ashoka and currently Director of TCG-Crest) for the ideas and years of experience in experimental physics that he brought to the laboratory. I would also like to thank Professor Bikram Phookun who contributed a great deal to the introductory sessions and also helped in editing the entire document and proof-reading it multiple times. This work would not have been possible without his guidance.

And lastly, I would like to express my appreciation to the first batch of physics students at Ashoka University who took this lab in the Spring of 2018 and who – despite never having the chance to use these manuals because they weren't yet completed – contributed greatly to my understanding of all the experiments detailed in this book. I state their names here for posterity: Aditya Singh, Aishwarya Jain, Anand Waghmare, Heer Shah, Nayanika Krishnan, Rahul Menon, Rashmi Gottumukkala, Riya Banerjee, Shwetabh Singh, Sreerag K P, Sreya Dey, Srinidhi Pithani, Vidur Singh, and Yajushi Khurana.

Philip Cherian
Teaching Fellow (2017 – 2020)

Contents

I Introductory Sessions **1**

1 Data Collection, Recording, and Interpretation **2**

1.1 Objectives 2

1.2 Introduction 2

1.3 Data Collection 2

1.4 The Experiment 6

2 Graphing **8**

2.1 Objectives 8

2.2 Introduction 8

2.3 Graphing using a pencil and graph paper 8

2.4 Graphing using your computer 10

3 Sensitivity to Errors in Measurement and Preliminary Error Analysis **25**

3.1 Objectives 25

3.2 Introduction 25

3.3 Systematic Errors 26

3.4 Vernier Calipers and the Screw Gauge 27

3.5 Random Errors 30
3.6 Errors in Time Measurements 31
3.7 Propagation of Errors 34
3.8 Using Error Analysis to Plan Your Experiment 35
3.9 The Experiment 36

II Experiments 37

1 The Soft Massive Spring 38

2 The Three-Terminal Black Box 50

3 The Incandescent Lamp and the Inverse Square Law 58

4 Diffraction of Light 65

5 Formation of Rainbow 75

6 The Carey Foster Bridge 89

III Appendices 95

1 Writing a Lab Report 96

1.1 Objective 96
1.2 Introduction 96
1.3 Essential Elements of a Report 97
1.4 Conclusion 101

3 More Advanced Error Analysis 114

3.1 Objectives 114

3.2 Introduction 114

3.3 Random Errors 115

Part I

Introductory Sessions

Session 1

Data Collection, Recording, and Interpretation

1.1 Objectives

1. To understand basic data collection, representation, and interpretation using a simple pendulum.

1.2 Introduction

In this preliminary experiment you will use a simple and familiar system, the simple pendulum, to understand measurement, data collection, and elementary data interpretation. In this experiment, as in many others that you will do in this lab, you will observe a system that is well understood. You will often want to know how a certain property of the system changes when one of the parameters of the system is changed. For example, in this experiment, you might ask: How does the the time period vary with the mass of the bob? How does the time period vary with the length of the pendulum? Alternatively, you may want to set up an experiment that leads to a particular value for a certain property. For example, you may ask: How long does the pendulum need to be for its period to be 1 second?

1.3 Data Collection

Data collection is the observing and recording of data using instruments. You will record data in a *log book* – preferably a thick and sturdy hard-covered notebook that lasts the entire semester (or, if you intend to major in physics, perhaps all your years at Ashoka).

1.3.1 Recording Raw Data

Raw Data vs Processed Data

Raw data is what you see on your measuring instruments. Recording it is the first important step in experimental physics. Raw data is distinct from processed data, and it is important never to forget the difference between them. For example, if you measure the time for 10 oscillations of a pendulum, it is the time for 10 oscillations that is your raw data. You may be interested in the time period of the pendulum (i.e. the time for 1 oscillation) and to do that you will have to divide the measurement you have made by 10; the moment you do this, you have performed an operation on it, i.e. you have processed your data.

> In experimental physics, you must never, under any circumstances, process your data on the fly before recording it: what you must record, first and foremost, is what you see on your measuring instruments. Later, you can process it and change it to any other form that suits you.

Instrument Parameters

The instrument gives you a number, which will in general have a certain dimension, with units. To make sense of the data you record, you will have to first record the units in which the numbers on the measuring device are expressed. For example, if you use a *mm* scale to measure the length of an object, you will need to record *mm* as your units before you take any actual readings. In a certain sense, the unit in which the number on the instrument is expressed is your first observation. (There is a reading that needs to be taken even before that – the date on which the experiment is being done!)

Tabulating Data

In physics we are interested in patterns. One of the most compact and effective ways of recording data so that patterns can be seen easily is the table. For example, we may want to find the answer to the following question: How does the time period of a simple pendulum vary with the mass of the bob? To answer this question, we will want to take many measurements of the time period, one (or more) for each mass of bob. It follows naturally that we should align each mass with the corresponding time period, with one reading following another in vertical progression. Such an arrangement is called a table. Spend some time deciding what data table to draw before beginning your experiment: it will help you decide what data to take.

There is no strict "right" way to tabulate data.[1] An example is shown in Table (1.1).

[1] However, there are several wrong ways.

Mass of bob (g)	**Time for 10 oscillations** (s)			**Time Period** (s)
	Trial 1	Trial 2	Trial 3	

Table 1.1: Sample data table

In the table above, the raw data are recorded in the columns for the *Mass of bob* and *Time for 10 oscillations*. Note how the unit in which the measurement is taken (read from the instrument used) is noted next to the title.

Data Ethics

It is essential that none of these readings be changed; if you wish to change a reading, strike it out gently, in such a way that it can still be seen, and write down, next to it or under it, the new reading. It is essential to have a record of all the data you have taken. Remember that this does not need to be excessively neat, only understandable to you and a trained physicist looking at your log book later, e.g. your TF or your instructor.

If you imagine that the purpose of the experiment is to get a certain result, then you will sometimes be tempted to get rid of "wrong" data, but if you remember that the purpose of the experiment is to learn the methods of experimental physics, you will realise that there is no such thing as wrong data.

This does not imply that you will always use all the data that you record. But, in general, when you discard certain data, or decide not to use it – a process that is called "flagging data" – you will have a reason for doing so.

Before you leave the lab show your data log to your TF or course instructor and ask him or her to initial it.

"How many readings should I take?"

This is the question we hear the most often. The answer is of course that **it depends**. It certainly depends on the experimental setup, and it depends on the different types of uncertainties present in each measurement.

While some scientific measurements are exact,[2] others – such as the sorts of measurements you will be doing in this lab – are not. When we make a measurement of a quantity, we generally assume that some exact or true value exists for each, based on how we define what is being measured. We attempt to find these quantities as best we can, with the available resources, keeping in mind what

[2] Counting the number of parents you have, for example.

we want to get eventually from the data and the effect that errors in measurement could influence our desired result (more on this in the chapter on error analysis). Appreciating this helps us understand how many readings to take.

You will notice that you usually obtain slightly different results on making multiple measurements of the same quantity. We will deal with this in great detail in the chapter on error analysis, but for now you may imagine that these are a result of random fluctuations about the "true" value. One way to increase your confidence in experimental data is to repeat the same measurement many times and take the average. (We will have more to say about this later, but let us understand it intuitively first.)

For example, one way to determine the time period (T) of a pendulum is to measure the time for 1 oscillation. But we all understand, intuitively, that it is perhaps better to measure the time for 10 oscillations and divide by 10. This is true. (Why this is true is not, if you think about it, so obvious – but we will go into that later.)

But we can find the time for 10 oscillations in two different ways: (i) we can find the time for 10 oscillations in a row; (ii) we can find the time for 1 oscillations 10 times. Do these two procedures amount to the same thing? It turns out that they do not – in the sense that the error associated with the two procedures are different. (Once again, why this is true is not obvious; let us accept it for now and understand it later.)

Given that these two procedures for measuring T are not identical, one may now ask the following question: If I want to measure the time for 100 oscillations, is it better to do that all at one go or divide it into 10 trials of 10 oscillations each? (The answer to this question is not obvious either!)

There are all kinds of subtleties in data collection that you will slowly come to grips with. The answer to the question that is the title of this subsection is not obvious, and you will learn it slowly.

Question: Can you explain why taking the time period for a 1000 oscillations may not give you an answer that is much more accurate than 100 oscillations?

1.4 The Experiment

1.4.1 Apparatus

1. Metallic bobs of different materials
2. A length of string
3. A cork with a slit
4. A retort stand with an attached protractor
5. A stopwatch
6. A scale
7. A weighing balance

1.4.2 Suggested Procedure

Question: Which are the different physical quantities in this problem that can be varied to potentially change the time period?

Part A

In this part of the experiment you will design a simple experiment to determine the variation of the time period with the mass of the bob.

1. Begin by deciding which variables you need to fix, and which variables you will change.
2. Draw out an appropriate table in your auxiliary notebooks. Mark out any important details that would help you remember what you've done when you re-read this. Remember to state not only what you have changed, but also what you have kept *fixed*.
3. Decide on the **number** of readings you will take. When you have arrived at a number, try to *justify* it.
4. Perform the necessary experiment, varying the relevant parameter. Note down your data.

Part B

In this part of the experiment you will design a simple experiment to determine the variation of the time period with the length of the string from the pivot to the centre of mass of the bob.

1. Begin by deciding which variables you need to fix, and which variables you will change.
2. Draw out an appropriate table in your auxiliary notebooks. Mark out any important details that would help you remember what you've done when you re-read this. Remember to state not only what you have changed, but also what you have kept *fixed*.
3. Decide on the **number** of readings you will take. When you have arrived at a number, try to *justify* it.
4. Perform the necessary experiment, varying the relevant parameter. Note down your data.

Part C

In this part of the experiment you will design a simple experiment to determine the variation of the time period with the angle of release of the bob.

1. Begin by deciding which variables you need to fix, and which variables you will change.
2. Draw out an appropriate table in your auxiliary notebooks. Mark out any important details that would help you remember what you've done when you re-read this. Remember to state not only what you have changed, but also what you have kept *fixed*.
3. Decide on the **number** of readings you will take. When you have arrived at a number, try to *justify* it.
4. Perform the necessary experiment, varying the relevant parameter. Note down your data.

Note: The repetition is – of course – intentional. We have found that students usually jump through these steps and – as a result – spend much of their time painstakingly collecting data that is of little or no use. It is essential that you spend some time deciding what exactly you want to collect, and how best you will represent it, before actually spending any time with the apparatus.

Question: In each of the above cases, which graph would be the best to plot? Why?

Question: It ought to be pretty clear that T should depend on the gravity of the Earth. So, ideally, we would like to vary the gravitational force as well, and see how T changes. Can you think of a way in which this can be done?

Question: On the basis of your experimental results (not what you learnt in school about the simple pendulum) what parameters do you find the time period T depends on? How confident are you of this conclusion? What is the best way to think of the parameter(s) on which T depends?

Question: What would be the effect – if any – of changing the **shape** of the bob on the time period? Justify your answer.

Session 2

Graphing

2.1 Objectives

1. To learn to plot graphs using (i) a pencil and graph paper, (ii) spreadsheets, and (iii) Python.

2.2 Introduction

Before you begin, read the article on graphing by Christopher Deacon. You don't need to grasp all the points in it right way, but as you grow in experience you will appreciate them.

While we will show you how to use Python, the purpose of this lab is not to familiarise you to coding – there is another course that does that. Everything here can alternatively be done using MS Excel.

> As with the other Physics courses, we will be using **Jupyter Notebook** for coding in Python. Go to the link and follow the instruction procedure to install Anaconda, a Python distribution that comes with Jupyter pre-installed. This installation takes a significant amount of time. As a result, you should get it done **before this session begins**.

2.3 Graphing using a pencil and graph paper

Graphing, like many other scientific activities, begins with a set of questions: we wish to find something, in this case a pattern, and we would like to plot our graph in such as way as to see and show that pattern most effectively. Some of these patterns may already be obvious in the data, e.g. you must have noticed that the time period T of the pendulum does not seem to depend on the

mass m of the bob. When you compare the columns for T and the length l of the pendulum, you see that T increases with l, but not in the most obvious way. Finally, when you look at how (and whether) T changes with the amplitude θ of the oscillation, you will see no obvious pattern, even though there is one hidden in the data.

2.3.1 Plotting: Stage 1

Let us begin by exploring the variation of T with respect to m and with respect to l. We will consider the variation of T with θ_o later.

1. On a graph sheet, draw the x and y axes. The x axis is normally used for the parameter that you control, sometimes called the independent parameter, and y axis for the parameter that you observe, sometimes called the dependent parameter. For example, when you plot T vs m, T is along y and m along x.
2. Along each axis, choose a range that (i) covers the values you use and observe, and (ii) uses a scale that assigns a reasonable value to each division of the grid. So far as possible use units that make the numbers for the variables (i) small and (ii) similar along both axes. (If you write the mass in kg and the time in s, you will understand what *not* to do.)
3. Plot the observed point, and then circle each point so that it visible.
4. If the pattern appears to be straight line, use a ruler to draw the line that seems intuitively the most appropriate. (We will see later what this means.) If the pattern seems non-linear, sketch free-hand the curve that seems, once again, most appropriate.

2.3.2 Plotting: Stage 2

You will have noticed that the way T changes with l is not linear. In fact, if you look carefully at the values, you will see that when l is doubled, T changes by about a factor of 1.4. Since 1.4 is close to $\sqrt{2}$, this suggests that it is T^2 that doubles when l is doubled. In other words, we expect a linear dependence of T^2 on l. To see if this true, plot T^2 against l.

2.3.3 Plotting: Stage 3

If T^2 depends linearly on l, T must depend linearly on $\sqrt{l}$. This implies that $\log T$ depends linearly on $\log l$. Make columns for $\log T$ and $\log l$ in your table, and make the corresponding plot.

Question: Compare the slopes of the $T^2 - l$ and $\log T - \log l$ graphs.

Question: What are the advantages of plotting the log of a quantity?

Question: Suppose that you are examining the variation of y with respect to x, it is sometimes useful to plot $\log y$ vs x (not $\log x$). What kind of dependence would this indicate?

2.4 Graphing using your computer

When you were drawing the $T^2 - l$ and $\log T - \log l$ graphs you were essentially using your artistic sense, which allowed you to judge what the *best* straight line is, i.e. the line that seems to best fit the points. Notice that the confidence with which you arrive, artistically, at the the best line depends on the scatter of the points about it. Thus, the greater the scatter, the less confident you are (or ought to be) of your artistic fit. It becomes imperative, therefore, to find an unambiguous, quantitative way of arriving at the best fit.

To do this we begin by understanding that a straight line is determined *completely* by just two points. In other words, when we have more than two points, there is no such thing as *the* line that fits the points. In mathematical language, the problem is *over-determined*. However, we can certainly ask what is the *best* line, i.e. the one that is most representative of all the points, or, in some sense (yet to be defined) closest to them.

To find the best line, we go back to the remark that a line is determined completely by two points. Another way of saying this is that a line is determined uniquely by two *parameters* – the slope m and the y-intercept c. So when we have many points, we can think of a host of lines with different with a range of different ms and cs. The question is: Which line is the best? To answer this question mathematically (rather than artistically) we need a quantitative measure of the **goodness of fit** of a line given more than two points. This measure of the goodness of fit must depend on m and c, and the process of finding the best line is thus the process of *optimisation* with respect to m and c. The mathematics of this is a little advanced, but by the end of this semester you will be able to work it out.

For the moment, you will allow the computer to use standard algorithms to find the best straight line for us. To do that, you have to learn how to get our computer to import the data we want to plot and use it in such a way as to find the best straight line. For that purpose we will first use Microsoft Excel and then Python.

Question: Show that if you have just two points, then m and c are uniquely determined.

2.4.1 Graphing with Microsoft Excel

Excel is a *spreadsheet*. Computer spreadsheets are powerful tools to analyse data and are used in a variety of different disciplines as they allow one to represent and analyse data easily and efficiently.

Tip

We will be describing using Microsoft Excel in this section. However other spreadsheet programs are very similar. For example, LibreOffice Calc and Google Spreadsheets are two other programs that can be used and possess more or less the same functionality.

Tabulating Data

Let's begin with a sample data table verifying Ohm's law, shown in Table (2.1).

Experiment n: **Verification of Ohm's Law for a large resistance**	
Data collected by: Philip Cherian, Expt. performed : 29/01/2019 to 05/01/2019	
Resistance $R1$; Multimeter *Victor* VC97	
Voltage (V)	**Current** (μA)
1	3
2	5
3	6
4	7
5	10
6	13
7	15
8	16
9	19
10	20
11	22
12	24
13	26
14	29
15	31
16	33
17	35
18	36
19	39
20	41

Table 2.1: Sample data for verification of Ohm's law for a large resistance.

Reproduce the above data table in MS Excel (be sure you learn how to include the μ symbol![1]).

Tip

Leave a few rows at the top of the spreadsheet for information about the data collected. In the above case we have included the information about who took the data and when, and also information about the equipment.

[1] You will often need to insert Greek characters (λ, θ, $\sigma \ldots$); add these using the `Insert` → `Symbol` option in MS Excel. Google Sheets does not have this functionality, but you can paste special characters from Google Docs.

Plotting and Formatting your Graph

You are now ready to begin plotting this data. Begin by selecting both columns and `Insert` a `Chart`. When you are asked to select a `Chart Type`, select the `XY (Scatter)` option with points **only**. This is important: you might be tempted to "join the dots", but that would be misleading: especially if the graph is not linear! When you are done with this, label your axes (**including units!**): most spreadsheet software take the first column to be the x-values and the second the y-values, but this can be changed. Your graph should look something like Figure (2.1).

	A	B
1	Experiment *n*: **Verification of Ohm's Law for a large resistance**	
2	Data collected by: Philip Cherian, Expt. performed : 29/01/2019 to 05/01/2019	
3	Resistance *R1*; Multimeter *Victor* VC97	
4	**Voltage (V)**	**Current (μA)**
5	1	3
6	2	5
7	3	6
8	4	7
9	5	10
10	6	13
11	7	15
12	8	16
13	9	19
14	10	20
15	11	22
16	12	24
17	13	26
18	14	29
19	15	31
20	16	33
21	17	35
22	18	36
23	19	39
24	20	41

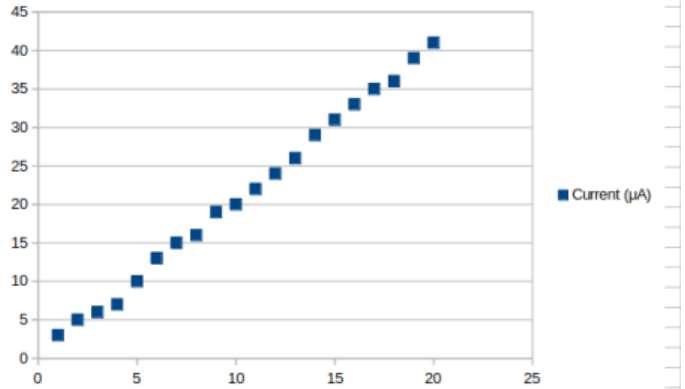


Figure 2.1: Sample graph of the data in Table (1.1) with much to be improved.

Excel's biggest user base is business, so default graph formats are mostly setup for that purpose. As a result, changes need to be made to bring it to a "scientific" format:

1. Remove any unnecessary white space. You can do this by adjusting the minimum and maximum values of the axes. Do this by right clicking the axes and selecting `Format Axes`.
2. Remove the legend on the right when possible: any information regarding the graph should be added in the caption. You can select it and press `Delete`.
3. Change the marking icon to a more scientific size and shape. You can right click the graph and select `Format Data Series` and edit the icon size and shape there.
4. Remove the grid-lines in the plot area.

You should now have a graph like the one shown in Figure (2.2).

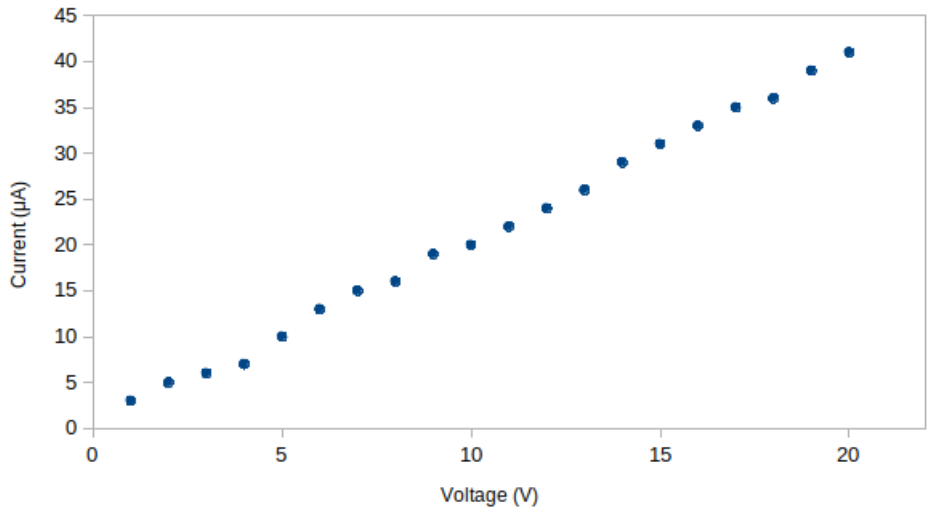


Figure 2.2: Same graph as in Figure (2.1), with some modifications made.

Adding a Trendline

It should appear to you now that this trend is quite linear. We can now use the computer's algorithm to get the "best" straight line between these points. Don't worry right now about how this is done, you will learn about this in some detail at a later point.

Right-clicking the data-points, select `Add Trendline`. In the dialog box that opens, choose a `Linear` regression type and select the checkbox that says `Show Equation`. You can also make the line dotted, and change its colour.

If your trendline equation looks something like this:

```
f(x) = 2.02556390977444 x + 0.231578947368423
```

then edit it to round it off to a more reasonable number of digits. You will understand how many digits is reasonable in the next chapter on error analysis, for right now, let us choose to keep three decimal places. You should now have a graph that looks like Figure (2.3).

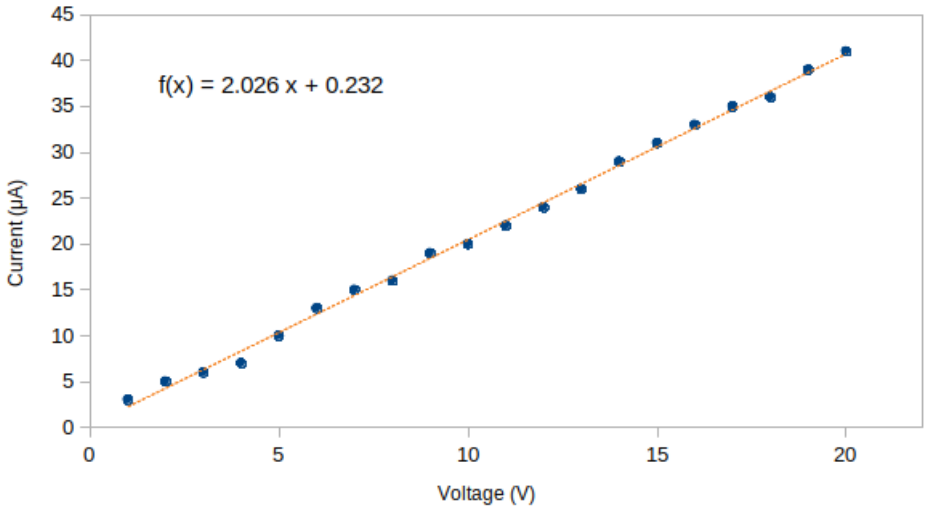


Figure 2.3: Same graph as in Figure (2.1), with a formatted trendline.

Tip

In Microsoft Excel, you can make the current graph style a "template", so that you don't have to keep making most of these changes every time you plot new data. To do this, right-click the chart, and select `Save as Template`. Choose an appropriate name for the chart template and save it. The chart template automatically appears in the `Templates` folder for charts. You can access it on the `All Charts` tab in the `Insert Chart` or `Change Chart Type` dialog box, where you can apply a chart template like any other chart type.

2.4.2 Graphing with Python

Programming goals

1. Getting the packages you need to get started
2. Plotting and formatting your first graph
3. Getting your data into the program
4. Adding and formatting a trendline

Getting the packages you need

Python `packages` are pre-written scripts that you can `call` with your program to help you do whatever it is you want to do. They come with pre-defined functions and values that you can use. For example, imagine you wanted the value of the constant π.

Run the cell below this one. You can run it either by clicking the `Run` button on the top (in the toolbar), or by pressing `Shift + Enter` simultaneously.

[ ]:
```
pi
```

You have now seen your first Python error. Get used to errors; this one will not be your last.

You have asked Python to do something, and it does not understand what it is you have asked of it. As a result, it tells you where *it* thinks there's a problem and what the problem is (it's often right, but not always).

In this case, it says the problem is on line 1 of the cell (`----> 1 pi`) and the error is that it does not understand what you mean by `pi`.

Now compile the cell below:

[ ]:
```
import math
math.pi
```

We have `imported` the `math` package, a collection of standard mathematical values and functions. It contains, among other things, the value of `pi`, which we call by calling `math.pi`.

Some packages have horribly long names, and it would be quite cumbersome to have to type them out everytime, so we give them 'nicknames'.

See if you can understand the following code snippet:

[ ]:
```
import math as m
m.pi
```

The NumPy package

NumPy (pronounced "Num Pie", not "Num Pee") is the fundamental package for scientific computing with Python. It contains almost all scientific functions that you require, and is optimised to be significantly faster than the usual Python functions for scientific operations.

If we wish to call any functions it has, we will be using the `np` nickname. This is not essential, but it is common. Try the following:

[ ]:
```
import numpy as np
np.pi
```

The MatPlotLib package

From the Matplotlib website:

"*Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code*".

We will be using it to plot all our data.

You will notice a strange line underneath this: `%matplotlib inline`

This is magic, don't worry about it. It's a command that needs to be there so that figures can be printed "inline" in Jupyter notebook and be stored in the document.

The SciPy package

SciPy ("Sigh Pie") is *another* library of open-source software for mathematics, science, and engineering. We will be using it for **curve fitting**.

You can now go ahead and import all the above packages using the cell below.

[ ]:
```
import numpy as np                      # Importing the NumPy package

import matplotlib.pyplot as plt         # Importing the Matplotlib package for plotting
                                        # "Magic" to display images inline
%matplotlib inline

import scipy as scp                     # Importing the SciPy package
from scipy.optimize import curve_fit # Importing the curve fitting module from SciPy
```

Plotting and formatting your first graph

Lists

Let's start off immediately with a simple plot. The basic syntax for creating line plots is `plt.plot(x,y)`, where x and y are lists ***of the same length*** that specify the (x, y) pairs that form the line.

We can define a list by simply enclosing its elements within square brackets `[]`, like so:

`listName = [element1, element2, element3]`

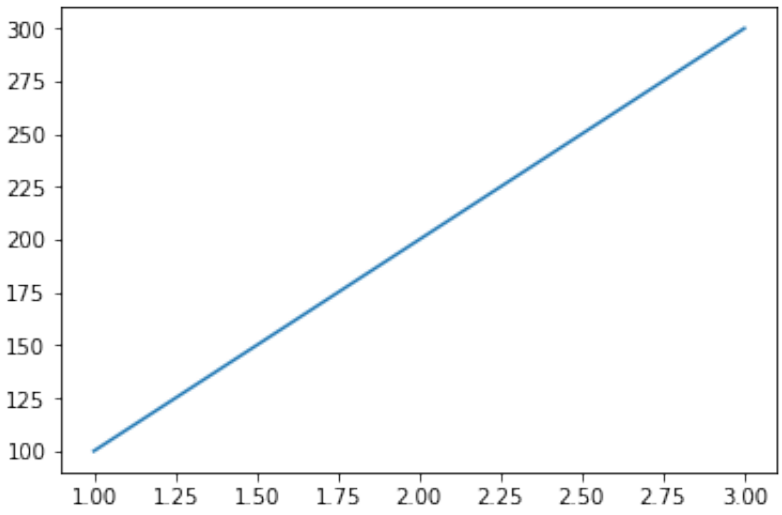
Exercise 1: *** In the following empty cell, define two lists, `xtest` and `ytest`, where:

`xtest` takes values 1, 2, and 3 `ytest` takes values 100, 200, and 300

[ ]:
```
# EXERCISE 1 === Fill in this cell as instructed above

xtest = # <== Insert a three-element list of x-values
ytest = # <== Insert a three-element list of y-values
```

Your first plot

Run the following code snippet: it will create a line plot with the above x and y values you have defined. You should get a graph like this:

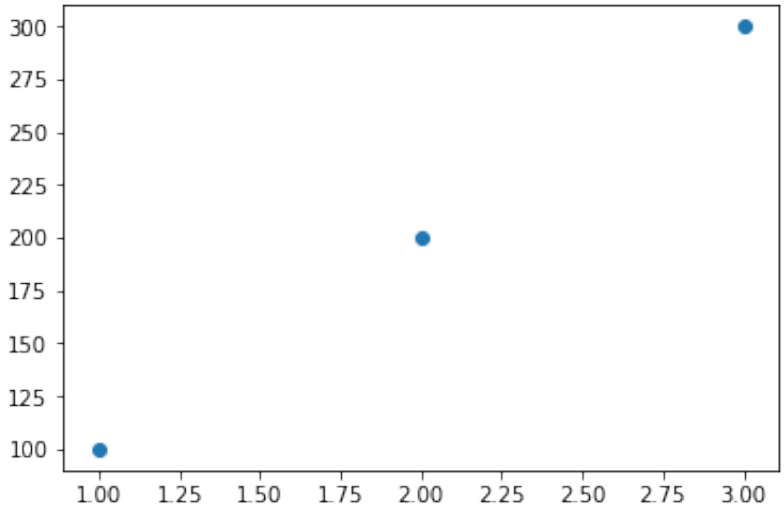


```
[ ]: plt.plot(xtest,ytest)     # Plotting a 'line' plot of ytest vs. xtest


     plt.show()               # This function simply "shows" the graph.
                              # it is not essential to understand how or why it works.
```

There you go, your first graph!

Of course, as mentioned in the previous section on plotting with MS Excel, **you will never plot a line from your data**. Instead, you will more often use the `plt.scatter(x,y)` command.

Exercise 2: *** In the following empty cell, insert one line to plot a `scatter` plot with the same data. After running it, you should get something like this:




```
[ ]: # EXERCISE 2 === Fill in this cell as instructed above

     #<== Insert ONE line here to plot a 'scatter' plot of ytest vs. xtest

     plt.show()       # This function simply "shows" the graph.
                      # it is not essential to understand how or why it works,
```

Formatting your graph

You will notice that it is still incomplete, you will need to label the axes. This can be done by adding modifying the `plt.xlabel` and `plt.ylabel` parameters, and re-plotting as shown below:

```
[ ]: # EXERCISE 2 === Fill in this cell as instructed above

     plt.xlabel('x values (unit)')
     plt.ylabel('y values (unit)')

     #<== Insert ONE line here to plot a 'scatter' plot of ytest vs. xtest

     plt.show()      # This function simply "shows" the graph.
                     # it is not essential to understand how or why it works.
```

You will often have to use scientific symbols in your graphs like λ and θ, and so it is helpful to learn how to do this.

Just type out `plt.xlabel(r'$\lambda$ ($\mu$ m)')` and `plt.ylabel(r'$\theta$')` in the cell below and compile it.

(The `r` present before the single quotes indicates that it is a "raw" string, and the greek letters are called using their LaTeX names. You can find out more **here**.)

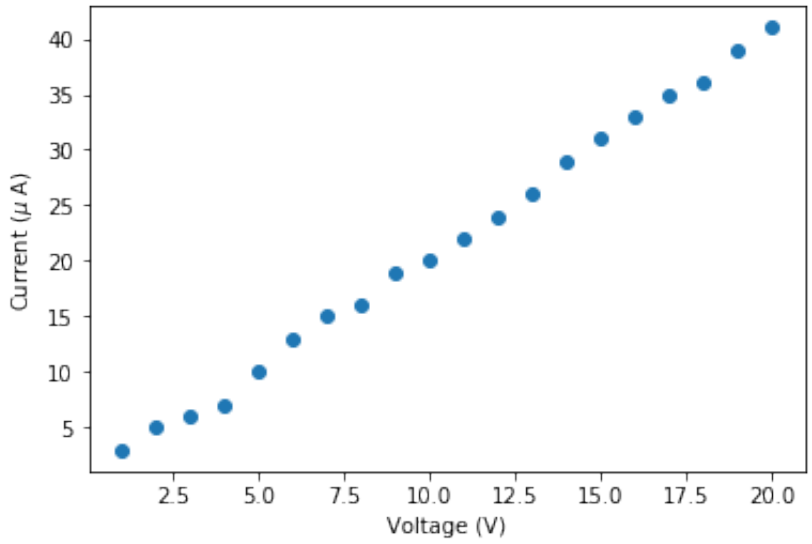
```
[ ]: # EXERCISE 2 === Fill in this cell as instructed above

     # <== Insert ONE line of code to to change the x label of the plot
     # <== Insert ONE line of code to to change the y label of the plot

     # <== Insert ONE line here to plot a 'scatter' plot of ytest vs. xtest
     plt.show()      # This function simply "shows" the graph.
                     # it is not essential to understand how or why it works.
```

What if you had two sets of data to plot for the same x values? Try out the snippet given below:

```
[ ]: xtest  = [1,2,3]           # A three-element list of x-values
     y1test = [100,200,300]     # A three-element list of y-values
     y2test = [50,150,250]
     plt.scatter(xtest,y1test) # Plotting a 'scatter' plot of y1 vs. x
     plt.scatter(xtest,y2test) # Plotting a 'scatter' plot of y2 vs. x
     plt.show()
```

Plotting your data

Exercise 3: *** Let's now create two simple lists of (x, y) data given in this week's handout, and plot a scatter plot of it. You should get something that looks like this:




```
[ ]: # EXERCISE 3 === Fill in this cell as instructed above

     xval = #<== Insert ONE line here with the x-axis data as a simple list
     yval = #<== Insert ONE line here with the y-axis data as a simple list

     plt.scatter(xval,yval)
     plt.xlabel(r'Voltage (V)')
     plt.ylabel(r'Current ($\mu$ A)')
     plt.show()
```

NumPy arrays

The "lists" that you've seen earlier (such as `[a,b,c]`) are those defined by default in Python. It turns out that the numpy package has its *own* version of lists, known as **arrays**. We will be using these arrays

instead of the standard lists, as they have been optimised for scientific operations.

Numpy arrays can simply be created by wrapping the earlier lists with the `np.array` command, i.e. `np.array([list])`.

Exercise 4: *** In the next cell, define `xval` and `yval` as NumPy arrays, with the same values as above. Nothing should change with your output.

[ ]:
```
# EXERCISE 4 === Fill in this cell as instructed above

# Do the same thing as the previous cell, only this time with NumPy arrays.

xval = #<== Insert ONE line here with the x-axis data as a NumPy array
yval = #<== Insert ONE line here with the y-axis data as a NumPy array

plt.scatter(xval,yval)

plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')

plt.show()
```

Importing data

Let's say you've got your data in MS Excel, and you want to use it in Python. If you have very little data, the easiest way to do this is to type it out yourself (as we've done above). However, if you have a large amount of data, it may be better to import a `.csv` file. CSV (or "Comma Separated Value") files can very easily be created from spreadsheets like Excel (go to `File`⟶`Save As`, and select it from the drop-down list).

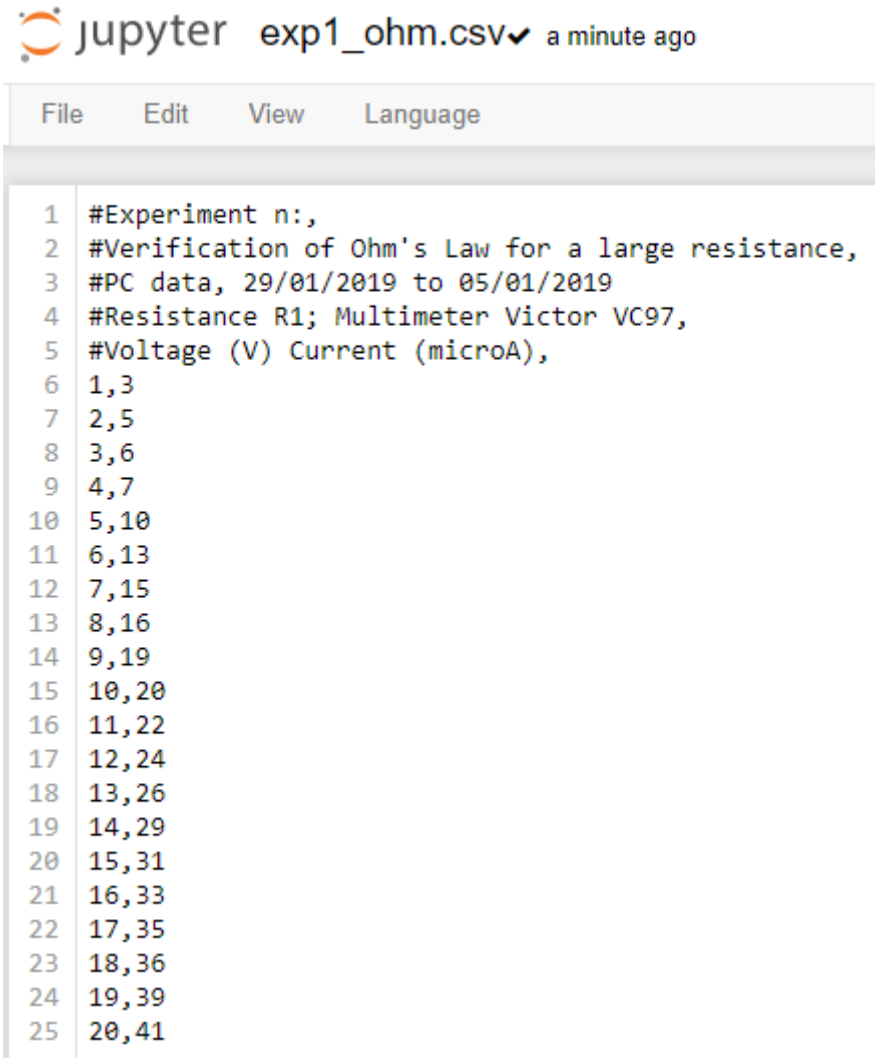
If you are asked to select a delimiter, choose `,`. This does not matter, but let us be consistent. Save the CSV file in the same folder as this Jupyter notebook, and remove all the lines that aren't the actual data.

Tip:

If creating CSV files is too difficult, here's a quick workaround:

1. On Excel, create a column with just your data values, separated by a comma. This can be done easily using the `Concatenate` function in Excel. For example: type `=CONCATENATE(A1,",",B1)` into the cell to merge them, separated by a comma.
2. Now copy the range you're interested in, and paste it in a `Something.txt` or `Something.csv` file and save it in the "data" folder that came along with this Notebook. ***

Sample data file In the "data" folder, you will see there is already a file called `exp1_ohm.csv` with sample data. You can access it in the `data` folder. In Jupyter, go to the `File` tab, and click `Open`. Navigate to the `data` folder and open `exp1_ohm.csv`. You will see something that looks like this:



You will notice that in each row, elements of the two columns are separated by a comma, and that the first rows with the names of these columns and other details have a `#` symbol before them. This is very important! When you import the data into Python, this `#` at the beginning tells the function that's importing the data to ignore these line. (Otherwise, these lines would be a string of letters, and other lines would be numbers, and this wouldn't work). Put the `#` symbol in front of any lines you'd like to ignore while importing.

You are now ready to import this data using a simple `loadtxt` function that's present in the `numpy` package. We're going to learn how to do this in a slightly roundabout way:

1. Call the `loadtxt` function of the `numpy` package. You will see that it throws an error, since you have not given it the name of the file to load.
2. Now, you will start writing out the name as a string. Write out the following first `"./"` and then press the `TAB` button on your keyboard. You will see that Jupyter will show you the files in the current working directory (where your code is stored)! Thus, the "current working directory" is called `..`
3. You need to get to the `data` folder which is one folder up. The standard way to get to the folder above it by using `../`. Type this out, and press the `TAB` button.
4. Select the `data` folder, and then press the `TAB` button again. Now select `exp1_ohm.csv`.
5. Running this will *still* cause an error, since you haven't told Python that the different columns are separated by `,`s. Add a comma after the string, and type out `delimiter = ","`. This tells Python that the data in each row is separated ("delimited") by commas.

Tip:

Using the TAB key to autocomplete commands is a fantastic trick, useful not only here but also on most UNIX-type machine terminals (like Macs). ***

Exercise 5: ***

Complete the steps detailed above and run them in the empty cell below, you should get an output that looks like this:

```
array([[ 1.,  3.],        [ 2.,  5.],        [ 3.,  6.],        [ 4.,  7.],        [
5., 10.],        [ 6., 13.],        [ 7., 15.],        [ 8., 16.],        [ 9., 19.],
[10., 20.],        [11., 22.],        [12., 24.],        [13., 26.],        [14.,
29.],        [15., 31.],        [16., 33.],        [17., 35.],        [18., 36.],
[19., 39.],        [20., 41.]])
```

This is a two-dimensional array (very much like a matrix). It is useful to learn how to manipulate them, and I've added a short optional section about them at the bottom. But for now, we won't really use them since there's a much simpler way to "unpack" your data directly in a format you can use.

[ ]:
```
# EXERCISE 5 === Fill in this cell as instructed above

#<== Insert ONE line here to load the file exp1_ohm.csv from the data folder
```

Unpacking data Wouldn't it be nice if Python understood that the first column was one variable's data and the second another variable's data? It turns out that the `NumPy` package allows for just this, using another option called `unpack`.

Using this option, you can equate **any number** of variables to this data, and Python will automatically send the first column to the first variable, the second to the second, and so on. Thus, if you did:

`x1,x2,x3,...,xn = np.loadtxt(...,unpack=True)`

to a file that contained `n` columns, the first column would be stored in `x1`, the second in `x2` and so on!

Exercise 6: *** In the cell below, write down the code to take the first column as `xpoints`, and the second column as `ypoints`.

[ ]:
```
# EXERCISE 6 === Fill in this cell as instructed above

#<== Insert ONE line here to load the file exp1_ohm.csv from the data folder,
# and unpacking by column int two variables `xpoints` and `ypoints`
```

Adding a Trendline

Adding a trendline in Python is not as obvious for the simple linear graphs as it was in Microsoft Excel, but provides much more functionality. The idea we will use is the following:

1. We will define a function (say) `f` which accepts some parameters (`x,a,b,c...`) (how many depends on what type of function we're fitting) and `returns` a value.
2. We will then call `scipy.optimize`'s `curve_fit` function with `f` and our data (`xpoints` and `ypoints`) which will automatically vary the parameters to give us the best values for `a,b,c...`

This is shown in the following snippet:

[ ]:
```
def f(x, a, b):          # Define a function `f` which `returns` a value of a*x+b
    return a*x + b       # This line makes sure that the function returns the above␣
 ↪value

par, covariance = curve_fit(f, xpoints, ypoints)

print("Variable par is this array:",par)
```

```
print("")
print("The slope is:",par[0], " and the intercept is: ",par[1])

m = np.round(par[0],3)  # We use the `numpy.round` function to round to 3 decimal␣
 ↪places
c = np.round(par[1],3)
```

The line `par, cov = curve_fit(f, xpoints, ypoints)` might require some explanation: the `curve_fit` function returns **two values** by default, an array with the *parameters*, which we have called `par` here, and an array with the *covariance* which is a statistical concept that you do not need to interest yourself with now.

All you need to know is that `par[0]` is the value of `a` (the slope of your line) and `par[1]` is the value of `b` (its intercept).

Let's use this now to create an array of "theoretical" y values, using the formula $y(x) = mx + c$ – a straight line.

[ ]:
```
ytrend = m*xpoints+c      # Creates an array of y values corresponding to the␣
 ↪xpoints,
                          # which satisfy the trendline given by the parameters in␣
 ↪[par]

plt.plot(xpoints, ytrend, '--',color="red") # Plot a red dashed line of ytrend vs.␣
 ↪xpoints
plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')
plt.show()
```

Now, we can simply *combine* the above two graphs, as shown below.

You can also add an equation using the `plt.text` command, which allows you to add a string of text (here, the variable `eqn`) at specified coordinates (in our case, $x = 4, y = 14$), with font size 12pt. If you're plotting a straight line, you don't need to change `eqn`, it takes the value of the slope and intercept defined earlier, converts them into strings and adds them to the graph. (Placing `{c:+}` simply states that if `c` is a positive number, it is made into a string `"+c"` instead of just `"c"`(see **here** for more).

[ ]:
```
plt.scatter(xpoints,ypoints)                        # Plotting the data-points
plt.plot(xpoints, ytrend, '--',color="red")         # Plotting the trend-line

slope_string = round(m,3)                           # Rounding off the slope to 3␣
 ↪digits
intercept_string= round(c,3)                        # Rounding off the intercept to 3␣
 ↪digits

eqn = 'y(x) = '+f'{slope_string}'+'x'+f'{intercept_string:+}'  # The equation string

plt.text(4, 14,eqn,fontsize=12)                     # Displaying the above string


plt.xlabel(r'Voltage (V)')                          # Formatting the axes
plt.ylabel(r'Current ($\mu$ A)')
plt.show();
```

FIN

Slightly more advanced topics

Two-dimensional arrays *(optional)* Sometimes you might want to include your data as an array of (x, y) points, for example something that looks like:

```
z = [[ 1.  3.],[ 2.  5.],[ 3.  6.],[ 4.  7.],[ 5. 10.],[ 6. 13.],[ 7. 15.],[ 8. 16.],[
9. 19.],[10. 20.],[11. 22.],[12. 24.],[13. 26.],[14. 29.],[15. 31.],[16. 33.],[17.
35.],[18. 36.],[19. 39.],[20. 41.]]
```

In this case, to plot it, you would first have to take all the xs, and then all the ys. We do this using the highly useful : symbol as follows:

[ ]:
```
z = np.loadtxt("../data/exp1_ohm.csv",delimiter=",")

print("The 2D array z:")
print(z)

xpoints = z[:,0]

print()
print("The slice of x values:")
print(xpoints)
```

The : symbol here acts like a **wildcard**. Suppose our array z looks as follows:

```
[[ 1.  3.]  [ 2.  5.]  [ 3.  6.]  [ 4.  7.]  [ 5. 10.]  [ 6. 13.]  [ 7. 15.]  [ 8.
16.]  [ 9. 19.]  [10. 20.]  [11. 22.]  [12. 24.]  [13. 26.]  [14. 29.]  [15. 31.]
[16. 33.]  [17. 35.]  [18. 36.]  [19. 39.]  [20. 41.]]
```

The : (called the *slicing* operator) gives us a 'slice' of the data: **all values in the column on the left**. (This is Column "0". Column "1" would be the one on the right.)

Exercise 7: ***

Complete the following cell: Write down a line of code to collect the y points and create an array `ypoints`.

[ ]:
```
# EXERCISE 7 === Fill in this cell as instructed above

#<== Insert ONE line of code here to get the data as a 2D array z

xpoints = z[:,0]
#<== Insert ONE line of code here to get the ypoints data
```

For those interested in more advanced fits *(optional)* In order to plot a polynomial trendline, we will evaluate this (polynomial) function at the different `xpoints` and plot a dotted line over the earlier graph.

In order to do this, we will be using the `numpy.polyval` function which accepts two variables – an array of coefficients ($[a_n, a_{n-1}, ..., a_2, a_1, a_0]$) and a value of x – and prints out $y(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_2 x^2 + a_1 x + a_0$.

Thus, calling `np.polyval([a,b],x)` would give you $y(x) = ax + b$. If we used an *array* of x points, we'd get an *array* of y values!

We'll try to plot the line $y(x) = -x + 4$ to see how it works. The parameters will be `[-1,4]`. The resulting array will contain the values of y at these points.

[ ]:
```
xvalues = [0,1,2,3,4,5]               # A sample array of x values
yvalues = np.polyval([-1,4],xvalues) # y values corresponding to -1*x + 4

print(yvalues)                        # Printing out the yvalue array
```

```
plt.plot(xvalues,yvalues,'--')        # Plotting yvalues vs. xvalues in a line plot,
                                      # with a dashed line (given by '--')
```

Using np.genfromtxt ***(very optional)*** Apart from the simple `loadtxt` function, `NumPy` also has a slightly more powerful `genfromtxt` function which allows you to deal with `.csv` files that have missing values and so on. However, at your level, I think there is no difference at all between the two, and `loadtxt` is easier to remember!

You can use the `unpack` function here as well, but for illustrative purposes, I have used the slicing operator `:`.

[ ]:
```
data = np.genfromtxt("../data/exp1_ohm.csv", delimiter=",") # Looks in the current␣
↪directory
                                                            # for the csv (or txt) file and
                                                            # imports it, with commas being
                                                            # treated as delimiters.

xpoints = data[:,0]                                         # Get the first column of data,
                                                            # saving it to xpoints

ypoints = data[:,1]                                         # Idem for second column and␣
↪ypoints.

plt.scatter(xpoints,ypoints)
plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')
plt.show()
```

Session 3

Sensitivity to Errors in Measurement and Preliminary Error Analysis

3.1 Objectives

1. To recognise different kinds of errors, in particular systematic errors, errors due to the least count of an instrument, random errors, and errors in time measurement.
2. To learn to use Vernier calipers and the screw gauge.
3. To begin to understand propagation of errors.
4. To use preemptive error analysis to plan an experiment.

3.2 Introduction

When observing the time period of the simple pendulum, you already became aware of the extent to which you are certain of your time measurement. In fact, all the measurements you made – of mass, of length, of time – were uncertain. In experimental physics it is of enormous importance to understand how different kinds of errors arise, because without an understanding of errors it is difficult to make sense of data. This is a vast and subtle subject, and so you will learn the elements of it slowly.

It is important to appreciate that in physics *an error is not a mistake.* An error in physics is a measure of how certain (and therefore how uncertain!) we are of a measurement. All measurements are to a greater or smaller extent uncertain; there is no such thing as a perfect measurement:

therefore, there is no such thing as an error-free measurement (unless it is something like the **number of oscillations or rotations**).

To broaden the scope or your understanding of errors, you will use not only the data you got while observing the oscillations of a pendulum, you will also learn to use two very useful instruments, the Vernier calipers and the screw gauge, which highlight beautifully certain kinds of errors.

Tip

Example: if you use a metre scale with a least count (smallest division) of 0.1 cm upside down to measure a pencil and misread the scale as 94.6 cm instead of 5.4 cm, this is a **mistake**.

However, if you note down the length as being 5.4 cm, but rightly note that it is not **exactly** 5.4 cm, but simply that your measuring device does not allow for any more precision, this is a measure of **uncertainty**.

We will expect you to have verified that you haven't done the former, and will take for granted from here on that no mistakes were made in the collection of data.

The complete statement of a measured value *must* include an estimate of the level of confidence associated with the value. This allows people to judge the quality of the experiment and allows for comparisons with other similar estimates of the same quantity, or with theoretical predictions.

A proper experiment must report both a "best" value and an uncertainty for each measured quantity. You will be expected to do this in all your experiments.

Without an uncertainty estimate, it is impossible to answer the basic scientific question: Does my result agree with a theoretical prediction or results from other experiments? This question is fundamental for deciding if a scientific hypothesis is corroborated or refuted.

3.3 Systematic Errors

Suppose that you are weighing yourself and the pointer is at 1 kg even when no-one is standing on the scales. Then, obviously, all measurements will differ from the correct weights by 1 kg. Such an error is called systematic, since it appears systematically in all readings. The only way to eliminate a systematic error is to identify its cause and eliminate it. Some devices allow such checking to be done relatively easily, e.g. a metre scale or a weighing scale, but many others do not, e.g. electrical devices with internal errors. Estimating possible errors due to such systematic effects will depend on your understanding of your apparatus and the skill you have developed for thinking about possible problems. However, if you get a value for some quantity that seems rather far off what you expect, you should think about the possibility of systematic error more carefully. You could end up trusting a device that you do not know is faulty. This happens very often, to the best of us.

Question: You use a ruler whose end is worn out, so that it effectively begins at 2 mm. How can you avoid the systematic error that would arise if you used the ruler naively?

Question: A systematic error can be either positive or negative. What is the difference and under what circumstances would they arise? (If you don't understand this point, you may end up doubling a systematic error in trying to eliminate it.)

3.3.1 Least-count-related Errors

When you make a length measurement with a ruler whose smallest division is 1 mm in size. If you imagine aligning an object of a certain length against a ruler, you can ensure that one end of the object is flush with a mark on the ruler but you cannot ensure that the other end is also flush with another mark – it general it will be somewhere between two successive marks, and you cannot tell exactly where it is; normally the best that you can do is say, "It's more than halfway between", or "It's less than halfway between". So the error associated with the least count of the instrument is normally between half and one full least count.

Because the least count of an instrument is a limitation, one can try to subdivide the region between two marks. It is obvious, however, that this is not going to work beyond a point, since we will simply not be able to tell where the end of the object lies.

A couple of clever designs *effectively* sub-divide the region. The most beautiful of these designs was first included in an instrument called Vernier calipers; another design is found in the screw gauge. You will learn how to use both of these instruments. (The essential design of the Vernier calipers is incorporated in many other instruments, and we then speak of it as having a Vernier scale.)

3.4 Vernier Calipers and the Screw Gauge

3.4.1 Vernier Calipers

You will regularly come across Vernier scales in the lab when using calipers, angular Vernier scales (in spectrometers), and travelling microscopes. Vernier scales allow you to read off a value more precisely than when using an ordinary scale. In this section, we will explain how this works.

First consider a "main scale" with a least count of 1 unit (you could imagine this is 1 mm, if you wish). This implies that any distance between, say, the 2 and 3 unit marks cannot be determined accurately. In other words, the best you could say is that an object is "2 and a bit" units. The Vernier scale allows you to **quantify** this "bit" to some extent.

The method is ingenious: instead of measuring this "bit" directly, its magnitude is translated into a a degree of coincidence between two scales, the main scale and a secondary one called the Vernier

scale. The Vernier scale has divisions that are **slightly** smaller than that of the main scale, such that n divisions on the Vernier scale have the same length as $n-1$ divisions of the main scale. For example, 9 divisions on the main scale may coincide with 10 divisions on the Vernier scale (see Figure (3.1)).

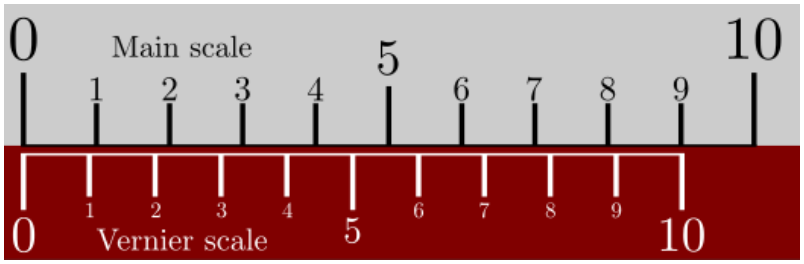
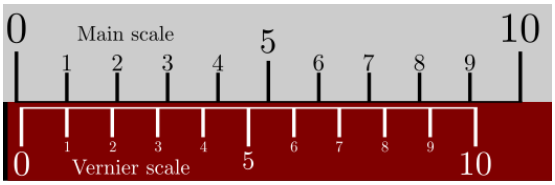


Figure 3.1: 10 Vernier scale divisions are set to coincide with 9 main scale divisions.

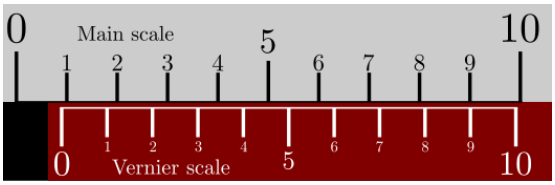
Question: What is the spacing between two Vernier scale divisions? Is it:

1. 1 `MSD` unit?
2. 0.1 `MSD` units?
3. 0.9 `MSD` units?

Let us now try to measure the smallest possible distance between 0 and 1 unit. Take a set of Vernier calipers and close the jaws completely. If there is no zero-error, the only two readings on the Vernier scale which match the main scale are 0 and 10, as shown in Figure (3.1).



(a) Just away from 0, when the coinciding division is 1, the distance is $1 \times \texttt{MSD} - 1 \times \texttt{VSD} = 0.1\,\texttt{units}$.



(b) Before crossing 1, the last coinciding division is 9, and the distance is $9 \times \texttt{MSD} - 9 \times \texttt{VSD} = 0.9\,\texttt{units}$.

Figure 3.2: Measurement with Vernier Calipers

1. Move the Vernier scale *slightly*, until the 1 on the Vernier scale coincides with the nearest main scale division (this is obviously the 1 on the main scale, see Figure (3.2a)).
2. You will notice that the jaws are slightly apart. The distance between them is of course the distance the 0 on the Vernier scale has moved. This is the least distance you can measure with this Vernier calliper.
3. But this distance is simply the difference between 1 Main Scale Division (`MSD`) and 1 Vernier Scale Division (`VSD`), since both the 1 marks coincide!

4. Thus, the spacing between the jaws of the calipers is now 1 MSD − 0.9 MSD = 0.1 MSD units!

5. We have thus been able to measure a distance of 0.1 units using two scales, one of least count 1 unit, and the other of least count 0.9 MSD units!

Similarly, if you went one division further, and had the 2 of the Vernier scale coincide with a main scale division, then the distance between the jaws would be $2 \times \texttt{MSD} - 2 \times \texttt{VSD} = 0.2$ MSD units. Thus, if the nth Vernier scale division coincides with a main scale division, the distance between the jaws is $n \times (\texttt{MSD} - \texttt{VSD}) = n \times \texttt{LC}$. Where LC is the Least Count of your Vernier calipers.[1]

Of course, up until right now we have been measuring distances between 0 and 1 unit. What about an arbitrary distance? Consider the example given in Figure (3.3).

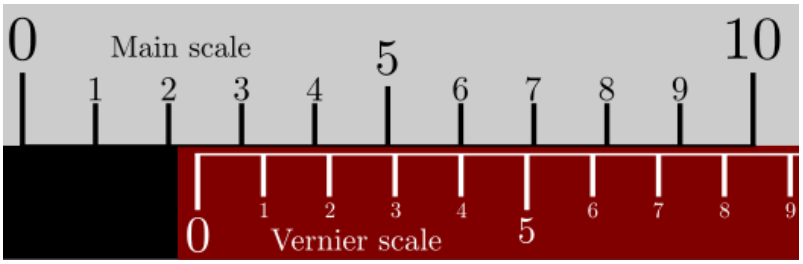


Figure 3.3: The main scale reading is more than 2 units. The coinciding Vernier scale division is 4 (the fact the main-scale reading with which it coincides is 6 is not important).

Here is the general procedure to take a reading using Vernier calipers:

1. Look for the main scale reading to the left of the the 0 of the Vernier scale, this is the Main Scale Reading or MSR; the total reading must be this much plus the extra bit between it and the 0 of the Vernier scale. In the example in Figure (3.3), the MSR is 2.

2. Now look to find the mark on the Vernier scale which most closely meets any mark on the main scale. This is the Vernier Scale Division or VSD, giving you the most precise digit. In this example, the VSD is 4 because the mark at the end of the 4th division on the Vernier scale coincides with a mark on the main scale (it does not matter which mark on the main scale coincides). If you are still confused, make a movie in your head in which you start with the 0 of the Vernier scale coinciding with a main scale division. Now, slowly, move the 0 of the Vernier scale: First, when the spacing between it and the last main scale division is 1 LC, the first VSD will coincide with a main scale division, then, when that spacing is 2 LC, the *second* VSD will coincide with *another* main scale division, and so on.

3. The total distance is given by

$$\texttt{distance} = \quad \texttt{MSR} + (\texttt{VSD} \times \texttt{LC}) \quad = 0.2 \quad + (4 \times 0.01) = 0.24$$

   In our case, distance = 0.24 MSD units.

[1] In our case, the Least Count is 0.1 MSD units.

4. Thus, using two scales – one of least counts 1 unit and another of least count 0.9 units – we have calipers capable of measuring up to (or down to!) 0.1 units! You have an *effective least count* that is much smaller than the least count of the main scale.

Question: Show that the least count of a set of Vernier calipers is given by:

1. $$\texttt{LC} = \texttt{Least count of main scale} - \texttt{Least count of Vernier scale}$$

2. $$\texttt{LC} = \frac{\texttt{Least count of main scale}}{\texttt{Number of Vernier scale divisions}}$$

3.4.2 The Micrometer Screw Gauge

The design of the micrometer screw gauge is different, and uses the fact that a linear movement that is imperceptible can be turned into an angular movement that is easily perceptible; the best way to understand this is simply to examine the screw gauge (which, unlike Vernier calipers, is easy to understand). The screw gauge uses a screw with an accurate and constant **pitch** (the amount by which the thimble moves forward or backward for one complete revolution) as an auxiliary scale marked on a rotatable thimble.

The micrometers in our laboratory have a pitch of 0.5 mm. The rotating thimble is subdivided into 50 equal divisions. The thimble passes through a frame that carries a millimetre scale graduated to 0.5 mm. The jaws can be adjusted by rotating the thimble using the small ratchet knob. This includes a friction clutch which prevents too much tension from being applied.

Only tighten the screw gauge by rotating the ratchet, otherwise you may damage the instrument. Stop rotating after you hear **three** clicks. **Do not** tighten the screw any further.

Question: Show that the least count of a screw gauge is given by

$$\texttt{LC} = \frac{\texttt{Pitch of a the screw}}{\texttt{Number of circular scale divisions}}$$

3.5 Random Errors

Let us go back the simple measurement of the length of an object with a ruler. We place the object against the ruler with one end flush against a certain division of the ruler, say 0 (it does not have to be 0, but usually is). Then we observe where the other end is. In general, it will be, as

mentioned earlier, at some point between two successive marks on the ruler. Let us imagine doing this repeatedly. Every time we place one end against 0, it will be at *slightly* different point, and so the other end will also move slightly, in a manner that we cannot control. If this movement is much smaller than the distance between two successive marks on the ruler, i.e. its least count, then this random movement will not be discernible. On the other hand, if the movement is greater than the least count of the ruler, then the length we measure will change from one measurement to the next. In the first case the error is determined by the least count of the ruler, and the length measured will not change from one measurement to the next. In the second case, however, it is clear that the length measured will change from one measurement to the next. We call this kind of error a *random error*.

Another example that may make the idea clearer is the following. Imagine that a sound of a fixed duration is played over and over again, and you are required to time the duration with a stopwatch. You will have to start and stop your stopwatch over and over again. It is easy to see that even though the duration of the sound remains the same, the duration that you measure will vary from one measurement to the next. Once again we have a random error. It should be intuitively clear that the magnitude of the error depends not on the duration of the sound but on the *spread* of your measurements. The number we assign to the random error must therefore be obtained from a measure of this spread. To do this, we must remove what does not matter, namely the actual duration of the sound. Generally we don't know the actual duration of the sound. So we do the best we can: we subtract the mean of our measurements from all of them, since the mean is our best guess of the actual duration. That leaves us a distribution of the variations of the measurements.

We are not really concerned with whether the variation is positive or negative. So we use the sum of the squares of the variation, i.e. $\sum_{i=1}^{n}(x_i - \overline{x})^2$, where x_i is any reading, $\overline{x}$ is the mean, and n is the total number of readings. Because the variations are both positive and negative, this sum grows not in proportion to n but in proportion to $\sqrt{n}$. (This is a subtle point; you will understand it later.) So the standard measure of the spread of the readings is the standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}}.$$

(Why $n-1$ instead of n in the denominator is another subtle point. Again, let it be for the moment. Anyway, the difference between n and $n-1$ is negligible for large n, which is the only case in which the idea of random errors is meaningful.)

3.6 Errors in Time Measurements

The time measurements you will encounter in the lab will be made by human beings (you) using clocks. If you compare the processes of measuring a duration to that of measuring a length, you will immediately see that you no longer have the luxury of aligning one end of your "object" (here a duration) with a physical scale: you must start your clock at a moment that you best judge to be the beginning of the duration and end it at what you best judge to be its end. Thus you automatically make two errors, one at the end of your time measurement and one at the beginning. The magnitude of the error has nothing to do with how long or short the duration is.

Is this error due to the least count of you clock? It is – partly; but that's not the whole story: there is another factor to take into account, your reaction time. As a human being you have to judge when when the duration begins and press the start button on your clock; both your judgement and the pressing of the button entail errors that are due to the finite time it takes for your brain and your body to react. Thus the errors at the beginning and the end of a measurement of duration are a combination of the the error due to the least count of your clock and your reaction time. If one is much larger than the other, it dominates and the other may be neglected.

Determining the least-count error is easy: it's just the last count of your clock. Determining the error due to finite reaction time is not easy. One way is to do many measurements of a duration that is known to be constant (e.g. the period of a pendulum) and look not at the average but the spread in the readings, a measure of which is the standard deviation. For most human beings have a reaction time of about 0.1 s, much larger than the least count of standard lab stop watches, which is usually 0.01 s. Thus the error in most of your time measurements will be dominated by your reaction time.

3.6.1 Time Measurement for Periodic Phenomena

You will often be interested in measuring a duration of a periodic phenomenon, e.g. the oscillation of a pendulum. If you assume that the period of the pendulum has a true value that does not change from one oscillation to the next, then there is a way in which you can dramatically reduce the error on the time period. Since the error in the measurement of a duration occurs only at the ends, you can simply measure the duration for many oscillations, e.g. 10 or 100, and divide by that number. So, if the total error due to your reaction time on a duration (i.e at both ends of a duration together) is 0.1 s and you measure the duration for 10 oscillations, the error associated with a single time period will be $0.1/10 = 0.01s$.

Given that this is true, you may ask why one doesn't simply measure the duration for 1000 oscillations (or some arbitrarily large number). One answer is that the oscillations will damp out. The other had to do with human fatigue: if you are asked to count 1000 oscillations, you may make an error in counting the number of oscillations, and that would lead to a much larger loss in precision than what you gained by increasing the number of oscillations. (Machines don't have this problem and it makes sense, when a mechanical counter is being used, to increase the number of oscillations as much as possible.)

A question that might arise is what happens if we measure the duration for 10 oscillations several times. Is this equivalent to measuring the duration for 100 oscillations? No. Now we have a pair of end points for each duration of 10 oscillations, and each point is associated with starting and stopping a clock. If we measure the duration for 100 oscillations (let us call this case 1), we have 2 end points, and if we measure 10 sets of 10 oscillations each (case 2), we have 20 end points, i.e. there are 10 times as many points at which errors can occur. From this we may at first conclude that the error in case 2 is 10 times the error is case 1, but that is not correct: the errors we make at the end points can be either overestimates of underestimates, i.e. positive or negative, and so as we stack up the end points there will be some cancellation. It an be shown – and at some point you will show this – that as the number of end points n increases, the error goes up as $\sqrt{n}$. Thus, if the error on time measurement with single pair of end points is Δt, the error on combining n time

measurements is $\sqrt{n}\,\Delta t$. Putting all this together, if the error on a single time measurement is Δt, the error on the time period T using 100 oscillations is $\frac{\Delta t}{n}$ and the error on T using 10 measurements of 10 oscillations each is $\frac{\sqrt{n}\,\Delta t}{n} = \frac{\Delta t}{\sqrt{n}}$. (Don't worry if this paragraph went over your head; you will understand it slowly, over the semester.)

It should be obvious from the discussion above that while time measurement is much more subtle than length measurement, it allows an almost arbitrary improvement in precision (reduction in error) on the time period itself (to be distinguished from the duration of a measurement) if the phenomenon is periodic. This can be used to great effect in certain devices, e.g. the Kater's pendulum. To ensure that this great advantage is not vitiated by incorrect counting of oscillations, it is very important to know exactly how to count the number of oscillations. A very common error is to count $n-1$ oscillations when one wants to count n. You will learn the correct procedure in the lab.

Tip

A cautionary tale: Neutrinos are weakly interacting nearly massless particles that are theorised to be moving close to – if not at – the speed of light.

In September 2011, scientists of the OPERA collaboration reported evidence that neutrinos they produced at CERN in Geneva and recorded at the OPERA detector at Gran Sasso, Italy, had travelled **faster** than light. If this were true, it would have violated one of the pillars of modern physics, the Theory of Relativity.

The neutrinos were calculated to have arrived approximately 60.7 nanoseconds sooner than light would have if it had traversed the same distance in a vacuum. Assuming the error were entirely due to random effects, scientists calculated that there was a 0.2-in-a-million chance that this may be a false positive.

In March 2012 it was confirmed that a fibre cable was not fully screwed in during data gathering, and this effectively decreased the reported flight time of the neutrinos by 73ns, making them seem faster than light. The corrected difference between the measured and expected arrival time of neutrinos (compared to the speed of light) was approximately 6.5 ± 15 ns. This is consistent with **no difference at all**, thus the speed of neutrinos is consistent with the speed of light within the margin of error.

Precision, Accuracy, and Statistics

1. **Accuracy:** is how closely a measurement comes to some "true" value. It determines how well we have eliminated ***systematic errors*** and mistakes in our measurements.
2. **Precision:** is how closely a set of measurements agree *with each other*.

To use a sports simile, if you're playing football and are accurate, your aim will always take the ball close to or into the goal. If, however, you harbour a personal grudge against the coach and succeed in beaning him on the head every time you get the ball, then you would be *precise*, but not

accurate.[2]

A caveat: In any physics experiment worth doing, you usually don't have the answer beforehand. This means you don't have a "true" value with which to compare your answer. This is why it is so important to be both precise (reduce all random errors present in your measurements) and accurate (reduce all systematic errors present in your measurement). You must also, however, come up with ways to check if your answer is accurate *without* trying to resort to a "textbook" solution.

Question: In the example of the "superluminal" neutrinos given earlier, was the error of the OPERA team one of inaccuracy, or of imprecision?

3.7 Propagation of Errors

In general we are interested not just in our observations but in quantities derived from them, i.e. in derived quantities. It is obvious that the a derived quantity will inherit some the errors in the observations from which it is derived. But to what extent will it do so? We will use a few examples to understand this.

1. We want the radius of a diameter, we usually measure its diameter. Here the observed quantity is the diameter d and the derived quantity is the the radius r. We have $r = d/2$. Thus if there is an error of $\pm\delta$ in measuring d (for example due to the least count of the measuring instrument), then, clearly the error in r is $\pm\delta/2$.

2. We want the perimeter p of a rectangle having measured its length l and its width w. $p = 2l + 2w$. If the errors in measuring l and w are $\pm\delta$, for each, the error in p is clearly $\pm 2\delta \pm 2\delta$, which ranges between -4δ and $+4\delta$, depending on the signs on the errors on l and w. So the error on p is $\pm 4\delta$.

3. We want the area A of a rectangle having measured its length l and its width w. $A_{\text{true}} = l\,w$. Suppose that the error in the measurement of l is $\pm\delta l$ and the error in measuring w is δw, then what we derive is $A_{\text{derived}} = (l \pm \delta l)(w \pm \delta w)$. If the errors are much smaller than measurements, we can neglect terms of order δ^2. Then, the value of A_{derived} ranges between $A_{\text{true}} - w\delta l - l\delta w$ and $A_{\text{true}} + w\delta l + l\delta w$. Notice that the expression for the error is rather complicated; we can simplify it if we express it in terms of the relative error. Since the error in A is $\delta A = A_{\text{derived}} - A_{\text{true}}$, you can easily show that

$$\frac{\delta A}{A} = \pm\frac{\delta l}{l} \pm \frac{\delta w}{w}.$$

   The quantity $\delta x/x$ is called the *relative error* in x.

[2] It is widely accepted that the point of football is to get the ball into the goal and not to traumatise your coach, which is why this example works.

4. We want the volume V of a sphere given a measurement of its diameter d. You can easily show, using the method given above, that

$$\frac{\delta V}{V} = \pm 3 \frac{\delta d}{d}.$$

Notice how the sensitivity of the relative error in V to the relative error in d depends on the *power* in the dependence of V on d. (Since $V = (4/3)\pi r^3$, $V \propto d^3$.)

5. We want the mass per unit length λ of a wire having measured its mass m and its length l, with . $\lambda_{\text{true}} = (m \pm \delta m)/(l \pm \delta l)$. If you go through the same process as in the last case considered, you will find that

$$\frac{\delta \lambda}{\lambda} = \pm \frac{\delta m}{m} \pm \left(-\frac{\delta l}{l} \right).$$

3.8 Using Error Analysis to Plan Your Experiment

Suppose that you want to measure the volume of a tile (whose length l and width w are much greater than its thickness t). Since $V = l \times w \times t$, you have

$$\frac{\delta V}{V} = \pm \frac{\delta l}{l} \pm \frac{\delta w}{w} \pm \frac{\delta h}{h}.$$

This formula shows you right away that there is not much point improving one measurement indefinitely if the dominant cause of error is another measurement. What we want is for all the measurements to contribute more or less equally. This immediately tells us that we must choose our instruments and number of measurements so that this goal is achieved.

Question: What instruments would you choose measure the length, width, and thickness of the tile so that the relative errors in them are comparable?

Question: Suppose $P = xy^3$, where x and y are the measured quantities. On which of them is the relative error in P more sensitive?

3.9 The Experiment

3.9.1 Apparatus

1. A pair of Vernier calipers
2. A Screw gauge
3. A set of ball bearings
4. A set of aluminium cylinders
5. A small measuring flask
6. A container with a spout

3.9.2 Description

1. In **Part A** you will measure the radius, height, and volume of different cylinders. You will then plot an appropriate graph from which you will extract the value of π.
2. In **Part B** you will measure the radius and volume of different ball bearings. You will then plot an appropriate graph from which you will extract the value of π.

3.9.3 Suggested Procedure

Part A

1. Decide on the best instrument and method to measure the following properties of the cylinders:
   (a) Radius
   (b) Height
   (c) Volume

   Question: How and where would you measure the radius of the cylinders? How many trials would you take?
2. Be sure to take a sufficient number of **well-sampled** readings.
3. Decide on an appropriate graph between the measured quantities from which you can extract the numerical value of π.

Part B

1. Decide on the best instrument and method to measure the radii and the volumes of the bearings.
2. Be sure to take a sufficient number of **well-sampled** readings.
3. Decide on an appropriate graph between the measured quantities to extract the numerical value of π.

Question: How does your estimate compare with the known value of $\pi \approx 3.14159$?

Part II

Experiments

Experiment 1

The Soft Massive Spring

Objectives

1. To determine the spring constant and the mass correction factor for the given soft massive spring by static (equilibrium extension) method.
2. To determine the spring constant and the mass correction factor for the given soft massive spring by dynamic (spring mass oscillations) method.
3. To determine the frequency of oscillations of the spring with one end fixed and the other end free i.e. zero mass attached.
4. To study the longitudinal stationary waves and to determine the fundamental frequency of oscillations of the spring with both the ends fixed.

Introduction

Springs are familiar objects with many everyday applications ranging from retractable ballpoint pens, to weighing scales and metro handles. Let us consider an ideal spring: such a spring has an equilibrium length – that is, a length when no external force is applied – and resists any change to this length through a restoring force.

In such ideal springs, the force required to stretch (or compress) is assumed to be proportional to the amount which the spring is stretched (with compression being negative stretching). When one end is fixed, then this stretch is just the displacement of the free end. From now on we will assume that one end is fixed (which implies that we must use the formulas carefully when both ends are free). This tells us that the restoring force i.e. $F \propto x$. By defining a constant of proportionality k, known as the spring constant, we can write down *Hooke's Law*:

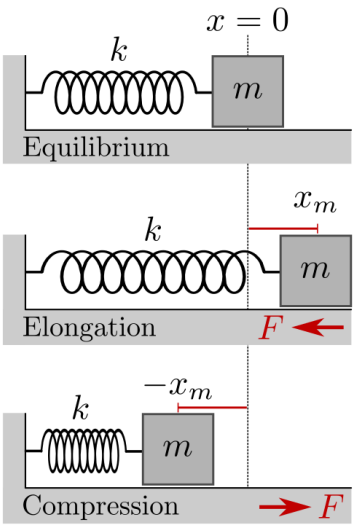


Figure 1.1: The restoring force experienced by the mass due to the spring is proportional to the amount the spring is stretched from its equilibrium position.

$$F = -kx, \tag{1.1}$$

where the negative sign shows that the force is in the direction opposite to the displacement of the end, i.e. it is a restoring force: when the spring is stretched, this force tends to compress it, and vice versa. Experiment reveals that for relatively small displacements, almost all *real* springs obey Equation (1.1).

Let us assume that the spring is placed horizontally on a frictionless surface and attached to a wall on one end and an object of some mass M on the other, which we can pull or push. If we displace the object by some x and let go, it will experience a restoring force due to the spring, which will change as the object begins moving. Combining Newton's Second Law and Equation (1.1) we can see that the acceleration of the object is given by

$$\begin{aligned} a &= \frac{F}{M} \\ \frac{\mathrm{d}^2x}{\mathrm{d}t^2} &= -\frac{k}{M}x \end{aligned} \tag{1.2}$$

where we have chosen $x = 0$ as the equilibrium position of the mass (in which the spring has its unstretched, or equilibrium, length).

It is not difficult to see that any function whose second derivative is itself (times a negative constant)

is a solution to the differential equation above. Both $\sin(\omega t)$ and $\cos(\omega t)$ are such functions. From the theory of differential equations, we know that the general solution to this differential equation is a linear combination of two linearly independent solutions. Thus we can write

$$x(t) = A\sin(\omega t) + B\cos(\omega t),$$

where ω – known as the frequency of vibration – must be equal to $\sqrt{k/M}$ for the left and right sides to match, and A and B are constants determined by the initial conditions (initial position and initial velocity). Observing the solution above, it should be clear to you (since sin and cos are periodic functions) that there must be some time T after which $x(t)$ comes back to itself, i.e. $x(t+T) = x(t)$. The motion is thus periodic, with period equal to that of the sinusoidal functions $\sin(\omega t)$ and $\cos(\omega t)$, which is

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{M}{k}}. \tag{1.3}$$

The mass thus oscillates about its equilibrium position with a time period T that is – at least for small amplitudes – independent of amplitude.

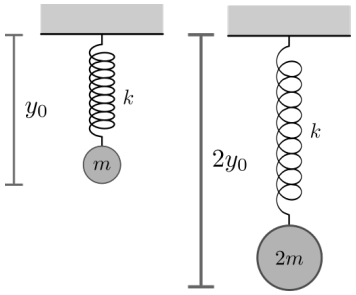


Figure 1.2: Hanging vertically, if a mass m causes the spring to extend to y_0, a mass of $2m$ will cause the same spring to extend to $2y_0$.

We could now imagine hanging the spring and mass vertically. While the spring is assumed to be massless (as it is ideal), the massive object experiences a downward force due to gravity of Mg. At static equilibrium, the spring would have displaced enough to balance this force. Thus,

$$y_0 = \frac{Mg}{k}. \tag{1.4}$$

We could similarly gently displace this spring vertically about this position, and we would again see oscillations, only this time the oscillations will be about the point y_0, which is the new equilibrium point.

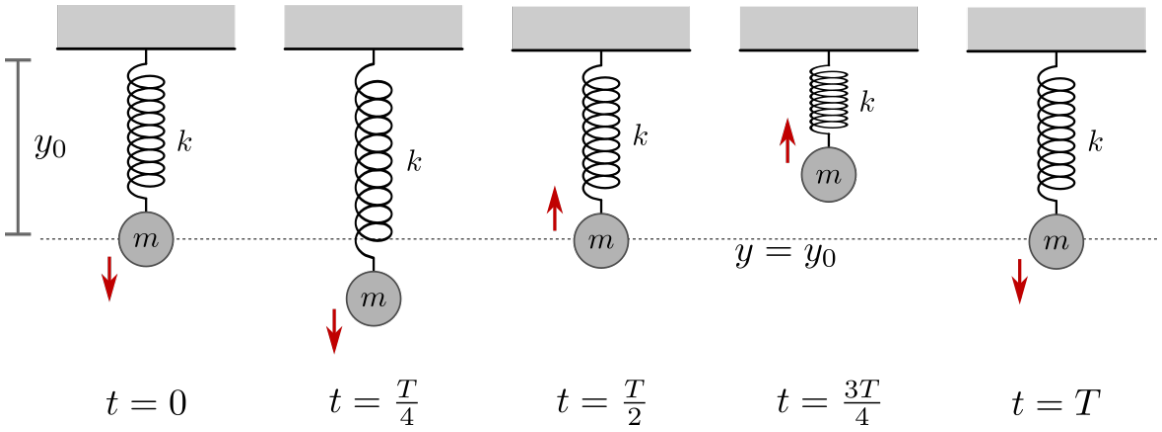


Figure 1.3: If the vertically hanging spring is displaced slightly, it will begin to oscillate with a time period T about the new equilibrium point y_0, as shown.

Question: Write out Newton's Law as a differential equation in this case (i.e. when gravity is included). Show that it can be written as:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}t^2} = -\omega^2 (y - y_0)$$

Question: By making a suitable substitution and comparing it with Equation (1.2), show that the solutions have the same form in a shifted variable.

Question: Show that the time period T is the same as in the preceding case (of "static" equilibrium).

Massive springs

So far, we have assumed our spring to be "massless", in that its mass may be neglected. Whether or not this is a reasonable assumption does not depend on the spring alone, but on external factors.

On inspection, it should be clear to you that the top and the bottom of a spring hung vertically do not experience the same downward or restoring forces, even at static equilibrium.[1] If this "droop" (caused by the mass of he spring) is negligible compared to the stretch caused by external masses, then the spring may be effectively considered as massless. Thus it depends roughly on the ratio of the masses.

[1] In general, a spring is not a "rigid" body.

The static case

Let us consider a point very close to the top of the spring: it will experience a (small) restoring force from the small amount of spring above it, and a much larger downward force from the mass of spring under it. Similarly, consider a point very close to the free end of the spring: it will experience a larger restoring force (as there is more "spring" above it) than downward gravitational force as there is very little mass under it. As a result, the spring does not stretch uniformly; the coils near the top are further separated than those near the bottom. We would thus expect some form of "correction factor" to the mass term in Equation (1.4).

Consider a spring with some mass m_s. Let L_0 be the length of the spring when it is kept horizontally with no forces extending it, and L_M be the length of spring when it is hung vertically with a mass M attached to its lower end. We can define the "equilibrium extension" S_M as

$$S_M = L_M - L_0. \tag{1.5}$$

It is possible to solve the problem theoretically and show that the massive spring acts effectively like an ideal spring with a mass $m_s/2$ attached to its end. In other words, if a massive spring is hung vertically, and an ideal spring with mass $m_s/2$ is attached to its end, they will both extend to the same distance.

Thus, Equation (1.4) is modified to

$$S_M = \left(M + \frac{m_s}{2}\right)\left(\frac{g}{k}\right) \tag{1.6}$$

where the factor $m_s/2$ is known as the **static** mass correction factor.

The dynamic case

Consider now the case of an ideal spring oscillating with a mass attached at its end. The oscillating mass M will have a kinetic energy

$$K_{\text{mass}} = \frac{1}{2}Mv^2$$

However, if the spring itself is massive, since different parts of it move at different velocities, it will also have some kinetic energy associated with it. To find the total kinetic energy of the spring, one could imagine an infinitesimal mass element on the spring moving at some velocity and integrate its contribution to the kinetic energy to get the total energy:[2]

[2]Though this is slightly advanced, it is not too difficult to do at your level. We urge you to attempt to show this theoretically.

$$K_{\text{spring}} = \frac{1}{6} m_s v^2 \tag{1.7}$$

This also explains why the spring oscillates even when it has no external mass attached.

Question: Show, from the earlier expressions for kinetic energy, that the spring behaves like an ideal spring with a mass of $m_s/3$ attached to it.

Thus, when an external mass M is attached to the spring, it behaves as if it is effectively an ideal spring with a mass $M + m_s/3$ attached to its end. Thus, Equation (1.3) will need to be modified.

The resulting time period T for the oscillations of a massive spring is given by

$$T = 2\pi \sqrt{\frac{\left(M + \frac{m_s}{3}\right)}{k}}. \tag{1.8}$$

The factor $m_s/3$ is the **dynamic** mass correction factor.

Note that this factor is different from the mass correction factor in the previous (static) case. The reason for this difference is that they arise from two different processes. In the first case, the effective mass comes from the fact that the different mass-elements comprising the spring experience different forces, which affects the effective *extension*. In the dynamic case where these different mass-elements are *moving*, the term $m_s/3$ comes from the fact that they have different velocities, which affects the effective *kinetic energy*.

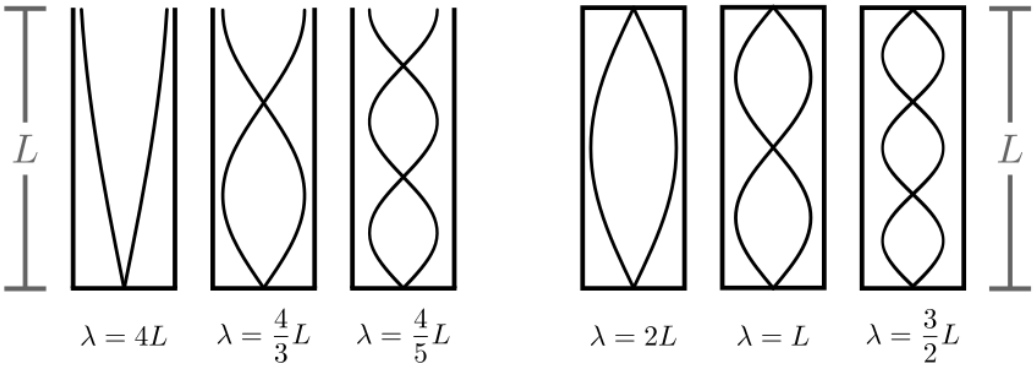
Question: Show that if the mass M that is attached is too large compared to m_s, Equations (1.6) and (1.8) give the same results as for an ideal spring.

When no additional mass is attached to the spring (i.e. $M = 0$), we can define a corresponding frequency for the massive spring

$$f_0 = \frac{1}{T} = \frac{1}{2\pi} \sqrt{\frac{3k}{m_s}}. \tag{1.9}$$

Standing waves on a massive spring

If we stretch the spring between two fixed points, it can be considered as a system with a uniform mass density (say, number of rungs per centimetre). When vibrated with some periodic forcing, this system has its own natural frequencies, very much like sound waves in a hollow pipe closed at both ends.[3]



(a) Waves on a pipe with one closed and one open end.

(b) Waves on a pipe with both ends closed.

Figure 1.4: Variation of amplitude in an air column: Sound waves are characterised by a series of alternate compressions and rarefactions. When an end is closed, the air particles there do not move, and thus that point is a node. Conversely, when the end is open, the particles have the maximum amplitude, and so that point is an antinode.

When this system is vibrated at one of these "natural" frequencies, standing (longitudinal) waves are seen to form, with clear nodes – i.e. points on the spring that remain completely stationary. Two important points can be noticed about the nodes in this case:

(a) They divide the length of the spring into equal parts.

(b) They increase by 1 as you move from one natural frequency to the next.

Question: Drawing out a diagram like Figure (1.4), show that the wavelength of a wave with n nodes is given by

$$\lambda_n = \left(\frac{2}{n+1}\right) L, \qquad n = 0, 1, 2 \ldots \tag{1.10}$$

where L is the length of the column and n the number of nodes (excluding the fixed end-points).

[3]The pipe being "closed" implies that the amplitude of the waves at both ends is zero. In this case, the top end of the spring is fixed, and the bottom end moves at an amplitude so small compared to the length of the spring, that it can effectively be considered to be fixed.

We know that waves satisfy the following equation that relates their wavelengths and frequencies,

$$v = f_n \lambda_n \tag{1.11}$$

where v is the velocity of the wave. As this is a constant for a configuration, it follows that

$$\begin{gathered} f_n = \frac{v}{\lambda_n} = \frac{v(n+1)}{2L} \\ f_1 = \frac{v}{2L}, \quad f_2 = \frac{v}{L}, \quad \ldots, \quad f_n = n f_1. \end{gathered} \tag{1.12}$$

Experimental Setup

Apparatus

1) A set of soft massive springs
2) A long and heavy retort stand with a clamp at the top end
3) A set of masses with hooks
4) A signal generator (*Equip-tronics QT-210*)
5) A dual output power amplifier with the connecting cords
6) A mechanical vibrator
7) A digital multimeter (*Victor VC97*)
8) A digital stopwatch (*Racer*)
9) Measuring tapes
10) A set of measuring scales (1.0 m, 0.6 m and 0.3 m)

Description

Digital Multimeter (*Victor VC97*) A multimeter is an instrument used to measure multiple parameters like voltage, current, and resistance. You will be using the multimeter to measure the frequency of the signal. You will have to use input sockets marked COM, V/Ω to do this. Note that the two input sockets marked mA and 10A are for the current measurement. You will not be using those. Connect the banana cables to COM and V/Ω, and select the Hz setting on the multimeter.

Signal Generator A signal generator is used to generate simple repetitive waveforms in the form of an alternating electrical wave. Typically, it will produce simple waveforms like sine, square, and triangular waves, and will allow you to adjust the frequency and amplitude of these signals. The instrument given to you generates sine, sawtooth, and square waveforms. The output may be taken from the respective output sockets through banana cables. The frequency can

be adjusted by turning the frequency dial and turning the Range knob to the appropriate multiplier. For example, turning the dial to 3 and selecting the 100X setting in the range knob would provide an output waveform with a frequency of 300Hz. Similarly the amplitude knob varies the amplitude of the output waveform.

Tip

The DC Offset button makes the signal oscillate about a constant non-zero DC voltage (instead of zero). Normally, this shouldn't affect the working of the setup. However, when a multimeter is connected to it to measure frequency, it will not be able to do so, as it is calibrated to measure signals alternating about zero.

Mechanical Vibrator The mechanical vibrator converts the electrical signals from the signal generator into mechanical vibrations, similar to how a speaker works. When using the spring as a uniform mass distribution, it should be clamped at one end of the stand, and its lower end should be clamped to the crocodile clip fixed on the vibrator. This end will be subjected to an up and down harmonic motion which will constitute the "forcing" of the spring. It must be ensured that the amplitude of this motion is small enough so that these ends could be considered to be fixed.

Power Amplifier The Power Amplifier is used to amplify the signals from the signal generator before it is sent to the mechanical vibrator. The reason for this is two-fold: the signal itself is not sufficiently strong, and even if it were, there is a risk that a signal with a large amplitude could damage the mechanical vibrator. In order to prevent this, the amplifier has been equipped with two fuses to protect the vibrator.

Precautions

- Don't overload the spring or you will stretch it beyond its elastic limit and damage it.
- Keep the amplitude of oscillations of the spring-mass system just sufficient to get the required number of oscillations.
- The amplitude of vibrations should be carefully adjusted to the required level using the amplitude knob of the signal generator so as to not blow the fuse in the power amplifier. The brighter and more frequently the indicator LEDs flash, the closer the fuse is to blowing.

Procedure

Part A

In this part, you will use the static method to determine the spring constant of a massive spring, by measuring the equilibrium extension of a given spring for different attached masses.

1. Measure the length L_0 of the spring keeping it horizontal on a table in an unstretched (all the coils touching each other) position.
2. Hang the spring to the clamp fixed to the top end of the retort stand. The spring extends under its own weight.
3. Take appropriate masses and attach them to the lower end of the spring.
4. Measure the length L_M of the spring in each case. (For better results you may repeat each measurement two or three times.) Thus determine the equilibrium extension S_M for each value of mass attached.
5. Plot an appropriate graph and determine the spring constant k and mass of the spring m_s.
6. Weigh the spring and compare its mass m_s with the one obtained from the graph experimentally.

Question: State and justify the selection of variables plotted on the x and y axes. Explain the observed behaviour and interpret the x and y intercepts.

Part B

In this part you will use the dynamic method to find the time period of a massive spring with different masses attached. The frequency of oscillations of the spring with the upper end fixed and the lower end free (i.e. with no attached mass) will be determined graphically through extrapolation.

1. Keep the spring clamped to the retort stand.
2. Try to set the spring into oscillations without any mass attached, you will observe that the spring oscillates under the influence of its own weight.
3. Attach different masses to the lower end of the spring and measure the time period of oscillations of the spring mass system for each value of the mass attached. You may measure the time for a number of oscillations to determine the average time period.
4. Perform the necessary data analysis and determine spring constant k and the mass of the spring m_s using the above data. Compare these values to those obtained in the Part A.
5. Also determine frequency f_0 for zero mass attached to the spring from the graph.

Part C

In this part, you will use a mechanical vibrator to force oscillations on the spring and excite its different normal modes of vibrations. Thus the longitudinal stationary waves will be set up on the spring, whose frequencies will be measured. The fundamental frequency f_1 in this case will be compared with f_0 obtained in **Part B**.

1. Keep the spring clamped to the long retort stand.
2. Clamp the lower end of the spring to the crocodile clip attached to the vibrator.
3. Connect the output of the signal generator to the input of the mechanical vibrator through the power amplifier, using a BNC cable.
4. Connect a the multimeter to the signal generator and set it to measure frequency.

   Tip

   While it might seem sensible to connect the multimeter to the mechanical vibrator, it is found that the amplification of the signal leads to a problem in detecting its frequency. Thus, it is better to connect it directly to the signal generator.

5. Starting from zero, slowly increase the frequency of the sinusoidal signal generated by the signal generator. At some particular frequency you will observe the formation of nodes: points on the spring which appear clearly visible as they are not moving.
6. Increase the frequency further and observe higher harmonics identifying them on the basis of the number of loops you can see between the fixed ends. (If you see n nodes – or fixed points – between the endpoints, there are $n + 1$ loops.)
7. Plot a graph of frequency for different number of loops versus the number of loops. Determine this fundamental frequency f_1 from the slope of this graph.
8. Compare this fundamental frequency f_1 with the frequency f_0 of the spring mass system with one end fixed and the zero mass attached (as determined in Part B) and show that

$$f_0 = \frac{f_1}{2}$$

Question: Can you think of why the two frequencies should be related by a factor of two? (You may use the analogy between the spring and an air column.)

References

[1] Rajesh B. Khaparde and H. C. Pradhan. *Training in Experimental Physics through Demonstrations and Problems*. English. First edition. Penram International Publishing, Jan. 2008. ISBN: 978-81-87972-34-1.

[2] Thomas C. Heard and Neal D. Newby. "Behavior of a soft spring". In: *American Journal of Physics* 45 (Nov. 1977). ADS Bibcode: 1977AmJPh..45.1102H, pp. 1102–1106. ISSN: 0002-9505. DOI: 10.1119/1.10956. URL: https://ui.adsabs.harvard.edu/abs/1977AmJPh..45.1102H (visited on 01/16/2022).

[3] Joseph Christensen. "An improved calculation of the mass for the resonant spring pendulum". In: *American Journal of Physics* 72.6 (June 2004). Publisher: American Association of Physics Teachers, pp. 818–828. ISSN: 0002-9505. DOI: 10.1119/1.1677269. URL: https://aapt.scitation.org/doi/abs/10.1119/1.1677269 (visited on 01/16/2022).

Experiment 2

The Three-Terminal Black Box

Objectives

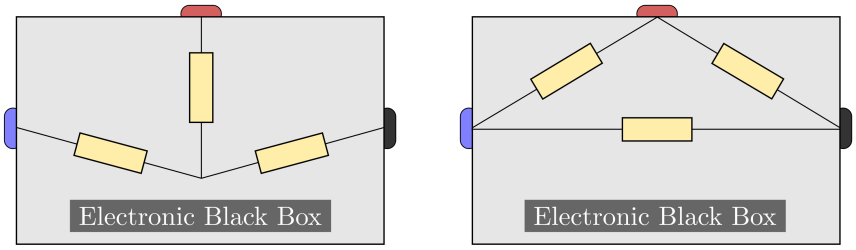
1. To study a three-terminal black box and identify the circuit within it along with the values of the electronic components.
2. To understand the responses of different active and passive circuit elements and their combinations, and to learn to recognise them.
3. To exercise your deductive and inductive powers, much as real physicists must do with real experiments.

Introduction

You are given a sealed box with three terminals marked A, B, and C which connect to a circuit inside. This circuit consists of **three** electronic components of different kinds, arranged between the three terminals in either a "star" or "delta" connection, as shown in Figure (2.1). You are expected to identify each component's type, value, and orientation with respect to the terminals.

You will be told beforehand if the connection is "star" or "delta" as you cannot distinguish them otherwise. This is because, if the circuit contains only resistors, every star connection has an equivalent delta connection and vice versa.

The black box will always contain **three** electronic components, but they need not all be distinct. Some components may be repeated; a black box may be composed of (i) three resistors, (ii) two resistors and a diode, (iii) a battery and two diodes, and so on. Each arm (see Figure (2.1)) may contain anything from all to none of the components.

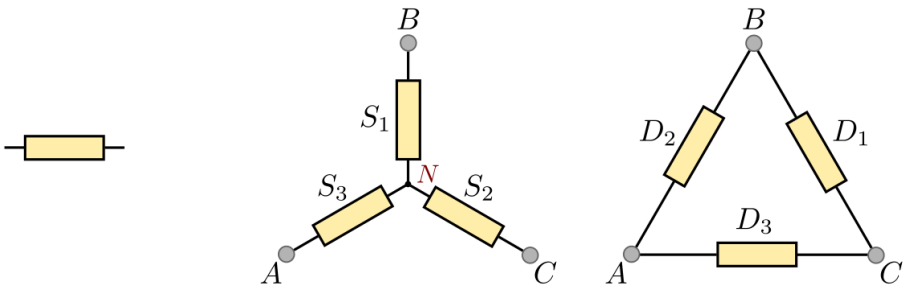


Star connection Delta connection

Figure 2.1: Different configurations possible within a blackbox: it will be specified if the box contains a star (*left*) or delta (*right*) connection. Do not attempt to open the box!

Star and Delta Connections

Standard 3-component circuits take on two major forms with names that represent the way in which the components are connected: a **Star** sometimes represented by Y, and a Delta connected network sometimes represented by a triangle or Δ.



(a) Component or combination of components (b) Star connection (c) Delta connection

Figure 2.2: In a star connection, one node is inaccessible and so only two components are connected at any time. In a delta connection, all the components are connected.

Both these forms, represented in Figure (2.2), are indistinguishable from each other, in that for every star connection there is an equivalent delta connection (with different values for the individual components) and vice versa. It is for this reason that you must keep in mind what form circuit in the black box has before drawing any conclusions.

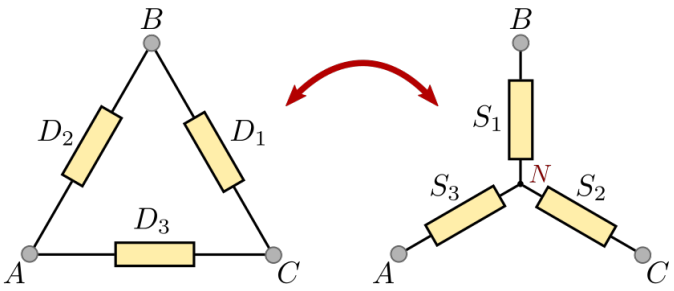


Figure 2.3: Every star connection has an *equivalent* delta connection. If the components happen to be resistors, explicit formulae may be derived for each arm.

If all the components are resistances, the following formulae will allow you to move between a star connection and its associated delta connection and vice versa:

Delta to Star Network:

$$P = \frac{AB}{A+B+C}$$
$$Q = \frac{AC}{A+B+C}$$
$$P = \frac{BC}{A+B+C}$$

Star to Delta Network:

$$A = \frac{PQ+QR+RP}{R}$$
$$B = \frac{PQ+QR+RP}{Q}$$
$$C = \frac{PQ+QR+RP}{P}$$

Passive circuit components

Passive circuit components are electrical components that cannot control the current in a circuit. They do not generate energy, but instead dissipate or store it.

Resistors A resistor is an electronic component used to oppose or limit the current in a circuit. It's behaviour is dictated by Ohm's law, which states that the voltage applied across the terminals of a resistor is directly proportional to the current flowing through it, with the constant of proportionality called the *resistance*, $V = IR$.

Capacitors A capacitor, made from two conductive plates with an insulator between them, stores electrical energy in the form of an electric field. It blocks DC signals (when fully charged) and allows the AC signals to pass through it. The charge stored in a capacitor is given by $Q = CV$.

When used with a resistor, the time a capacitor takes to charge or discharge is measured in units of an intrinsic time scale, known as the time constant $\tau = RC$ of the circuit.

Active circuit components

An active device is any type of circuit component with the ability to electrically control electron flow in the circuit.

Batteries Charges can be separated to produce a voltage. A battery uses a chemical reaction to produce energy and separate opposite sign charges onto its two terminals. As the charge is drawn off by an external circuit, doing work and finally returning to the opposite terminal, more chemicals in the battery react to restore the charge difference and the voltage.

p-n junction Diodes These are two-terminal semiconductor devices, which allow the electric current in only one direction while blocking it in the reverse direction. A diode has an anode (or positive end) containing positive charge carriers called "holes", and a cathode (or negative end) containing negative charge carriers (electrons). The interface between these ends forms a region without any charge carriers called the **depletion layer**.

The process of applying the external voltage to a p-n junction semiconductor diode is called biasing. If the diode is **forward biased** – that is, when a positive voltage is applied to the anode and a negative voltage to the cathode – it allows the charge carriers, and hence current, to flow. On the other hand, if the diode is **reverse biased** – negative to the anode and positive to the cathode – it blocks the current flow. In the conventional symbol for a diode, the arrowhead indicates the conventional direction of electric current when the diode is forward biased.

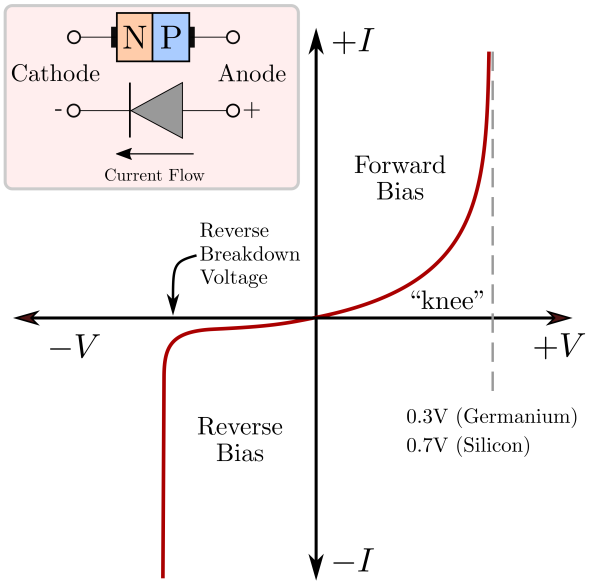


Figure 2.4: A diode and its current-voltage characteristics: for small (positive) values of input voltage, the output current varies exponentially.

Unlike a resistor, a diode does not behave linearly with respect to the applied voltage as the diode has an exponential current-voltage (I-V) relationship and therefore cannot be described

by simply using an equation such as Ohm's law. The reason for this is that the width of the depletion layer decreases with increase in positive voltage. After a certain "knee" voltage (0.3 V for Germanium semiconductors and 0.7 V for Silicon semiconductors) the width of the barrier effectively goes to zero, and the diode acts like a short circuit with zero resistance. When a junction diode is reverse biased the thickness of the depletion region increases and the diode acts like an open circuit blocking any current flow (except for a very small leakage current), until an "avalanche" voltage when the device undergoes a breakdown and gets shorted, leading to maximum current flow in the *opposite* direction.

Experimental Setup

Apparatus

1. A black box with three connecting terminals marked A, B, and C
2. A variable DC power supply (*Keltronix*)
3. Two digital multimeters (*MECO 603*)
4. A signal generator (*Testronix 72*)
5. A Digital Storage Oscilloscope (*GW INSTEK–GDS-1102U*)
6. A digital stopwatch (*Racer*)
7. A resistor ladder
8. Connecting wires

Description

Digital Multimeter (*MECO 603*) A multimeter is an instrument used to measure multiple parameters like voltage, current, and resistance. In this experiment you will use the given digital multimeters (Model *MECO - 603*) ***only*** for the measurement of DC and AC voltage and current. You will have to use input sockets marked COM, V/Ω and mA (or 20A) for the required measurements. Note that there are two input sockets marked mA and 20A for the current measurement. The socket marked mA may be used for measuring current below 250 mA and socket marked 20A may be used for measuring current up to 20 A. You will have to select the appropriate function and the range using the rotary switch provided on the multimeter. The value of voltage or current is displayed on the LCD screen. The multimeter is turned off by turning the dial to the appropriate setting.

Variable DC Power Supply (*Keltronix*) The DC Variable Power Supply can either act as a source of constant voltage (CV), or constant current (CC), indicated by the two LEDs present on it. We will be using it as a constant voltage source, so make sure the CV LED is lit. An ideal voltage source is one that produces a fixed voltage irrespective of the current the load

(your circuit) demands of it. Of course, this is not realistic; the supply given to you can supply a maximum potential difference of 15V, and a maximum current of 1A.

The voltage and the current in the circuit can be changed using the three knobs: voltage coarse, voltage fine, and current. The given power supply also displays the values of the output voltage and current. Do not use these values since the multimeter will be more reliable.

If the current knob is set to maximum, the supply acts as a constant voltage source. The constant voltage it supplies can be varied using the voltage knobs, and is independent of the load (your circuit) attached to it.[1]

> If the current your circuit draws at any voltage is beyond 1A, the power supply will not be able to supply it. This could happen if you short the two terminals of the power supply (don't!) or if you are close to the knee voltage of a diode (as in this case, the circuit has effectively zero resistance). In this case, the power supply's LED will shift from CV to CC, indicating that it is providing the maximum current possible. Avoid this situation, as it could damage the power supply.

Signal Generator (*Testronix 72*) A signal generator is used to generate simple repetitive waveforms in the form of an alternating electrical wave. Typically, it will produce simple waveforms like sine, square, and triangular waves, and will allow you to adjust the frequency and amplitude of these signals. The instrument given to you generates sine and square waveforms. The output may be taken from the respective output sockets. The frequency can be adjusted by turning the dial and selecting the appropriate Frequency Multiplier button. For example, turning the dial to 3 and selecting the 10k button would provide an output waveform with a frequency of 30kHz. Similarly the amplitude switch, used in conjunction with the appropriate buttons, can be used to adjust the amplitude of the output waveform from 0V to 30V.

Precautions

- The use of a multimeter to measure the resistance is strictly prohibited.
- Do not try to open the black box, use only terminals A, B, and C for the connections.
- Passing more current through the multimeter's (250 mA or 20A) sockets than they can handle will cause its fuses to burn out, leading to an open circuit. Use an appropriate resistance to limit the current in the circuit. The current should not exceed 500 mA.
- **Do not** short the outputs of the power supply, it could damage the equipment.
- When drawing circuit diagrams use the standard symbols.
- The digital multimeter, the DC power supply or the signal generator should be turned off if not in use.

[1] Similarly, if the voltage knob is set to maximum, the supply acts as a constant current source.

Procedure

Part A

The black box will contain only combinations of resistors, p-n junction diodes, and batteries in a ***star*** connection.

Question: Since not all of the elements will necessarily be present in a box, which is the first element you will test for?

1. Choose the first element that you would like to detect (or eliminate), and design a method to do this.
2. Repeat this for all the components.

   Question: How could you tell which way a diode was oriented?

   Question: How would you differentiate between a diode and a resistor?

3. Give your answer in the form of a circuit diagram showing the terminals A, B, and C clearly.

Tip

You will need to give *all* the information required in order to reconstruct the circuit. For example, you will need to determine the values of the resistors, the orientation of the diodes, and the voltage and orientation of the battery, if any of the above are used in the circuit.

Clearly report the procedural steps you have taken with reference to (i) your complete plan of experiment, (ii) the data you will collect, (iii) how the collected data will be analysed and (iv) how your analysis will be interpreted. Your reporting should be comprehensive, and all important steps should be mentioned in your answer sheet.

Parts B, C, & D *(optional)*

- **Part B:** The black box will contain only combinations of resistors, p-n junction diodes and capacitors in a ***star*** connection.
- **Part C:** The black box will contain only combinations of resistors, p-n junction diodes, and batteries in a ***delta*** connection.
- **Part D:** The black box will contain only combinations of resistors, capacitors, and inductors in a ***star*** connection.

References

[1] Rajesh B. Khaparde and H. C. Pradhan. *Training in Experimental Physics through Demonstrations and Problems*. English. First edition. Penram International Publishing, Jan. 2008. ISBN: 978-81-87972-34-1.

[2] *PN Junction Diode and Diode Characteristics*. en. Aug. 2013. URL: https://www.electronics-tutorials.ws/diode/diode_3.html (visited on 01/16/2022).

[3] *Relationship between Voltage, Current and Resistance*. en. Aug. 2013. URL: https://www.electronics-tutorials.ws/dccircuits/dcp_1.html (visited on 01/16/2022).

[4] *Star Delta Transformation and Delta Star Transformation*. en. Aug. 2013. URL: https://www.electronics-tutorials.ws/dccircuits/dcp_10.html (visited on 01/16/2022).

[5] *Beginners Guide to Passive Devices and Components*. en. Sept. 2013. URL: https://www.electronics-tutorials.ws/blog/passive-devices.html (visited on 01/16/2022).

[6] *RC Charging Circuit Tutorial & RC Time Constant*. en. Aug. 2013. URL: https://www.electronics-tutorials.ws/rc/rc_1.html (visited on 01/16/2022).

[7] *Electronic Components used in Electrical and Electronic Circuits*. en-US. Oct. 2014. URL: https://www.elprocus.com/major-electronic-components/ (visited on 01/16/2022).

Experiment 3

The Incandescent Lamp and the Inverse Square Law

Objectives

1. To understand the current-voltage characteristics of an incandescent lamp's filament.
2. To understand the dependence of the lamp's power on the resistance of its filament.
3. To observe and understand the variation of intensity with distance for a point source.

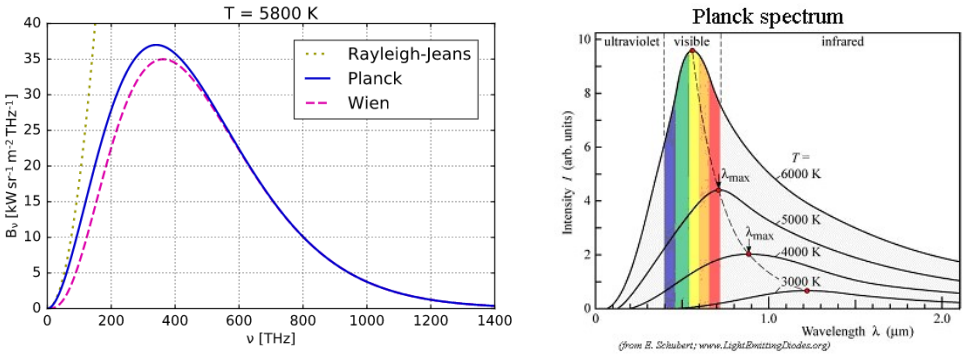
Introduction

Thermal radiation is the conversion of thermal energy into electromagnetic energy. You should remember from earlier courses that the **temperature** of an object is a statistical property that – at microscopic scales – is a reflection of the average kinetic energy of the atoms or molecules present within that object. Thus, an increase in temperature would cause these atoms or molecules to start 'jiggling' about more erratically (at least on an average). The higher the temperature, the more rapid this motion. This irregular thermal motion causes the charges that comprise the atoms and molecules to oscillate.

A famous result in electrodynamics states that accelerating charges radiate electromagnetic radiation. Each oscillation at a particular frequency can be considered a tiny "antenna" that emits and receives electromagnetic radiation. Of course, since these oscillations are due to random collisions within the object, there is a *continuous range* of possible frequencies of oscillations and thus of the emitted electromagnetic radiation. An object that emits radiation over a wide range of frequencies is known as a *black body*, and the characteristic spectrum of radiation it produces is called black-body radiation. For example, as a piece of iron is heated to increasingly high temperatures, it first glows

red, then yellow, and finally white. In short, all the colours of the visible spectrum are represented. In fact, the sensation of heat you get from the iron even before it begins to glow red is an indication of the emission of infrared radiation.

The spectrum of a blackbody was known quite well experimentally, however no theoretical approach could describe it accurately. There were two different laws, each of which fit the experimental data in different regimes, but neither of which explained the full spectrum. For low frequencies (or long wavelengths), the law was the "Rayleigh-Jeans law", while for high frequencies (or short wavelengths) it was known as "Wein's displacement law".



(a) Classical approximations (b) Variation with temperature

Figure 3.1: The Planck Spectrum for Blackbodies

The problem was solved by the German physicist Max Planck who realised that classical physics could not be used for such blackbody radiation, and thereby ushered in Quantum Mechanics by postulating that the energies were absorbed and emitted in specific quanta. The resulting spectrum was found to be characterised by one single parameter: the temperature of the object, and was found to fit a remarkable number of spectra from the sun to the Cosmic Microwave Background Radiation left as a remnant from the Big Bang.

Question: It is often said that the Cosmic Microwave Background Radiation has a "temperature" of 2.73 K. What do you think this means?

An incandescent lamp is composed of a wire filament protected from oxidation with a glass or fused quartz bulb that is filled with inert gas or a vacuum. If an electric current is passed through such a wire, it will begin to get heated. If the wire is heated to a sufficiently high temperature, it will give off light in the visible regime. This is the basis of a resistive incandescent lamp. The high melting point of tungsten (3680 K) and its low vaporisation pressure makes it a suitable material to be used as the filament of almost all incandescent lamps. The brightness and the overall colour of emitted light depends on the temperature of the filament for a given lamp, which is a non-linear resistive element.

At the heart of the lamp is a resistive element, which should follow Ohm's Law. Thus, when a voltage is applied across the lamp, a current flows through it, and the two are related by

$$V = RI. \tag{3.1}$$

However, as more current flows through the the filament, it gets heated which in turn changes its resistance. Thus, a more appropriate way of writing the above equation for the filament is

$$V = R(I)I \tag{3.2}$$

where R, despite being an "instantaneous" resistance, is not a constant value. Suppose we don't know the exact form of $R(I)$. One assumption might be to assume some sort of power law behaviour. Thus,

$$V = KI^a \tag{3.3}$$

where K and a are constants, and V is the voltage across the lamp and I the current passing through it.

Similarly, the power in the circuit can be calculated by the formula:

$$P = VI. \tag{3.4}$$

For a normal resistor which follows Equation (3.1) (known as Ohm's law), we could rewrite the equation as

$$P = VI = I^2R = \frac{V^2}{R},$$

however, in the case of an incandescent lamp where the ratio between the current and voltage is not a constant, this is no longer necessarily true. One finds for the resistive lamp, the empirical relation

$$P = CR^n$$

where P is the power supplied to the lamp, R is the resistance of the filament of the lamp, and n and C are constants whose values depend on the material used for the filament of the lamp. The above relation yields better results when the temperature of the filament of the lamp is approximately above 1800 K for the given lamp. For such temperatures the resistance R of the lamp is found to be directly proportional to the absolute temperature T of the filament of the lamp.

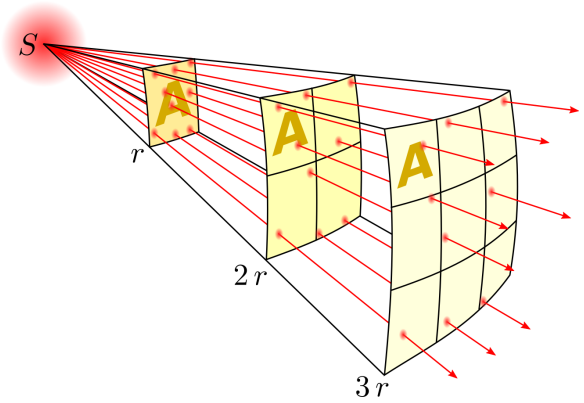


Figure 3.2: Motivating the inverse square law: Suppose S represents the light source, and the lines represent rays of light emanating from it. The total number of rays remains constant, as that represents the "strength" of the source. However, the density of these rays (that is, the rays per unit area) falls off with distance. If we imagine spheres of different radii (r, $2r$, and $3r$) centred around the position of S, then the density of rays on the surface of each of these spheres falls as the square of the radius, since the surface area of a sphere increases with the square of the radius. Thus, one would expect the intensity of the point-source to fall as an inverse-square law.

Point sources

A **point source** is a single identifiable localised source of, say, light. Such a source has negligible extent, distinguishing it from other source geometries, and are called point sources because in mathematical modelling, they can usually be approximated as a mathematical point to simplify analysis. The actual source need not be physically small, if its size is negligible relative to other length scales in the problem. For example, in astronomy, stars are routinely treated as point sources, even though they are in actuality much larger than the Earth.

For light or sound waves, a point source radiates the same intensity of radiation in all directions. That is, it has no preferred direction of radiation and radiates uniformly in all directions over a sphere centred on the source.

Experimental Setup

Apparatus

1. A DC variable power supply (*Optochem*)
2. Two digital multimeters with probes (*MECO 603* and *Victor VC97*)
3. A 12 V, 21W incandescent lamp
4. A stand to hold the lamp

5. A Resistor Ladder
6. Connecting cords
7. A convex lens
8. A photodiode

Description

Digital Multimeter (*MECO 603*, *Victor VC97*) A multimeter is an instrument used to measure multiple parameters like voltage, current, and resistance. In this experiment you are given two digital multimeters (Models *MECO - 603* and *Victor VC97*).[1] You will have to use input sockets marked COM, V/Ω and mA (or A or 20A) for the required measurements. Note that there are two input sockets marked mA and 20A (10A for the Victor multimeter) for the current measurement. The socket marked mA may be used for measuring current below 250 mA and socket marked 20A (or 10A) may be used for measuring current up to 20 A (or 10A). You will have to select the appropriate function and the range using the rotary switch provided on the multimeter. The value of voltage or current is displayed on the LCD screen. The multimeter is turned off by turning the dial to the appropriate setting.

Variable DC Power Supply (*Optochem*) The DC Variable Power Supply can either act as a source of constant voltage (CV), or constant current (CC). An ideal voltage source is one that produces a fixed voltage irrespective of the current the load (your circuit) demands of it. Of course, this is not realistic; the supply given to you can supply a maximum potential difference of 15V, and a maximum current of 2A.

The voltage and the current in the circuit can be changed using the three knobs: voltage coarse, voltage fine, and current. The given power supply also displays the values of the output voltage and current. Do not use these values since the multimeter will be more reliable.

If the current knob is set to maximum, the supply acts as a constant voltage source. The constant voltage it supplies can be varied using the voltage knobs, and is independent of the load (your circuit) attached to it.[2]

> If the current your circuit draws at any voltage is beyond 2A, the power supply will not be able to supply it. This could happen if you short the two terminals of the power supply (don't!) or connect it in any other way to a circuit that has effectively zero resistance. In this case, the power supply's output current will be 2A, indicating that it is providing the maximum current possible. Avoid this situation, as it could damage the power supply.

Photodiode A photodiode is a semiconductor device that converts light into an electrical current. The current is generated when photons of sufficient energy are absorbed in the photodiode, creating an electron-hole pair through the photoelectric effect.

[1] The Victor multimeter has a greater current sensitivity.
[2] Similarly, if the voltage knob is set to maximum, the supply acts as a constant current source.

Precautions

- Do not apply a voltage more than 12 V to the lamp.
- Clamp the lamp carefully on the retort stand so that the wires do not get shorted accidentally.
- Use the digital multimeter for the measurement of DC voltage and current. Choose the appropriate range for the measurement.
- The incandescent lamp should not be kept connected to the power supply for long duration.
- Passing more current through the multimeter's (250mA or 20A) sockets than they can handle will cause the fuses in the multimeter to burn out, leading to an open circuit. Use an appropriate resistance to limit the current in the circuit.
- When drawing circuit diagrams use the standard symbols.
- The digital multimeter and the DC power supply should be turned off if not in use.

Procedure

Part A

In this part, you will design an experiment to study the VI characteristics of an incandescent lamp.

1. Begin by drawing the necessary circuit diagram, and connect the lamp to the power supply through a multimeter working as an ammeter.

   Make sure you know how much current is flowing through the circuit so you know whether to use the 250mA or the 20A plugs on the multimeter.

   Question: How would you find the maximum resistance of a 12V, 21W lamp?

2. Connect a multimeter as a voltmeter across the lamp.
3. Vary the voltage across the lamp and measure the current.
4. Plot an appropriate graph to determine K and a in Equation (3.3).

Question: Describe qualitatively how you would expect the resistance of the lamp to vary with the current. Explain why you think this is the case.

Part B

1. Design an appropriate procedure to determine the value of n. You may use the data collected in **Part A**, or you may perform new measurements.
2. Plot an appropriate graph to determine C and n.

Part C

In this part, you will design an experiment to create a point source and then study the variation of intensity of the light with distance. For this you have to use the given photodiode with a digital multimeter to measure the intensity of light. We assume that the relation between the intensity of light incident on the sensing area of the photodiode and its output current is linear.

1. Use the converging lens to create a point source. You may use the screen with a hole in it to allow the passage of a small amount of light of this "point source" that may then be detected by the photodiode. The screen is now your point source.
2. Connect the output of the power supply to the given lamp and adjust the voltage applied to the lamp to be around 12 V.
3. Keep the photodiode initially as close as possible to the screen acting as your source.
4. Study the variation of current in the photodiode with the distance d between the source and the photodiode. (You have to take readings for the value of d varying from around 7.0 cm to around 50.0 cm.) You may have to correct your readings to account for the ambient light falling on the photodiode.
5. Plot an appropriate graph to show the variation of the intensity of light with distance d.

Question: What type of graph would be the most efficient to determine the variation of the intensity of light with distance? Why?

Question: After plotting this graph, you may find that points below a certain value (say, 20.0 cm) do not behave as you would expect them to. Can you explain why this is the case?

References

[1] E. Fred Schubert. *Light-Emitting Diodes*. 2nd ed. Cambridge: Cambridge University Press, 2006. ISBN: 978-0-521-86538-8. DOI: 10.1017/CBO9780511790546.

[2] *Photodiode : Construction, Types, Working & Its Applications*. en-US. Oct. 2016. URL: https://www.elprocus.com/photodiode-working-principle-applications/ (visited on 01/16/2022).

Experiment 4

Diffraction of Light

Objectives

1. To determine the wavelength λ of the light emitted by a laser source by studying the diffraction of light due to plane diffraction gratings.
2. To determine the width of the given single slit by studying its diffraction pattern.
3. To determine the diameter of a given wire by studying its diffraction pattern.
4. To determine the size of the circular aperture by studying its diffraction pattern.

Apparatus

1. A 10 mW semiconductor red laser source with a mount
2. A 5 mW DPSS green laser source with a mount
3. A set of plane diffraction gratings of different grating spacings with a holder and mounts
4. A single helix (spring) set in a holder with a mount
5. A double helix set in a holder with a mount
6. Measuring tapes
7. A set of screens
8. A single slit of fixed width mounted on a slide
9. Two multiple slits and a set of circular apertures mounted on slides
10. A spirit level

Introduction

In *Opticks* (1704), Issac Newton wrote "Light is never known to follow crooked passages nor to bend into the shadow". He explained this by describing how particles of light always travel in a straight line, and how objects kept in the path of the light cast a shadow because the particles can never spread out behind the object. However, a set of experiments on the propagation of light through small apertures performed by Francesco Grimaldi, Augustine Fresnel, Thomas Young and a few others firmly established that light actually enters into the shadow region with a definite pattern when it passes through around an edge. The resulting pattern depends on the relative size of the aperture or obstacle and the wavelength of light. If the size is much larger than the wavelength, the bending will be almost unnoticeable. However, if the two are similar in size, the diffraction will be considerable.

In this experimental problem, we will use a low power solid-state laser as a source of an intense beam of monochromatic light. When light from a distant source (or a laser source) passes around a thin aperture or through a narrow aperture and is then intercepted by a viewing screen, the light produces a pattern on the screen called a *diffraction pattern*. When such a beam is incident on various diffracting components like a plane diffraction grating, a single slit, a wire mesh or a two-dimensional diffraction grating, the light emerging from these components show a variety of interesting diffraction patterns. This pattern consists regions of maximum and minimum intensities, which characterise the diffracting object.

Theory

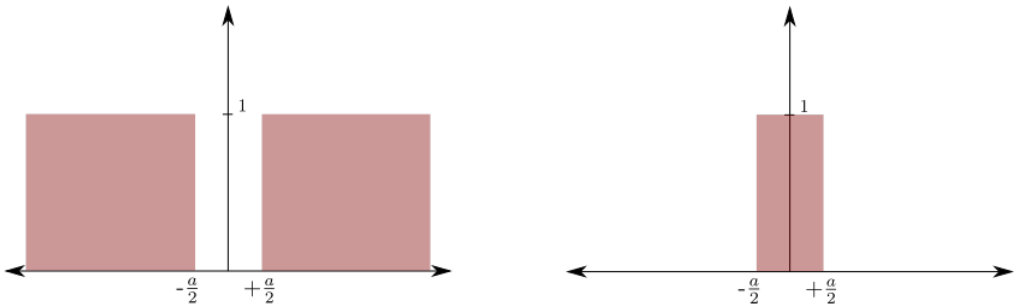
Babinet's Principle

Two objects are called *complementary* if one of them is transparent where the other is opaque and opaque where the other is transparent. An example of this is given in Figure (4.1).

Babinet's principle states that

"Complementary objects **produce the same diffraction pattern**, except for the intensity of the central maxima."

Henceforth we will not differentiate between the patterns made by such complementary objects. We will thus speak of a single slit (or of a thin wire) having the 'same' diffraction pattern, keeping in mind that the intensities of the central maxima are different.



(a) A thin wire (b) A narrow slit

Figure 4.1: Transparency of a thin wire and of a single slit.

Question: What are the complementary objects of the following:

1. A transparent sheet of paper.
2. A circular aperture of diameter d.

The Single Slit (or Thin Wire)

When a (monochromatic) beam of light such as a laser is incident on a narrow single slit, the light emerging from the slit shows a diffraction pattern on a screen. The distribution of the intensity of light received on a screen show a pattern of varying intensity consisting of a bright central maximum with alternate minima and maxima of decreasing intensity on either side, known as a *Fraunhofer diffraction pattern.*

Similarly, instead of a slit of width a, supposing we had a *wire* of diameter a placed as an obstruction to the laser beam. The resulting intensity pattern as observed on a screen almost exactly the same as that observed for a single slit. This intensity distribution can be written as a function of the angle θ as

$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2, \qquad \beta = \frac{\pi a \sin\theta}{\lambda}$$

This function has been represented in Figure (4.2).

Moving away from the central spot, when $\sin\beta = 0$ (but $\beta \neq 0$), the intensity vanishes! The positions of these **minima** (zeros) of the intensity distribution pattern are given by the relation

$$a\sin\theta_m = \pm m\lambda \qquad \text{for m} = 1, 2, 3, \ldots, \tag{4.1}$$

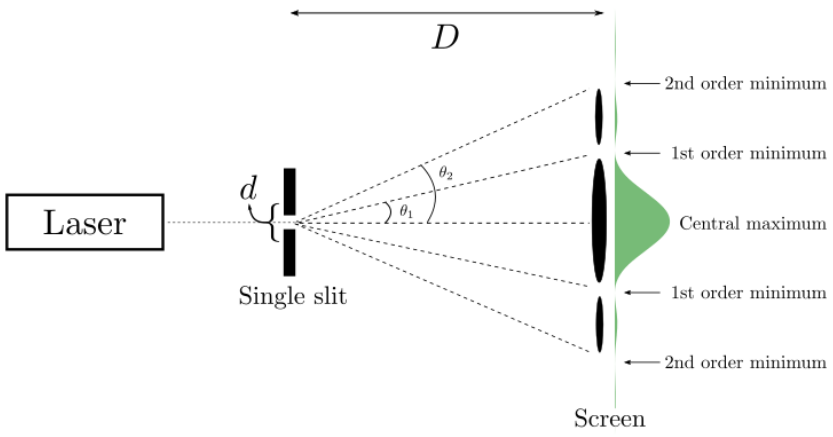


Figure 4.2: The diffraction pattern produced by a single-slit aperture (or thin wire).

where λ is the wavelength of the incident light, a is the width of the slit and θ_m is the angle corresponding to mth minimum. Here the $\pm$ refers to either side of the central spot ($\theta = 0$).

The Double Slit

We can now imagine two single slits of the same width next to each other. Such an arrangement is called a double-slit aperture. We could think of this aperture (shown in Figure (4.3)) as a combination of a single slit of width d, and a thin wire of width a.

This situation is slightly more complicated than the previous one; the thin obstacle of width a will produce a diffraction pattern like the one in the previous section. However, the light from the two slits on either side will **interfere**, producing an overall interference pattern which also has its own maxima and minima! The mathematics of this is too difficult to understand right now; you will cover it in a later laboratory.

The resultant intensity distribution is given by

$$I(\theta) = I(0)\cos^2\delta\left(\frac{\sin\beta}{\beta}\right)^2, \qquad \beta = \frac{\pi a\sin\theta}{\lambda}, \quad \delta = \frac{\pi d\sin\theta}{\lambda} \tag{4.2}$$

For a screen placed at a large distance D from the wire, the positions of the minima on the screen are observed at

$$\begin{aligned} x_{\pm n} &= \pm n\frac{\lambda D}{a}, && \text{due to diffraction,} \\ x_{\pm m} &= \pm\left(m-\frac{1}{2}\right)\frac{\lambda D}{d}, && \text{due to interference} \end{aligned} \tag{4.3}$$

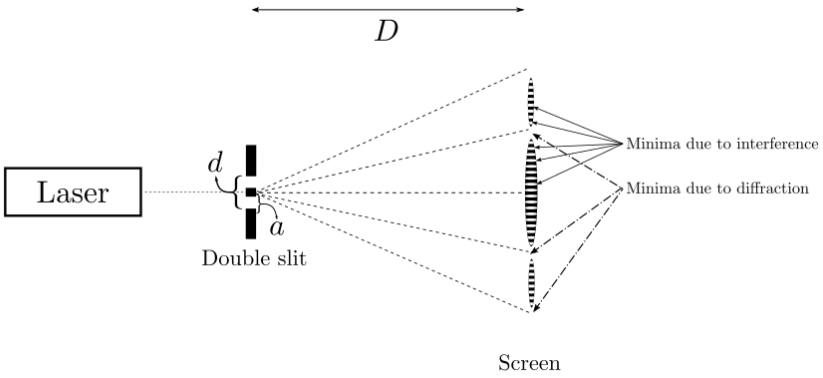


Figure 4.3: The diffraction pattern produced by a double-slit aperture is a combination of two patterns (diffraction due to a single wire and interference due to two slits).

Question: Look at the above condition for the fringes produced due to diffraction in Equation (4.3). Is it different from the result obtained in the previous section (Equation (4.1))? If yes, why? If no, why not?

Question: What is the "complementary" object to a double-slit?

Plane Diffraction Grating

A transmission diffraction grating consists of a large number of slits separated from one another by an opaque region. The grating concentrates the diffracted light along a particular direction in contrast to the single slit, which has a rather broad diffraction maximum. The maxima (bright intense spots) produced by a grating are usually called the principal maxima. They are quite intense and are also widely separated; what cannot be detected visually are the large number of secondary maxima which lie between neighbouring principal maxima.

The expression, relating wavelength λ of light used and the grating spacing d, with angle of deviation θ is

$$d\sin\theta_m = m\lambda, \qquad \text{for m} = 1,\ 2,\ 3,\ldots$$

In the above expression, m represents the order of **maxima** points and the angle θ_m corresponds

to mth order maximum intensity point. This relation is valid for a single slit and for the wire-like obstacle.

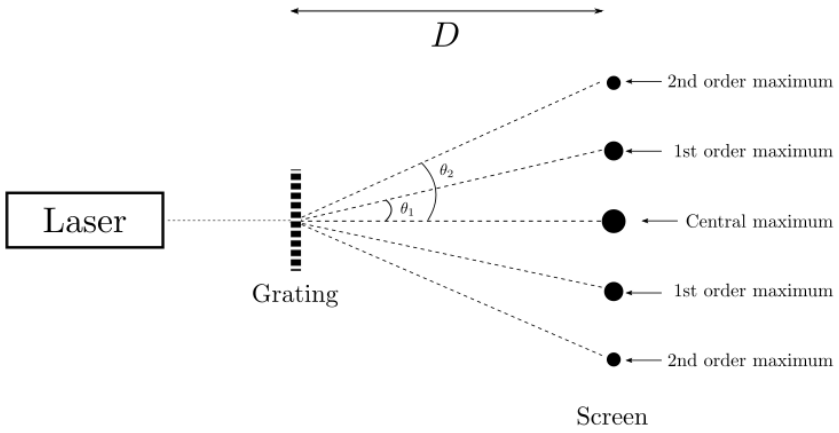


Figure 4.4: The diffraction pattern produced by a plane diffraction grating.

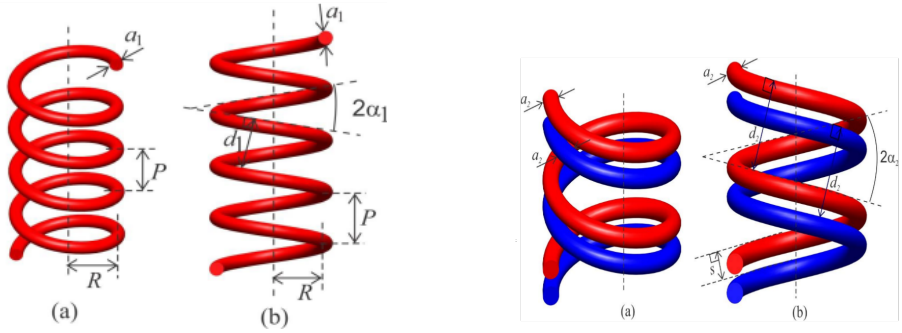
The Circular Aperture

The diffraction pattern due to a circular aperture (known as an *Airy diffraction pattern*) is similar to a single slit diffraction but the mathematics involved is more complicated which gives the expression nearly identical to that of the single slit. Hence we may apply the same expression to the diffraction due to a circular aperture,

$$d \sin \theta_m = \overline{m}\lambda, \qquad \text{for } m = 1, 2, 3, \ldots$$

where d is the diameter of the circular aperture and θ_m is the angle of deviation for the mth dark ring. The variable $\overline{m}$ has the following values:

$$\begin{aligned} m &= 1 \qquad \overline{m} = 1.22 \\ m &= 2 \qquad \overline{m} = 2.23 \\ m &= 3 \qquad \overline{m} = 3.23 \\ m &= 4 \qquad \overline{m} = 4.24 \end{aligned}$$

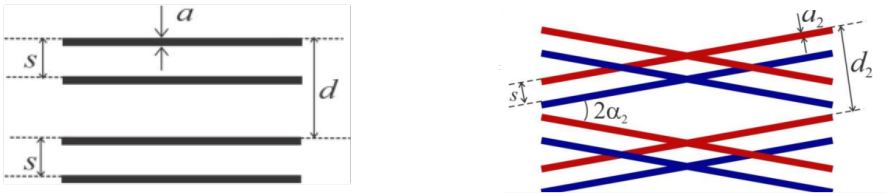


(a) Schematic of a single helix (like RNA) (b) Schematic of a double helix (like DNA)

Figure 4.5: Schematics of different types of helices. `Source: [1]`

Single and double helices

Now consider a set of four identical wires, the net intensity distribution is a combination of diffraction from each wire and interference due to pairs of wires and hence depends on a, d and s (see Figure (4.6a)). In other words, the combination of three different intensity patterns is observed.



(a) Projection of four identical wires (b) The double helix given in the sample.

Figure 4.6: Diffraction patterns of obstacles with three length scales. `Source: [1]`

Description

In **Part A**, you will determine the wavelength of light of a laser using a set of diffraction gratings.

In **Part B**, you will use the laser whose wavelength you have just determined to measure the width of a single slit.

In **Part C**, you will study diffraction pattern of a circular aperture.

In **Part D (*optional*)**, you will study the diffraction pattern of a single helix.

In **Part E (*optional*)**, you will study the diffraction pattern of a double helix.

Procedural Instructions

Part A

Observe effect of colour and also of white light on the diffraction pattern obtained by a suitable grating. Then choose an appropriate diffraction grating and perform the measurements to determine the wavelength λ of the laser.

Question: Estimate the error in the value of the wavelength of light.

Question: What are the sources of error in the above-determined value of λ? What measures should be taken to minimise these errors?

Question: Tilt the grating at an angle. How does this affect the diffraction pattern?

Now repeat this procedure for different gratings (with different values of d), and calculate wavelength λ as accurately as possible.

Part B

Design and perform the necessary experiment with a single slit of fixed width and determine the width d of the given single slit.

Question: Tilt the slit at an angle. How does this affect the diffraction pattern?

Part C

Now take the given circular aperture as the diffracting object and determine the diameter of the circular aperture.

Part D

Study of the diffraction pattern due to a helical spring and determine pitch of the spring and thickness of its wire.

Question: Can you explain the form of the diffraction pattern observed? Does this part of the experiment relate in any way to Part B?

Part E

Study of the diffraction pattern due to a double helix (as in our DNA) and determine all its parameters, as shown in Figure (4.6b).

Question Describe the diffraction pattern obtained if you use a laser source to illuminate

(a) A fine wire mesh,

(b) A square aperture,

(c) A rectangular aperture.

Precautions

1. Never look directly at a laser beam with the naked eye. It may damage the eye permanently.
2. Never point the laser at anyone else, for the same reason.
3. Never point an optical device at a laser beam. It could damage the internal sensors.
4. Never place highly reflective objects (such as rings, watches, and glassware) in the path of the laser beam.
5. For proper working of laser, it should be kept on throughout. Do not put it off until you complete all your readings, but if you do not need the laser beam for measurements or alignment, use a light-blocking screen to block the beam.
6. Treat the laser source as you would any other electrical device: It should never be tampered with while the power cord is connected.

References

[1] Praveen Pathak, Charudatt Kadolkar, and Manish Kapoor. *Diffraction due to Helical Structure*. 2015. URL: http://www.ipho2015.in/images/qna/ind/EI-Questions.pdf (visited on 01/17/2022).

[2] Rajesh B. Khaparde and H. C. Pradhan. *Training in Experimental Physics through Demonstrations and Problems*. English. First edition. Penram International Publishing, Jan. 2008. ISBN: 978-81-87972-34-1.

[3] Eric Stanley. "The use of crossed diffraction gratings to generate a three-dimensional bundle of diverging beams". In: *American Journal of Physics* 54.10 (Oct. 1986). Publisher: American Association of Physics Teachers, pp. 952–952. ISSN: 0002-9505. DOI: 10.1119/1.14803. URL: https://aapt.scitation.org/doi/10.1119/1.14803 (visited on 01/16/2022).

[4] Francis Jenkins and Harvey White. *Fundamentals of Optics*. English. 4th edition. McGraw Hill Education, July 2017. ISBN: 978-1-259-00229-8.

Experiment 5

Formation of Rainbow

Objectives

1. To study different orders of rainbows formed by a drop of water.
2. To study specifically the second-order rainbows for liquids of different refractive indices.
3. To determine the refractive index of an unknown liquid using the second-order rainbow it forms.

Introduction

The rainbow is probably one of the most spectacular light shows observed naturally in the sky. The conventional explanation to the formation of a rainbow is that sunlight – which is composed of a spectrum of colours – is separated into its constituent colours by the large number of water droplets present in the air on a rainy day, very much like the dispersion of light by a prism.

However, as we shall see, this is not the full story. To completely understand the formation of the rainbow and its location in the sky, we need to take into account the fact that the droplets are actually spherical, and that the incident light may be reflected many times within the drop before exiting it.

In the case of a primary (or first order) rainbow, the sunlight incident on a raindrop is refracted at the front surface, reflected at the back surface and refracted again as it emerges into the air.

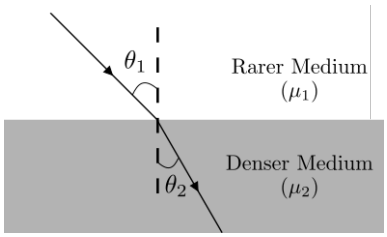
However, it is possible for sunlight to be reflected internally twice, before exiting the drop. While the intensities of such light rays will be much weaker than those that have been reflected only once, it is possible to see such 'secondary' (or higher-order) rainbows in nature. Secondary rainbows – as we shall see – are formed at angles that are further out from their primary counterparts. Thus, we

speak of the "order" of a rainbow as the number of times light has been reflected within the drop.[1] Higher-order rainbows are almost never seen naturally, as they are weaker than the background brightness of the sky and are usually positioned inconveniently with respect to the sun; if you are lucky, you can sometimes see a second-order rainbow.

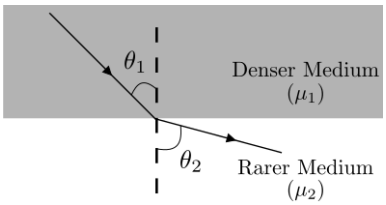
An instructive way to understand the formation of rainbows is to actually study the dispersion of white light by a spherical drop of water, as we shall in the present problem. (Note, however, that we use the word "rainbow" very loosely, to mean a "rainbow-like *spectrum*" rather than the more familiar "rainbow" in the sky.)

Theory

We will be dealing with the propagation of light between different transparent materials. When light moves from one medium to another, it changes direction through a process known as *refraction*. Thus, at least for the purposes of this experiment, we can characterise different transparent media by their refractive indices: a dimensionless number (μ) which describes how much light bends when it enters the medium from vacuum.



(a) $\mu_1 < \mu_2$:
The refracted ray is deviated towards normal.



(b) $\mu_1 > \mu_2$:
The refracted ray is deviated away from normal.

Figure 5.1: The propagation of a ray of light between media of different refractive indices.

The relationships between the angles θ_1 and θ_2 (shown in Figure (5.1)) is given by the Sahl-Snell–Descartes law (usually called Snell's law):

$$\mu_1 \sin\theta_1 = \mu_2 \sin\theta_2. \tag{5.1}$$

As light passes from a rarer to a denser medium (as shown in Figure (5.1a)), the ray of light bends *towards* the normal, while as light travels from a denser to a rarer medium (Figure (5.1b)), the ray of light bends *away* from the normal.

[1] **Note:** While the light *is* reflected inside the drop, it is *not* due to total internal reflection. The angles that the rays of light make with the normal inside the drop are less than the critical angle for the water-air interface.

The Primary Rainbow

Consider Figure (5.2) for first order ($K = 1$) rainbow formed by a spherical drop of liquid with refractive index μ.

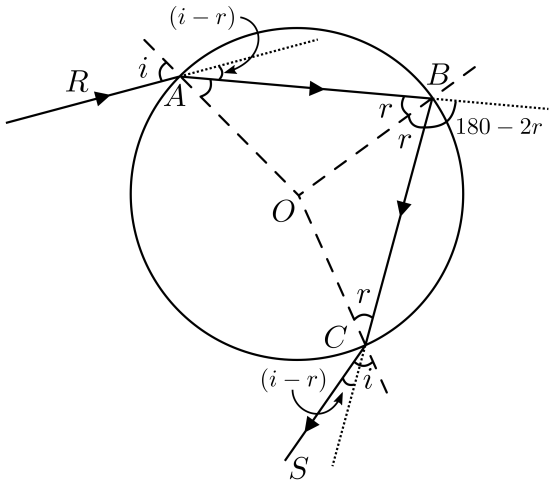


Figure 5.2: Ray diagram for the formation of a first order rainbow, with angles marked.

Let $RABCS$ be the path of a ray of incident monochromatic light, which comes out of a spherical droplet of liquid after suffering two refractions at A and C and one reflection at B.

Let i be the angle of incidence and r be the angle of refraction at A. It should be clear to you that r is also the angle of incidence and reflection at B.[2] At the point C, since the angle of incidence inside the droplet is r, the angle of emergence (at C) will be i.

Question: Show that at the point C, the angle of incidence is r and, using Equation (5.1), that the angle of emergence is indeed i.

We can now calculate the total *deviation* of the ray as it passes through the drop. From Figure (5.2) it should be clear that at the point A, if not for the drop the ray would have travelled along the dotted line. Thus, the deviation is $(i - r)$. Similarly, upon reflection at B, the ray deviates by $(180 - 2r)$. And finally, at C, the ray again deviates by $(i - r)$. Thus, the total deviation is:

$$\phi_1 = 2(i - r) + (180 - 2r). \tag{5.2}$$

[2]The triangle ABO is an isosceles triangle.

If we now consider the second-order rainbow as shown in Figure (5.3). The difference between this case and the previous one is that there are now two reflections within the drop. The deviations due to each of the reflections will be the same, and thus the total deviation due to both the reflections is $2 \times (180 - 2r)$.

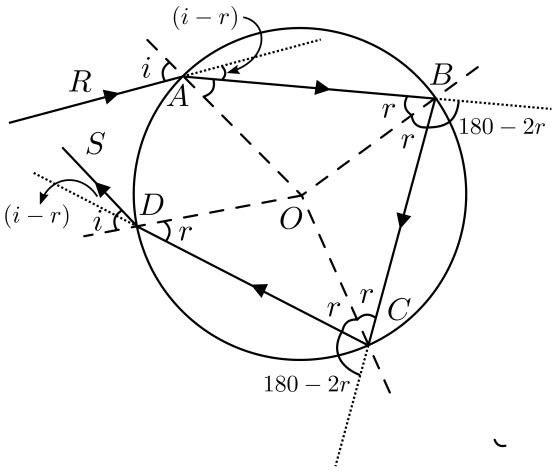


Figure 5.3: Ray diagram for the formation of a second-order rainbow, with angles marked.

We can now generalise this: for a K^{th} order rainbow (corresponding to K internal reflections), the deviation is $K \times (180 - 2r)$. As there are only two refractions (one at the point of incidence, and one at the point of emergence), the total deviation ϕ_K suffered by the emergent ray after K internal reflections will be

$$\phi_K = 2(i - r) + K(180 - 2r),$$

or more simply,

$$\phi_K = 180K + 2i - 2r(K + 1) \tag{5.3}$$

The angle of minimum deviation

Let us go back to look at the first order rainbow. We have:

$$\phi_1 = 180 + 2i - 4r$$

Question: Show that the angle φ between the initial and and final ray is given by

$$\varphi = 4r - 2i \tag{5.4}$$

Using Equation (5.1), we can express r in terms of i and μ. For a drop of refractive index μ, we can rewrite it as:[3]

$$\mu = \frac{\sin i}{\sin r} \tag{5.5}$$

or equivalently,

$$r = \sin^{-1}\left(\frac{\sin i}{\mu}\right). \tag{5.6}$$

The calculation of φ shows that it increases from zero at $i = 0$, to a maximum value of 42°, and then drops to about 14° at $i = 90°$. Consequently, the angle ϕ_1 *decreases* from 180° to a minimum value of 137.6° (as can be seen in Figure (5.4)). At this angle, the emergent ray is deviated the *least* from the direction of the incoming sunlight.

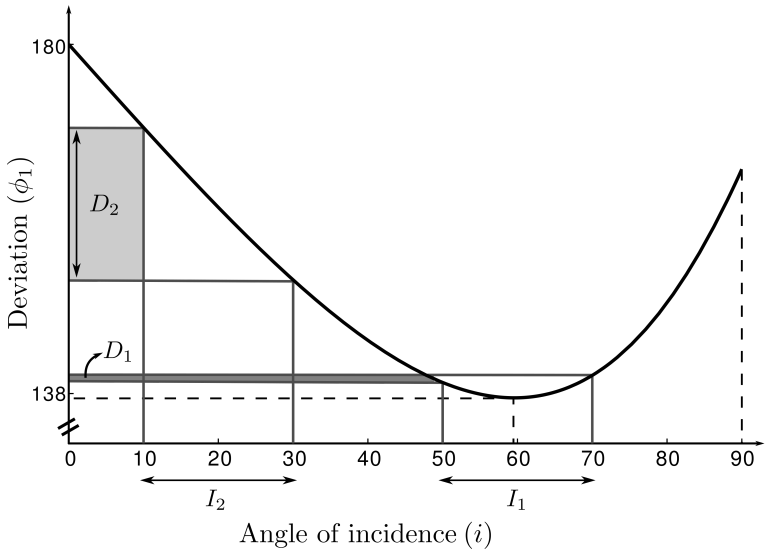


Figure 5.4: The angle of deviation approaches a minimum value of around 138° at $i = 59.4°$ ($r = 40.2°$ and $\varphi_1^{\max} = 42°$). The significance of this minimum is that rays from a range I_1 around the minimum "bunch" closer together than those from a range I_2 centred around some other point.

[3]The refractive index of air can be taken to be 1.

However, for us to actually *see* the rainbow, the different rays would nee to be "bunched" together, producing a higher concentration of rays that can be detected by our eyes. This is the significance of the angle of deviation being minimum: as shown in Figure (5.4), consider an interval I_1 centred on the minimum and an interval I_2 of the same width centred elsewhere. The range of deviation angles for values in I_1 (given by the interval D_1) is much smaller than the range of deviation angles for values in I_2 (given by the interval D_2). Thus, rays that hit the droplet with angles i in I_1 stick closer together than rays in I_2.

The reason that the rainbow appears coloured is that water has a slightly different index of refraction at different wavelengths causing this angle of minimum deviation – known as the "rainbow" angle – to vary with wavelength. In water droplets the angle of minimum deviation is 137.6° for red light and 139.4° for blue light. The arc of the rainbow comes from the fact that in order to see a rainbow, we must look at any direction which is at an angle of roughly 138° from the sun's direction,[4], and this condition describes a circle as shown in Figure (5.8). Thus, the rainbow is formed at a particular angle in the sky, and this angle is a function of the refractive index of the droplet.

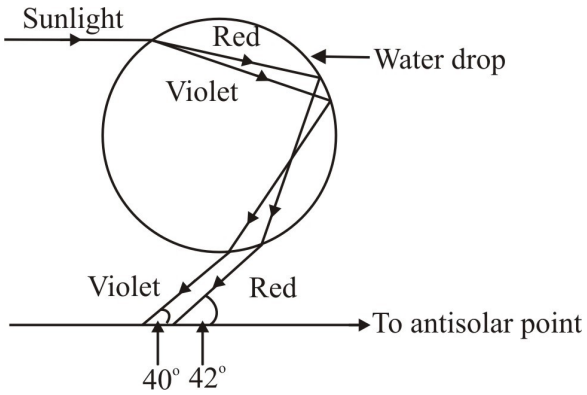


Figure 5.5: Ray diagram showing the dispersion of light within the droplet. The rainbow is thus seen roughly between 40° and 42° from the anti-solar point.

We will thus observe the rainbow at the minimum value of ϕ_K from the sun,[5] which we can find by enforcing the condition[6]

$$\frac{\mathrm{d}\phi_K}{\mathrm{d}i} = 0.$$

[4]Or 42° from the antisolar point.

[5]Or $\varphi_K = 180 - \phi_K$ from the anti-solar point: in other words, the K^{th} rainbow will appear at an angle φ_K from the line joining the tip of your head to the tip of your shadow's head.

[6]That this condition represents the case of minimum deviation is clear from the fact that if $\frac{\mathrm{d}^2\phi_K}{\mathrm{d}i^2}$ is calculated, it is found to be positive.

Figure 5.6: The rainbow in freshwater raindrops is extended below the horizon by a rainbow in seawater spray. The refractive index of saltwater drops causes the radius of the sub-horizon rainbow to be 0.8° less than that of the above-horizon freshwater rainbow. Solar elevation is 32.5°. `Photograph taken by J. Dijkema in the Pacific Ocean, 800 km southeast of Japan, during 1981.` © `G.P. Können.` `Source:` [1]

Question: Suppose for a moment that the refractive index of materials **didn't** depend on the wavelength of light. In this case, rainbows would:

1. Remain unchanged.
2. Not be formed.
3. Simply be a bright white band of light.

Question: In Figure (5.6), is the refractive index of salt water more than or less than that of the raindrops?

Differentiating Equation (5.3), we have

$$\begin{aligned}\frac{\mathrm{d}\phi_K}{\mathrm{d}i} &= 2 - 2\frac{\mathrm{d}r}{\mathrm{d}i} - 2K\frac{\mathrm{d}r}{\mathrm{d}i} \\ &= 2 - 2(K+1)\frac{\mathrm{d}r}{\mathrm{d}i} \\ \frac{\mathrm{d}\phi_K}{\mathrm{d}i} &= 0 \Rightarrow 2 - 2(K+1)\frac{\mathrm{d}r}{\mathrm{d}i} = 0\end{aligned}$$

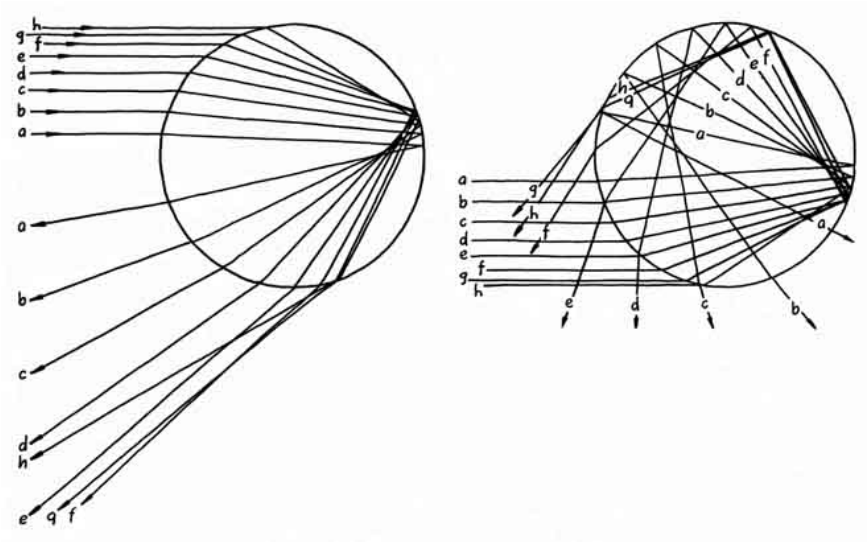


Figure 5.7: Ray paths forming a primary rainbow (*left*) and a secondary rainbow (*right*): The rays f and g in the primary rainbow and g and h in the secondary rainbow are 'bunched' together as they are close to the angle of minimum deviation. `Source: [2]`

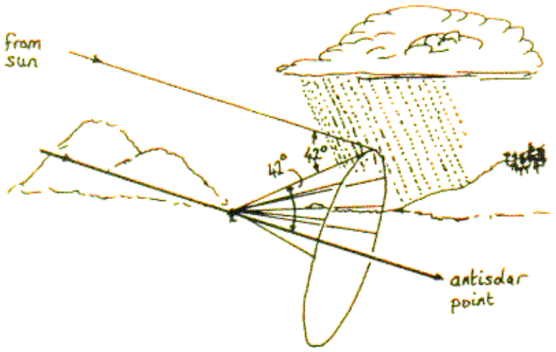


Figure 5.8: Angular position of the primary rainbow in the sky.

Hence,

$$\frac{\mathrm{d}r}{\mathrm{d}i} = \frac{1}{K+1} \tag{5.7}$$

Thus, for a K^{th} order rainbow at the angle of minimum deviation, Equation (5.7) relates the angle of refraction to the angle of incidence.

We can use Equations (5.6) and (5.3) to relate the total deviation ϕ_K for the K^{th} order rainbow to the angle of incidence i and the refractive index μ:

$$\phi_K = 180K + 2i - 2(K+1)\sin^{-1}\left(\frac{\sin i}{\mu}\right) \tag{5.8}$$

In particular, for $K = 2$, one can show – using Equations (5.8) and (5.11) – that

$$\cos\left(\frac{\phi_2}{6}\right) = \frac{1}{\mu}\left(3\cos\left(\frac{i+180}{3}\right)\cos i + \sin\left(\frac{i+180}{3}\right)\sin i\right) \tag{5.9}$$

We could also derive a relation between the angle of incidence and the refractive index for different orders of rainbows. Differentiating Equation (5.5) with respect to i, we obtain,

$$\cos i = \mu \cos r \frac{\mathrm{d}r}{\mathrm{d}i}$$

Substituting the expression for $\frac{\mathrm{d}r}{\mathrm{d}i}$ from Equation (5.7) in the above equation, and using the Snell's law relation and the trigonometric identity, $\sin^2 x + \cos^2 x = 1$, we have

$$\cos i = \frac{\sqrt{\mu^2 - \sin^2 i}}{K+1} \tag{5.10}$$

Squaring both sides, and using the same trigonometric identity again, the equation $(K^2+2K)\cos^2 i = (\mu^2 - 1)$ is obtained. Hence, we have

$$\cos i = \sqrt{\frac{\mu^2 - 1}{K(K+2)}}. \tag{5.11}$$

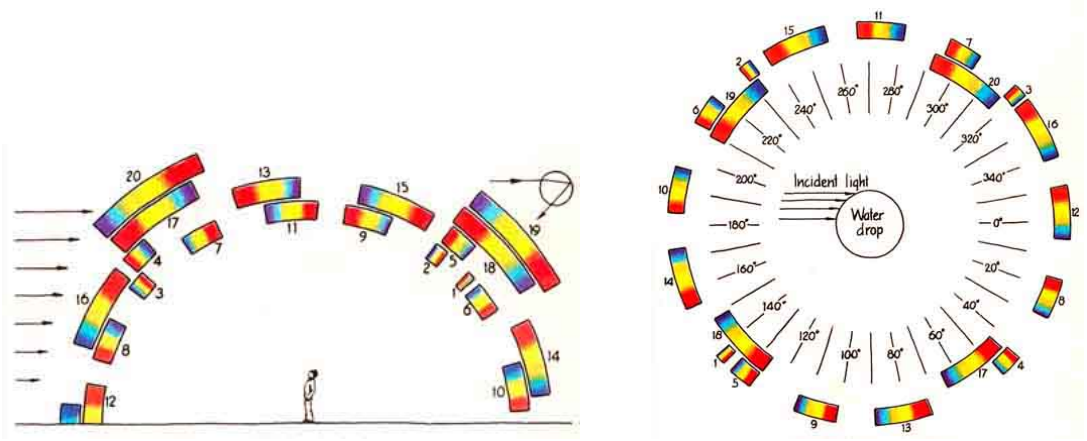


Figure 5.9: Rainbows represented as they would appear in the sky (*left*) and in this experiment (*right*), relative to the incident beam of light. `Source: [2]`

Experimental Setup

Apparatus

1. A modified spectrometer
2. An incandescent lamp
3. A power supply for the lamp
4. A holder for the syringe
5. Four small syringes
6. Four small beakers
7. Four small Petri dishes
8. A magnifying reading torch
9. Water, glycerine, clove oil, and an unknown liquid A (5 ml each)
10. A spirit level
11. Light blocking screen

Description

The (modified) spectrometer A spectrometer is an optical instrument, which may be used to study the spectrum of light, and can in particular be used to measure angles. It consists of four parts: a collimator, a telescope, a prism table and a circular scale disc.

The **collimator** is used to obtain a parallel beam of light. The collimator is a tube with a lens at the back (called the collimator lens) and an adjustable slit at the front. The distance between the slit and the collimator lens can be adjusted with the help of a knob fixed to its body. The **telescope** is used to collect and observe parallel light. The telescope is another tube with an arrangement of lenses (called the objective and the eyepiece) at the ends. The distance between the objective and the eyepiece may also be adjusted. The **prism table** is a circular table with levelling screws, which is mounted at the centre of the circular scale disc. The prism table is used to mount optical components (prisms, obviously, but also gratings, etc.). The **circular scale disc** is used to measure and record the position of the telescope arm and thus measures the angle of displacement. The circular scale disc has three scales, a circular main scale and two identical vernier scales. The vernier scales are fixed along two windows to the prism table and the circular main scale moves with telescope arm.

Modifications: In the modified spectrometer, the mounting base of the collimator arm has been extended to move the collimator away from the circular base and hence increase the range of movement of the telescope. The collimator is provided with an extra adjustable slit mounted externally on its back. This extra slit allows us to block the light falling on one side of the liquid drop so that only one half of the liquid drop is illuminated. The telescope does not have the regular lens arrangement, but instead the eyepiece is replaced by a pinhole and

the objective is replaced by a lens of suitable focal length. This is done to obtain a magnified image of the drop. The circular scale disc is kept as it is.

A specially designed holder is fixed on the prism table, within which a syringe may be clamped. The holder is designed so that when the syringe is clamped properly and the drop is obtained, the drop is seen from all the angular positions of the telescope exactly at the centre of field of view.

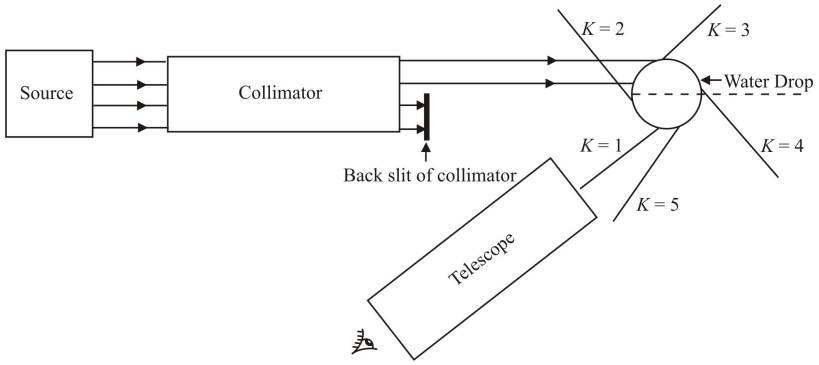


Figure 5.10: Schematic of the experimental setup

Halogen lamp A high power (150 W, 24 V) tungsten filament halogen lamp (such as those commonly used in projectors) is provided as a source of white light. This lamp has a small filament, through which a high current (6 A) can be passed. It is mounted in a wooden box, on a stand whose height may be adjusted. The wooden box has four windows through which the light may be observed. A small fan is fixed inside the box to prevent excessive heating. The lamp has a specially designed 24 V (AC) power supply which powers it. The intensity of light emitted by the lamp may be changed on the supply. This is essential for observing the higher-order rainbows, as one may need to have white light of higher intensity.

Miscellaneous A set of four small syringes, beakers, and Petri dishes are provided. Beakers may be used to store different liquids. Petri dishes can be placed on the prism table and are used to collect the liquid that falls from the syringe. Four different liquids (water, glycerine, clove oil, and an unknown liquid A) are also supplied. A magnifying reading torch is provided to read the circular scale, and a light blocking screen may be used to block the stray light falling on the drop. A spirit level may be used for levelling the prism table.

Useful Data

- Refractive index of water = 1.33
- Refractive index of glycerine = 1.47
- Refractive index of clove oil = 1.53

Precautions

- Avoid observing the source directly through the telescope for long time.
- Clamp the syringe on its holder carefully, so that it is held properly.
- Use different Petri dishes, beakers, and syringes, for different liquids to avoid mixing.
- Light reflected directly from the external surface of the drop will produce bright white glare spots that will hinder your observations. You will have to account for this and identify the rainbows carefully.
- The measurements have to be performed in a dark room.
- Use only one window of the spectrometer for recording the readings. Use the same one throughout your experiment.

Experimental Problem

Part A

In this part, you will study the formation of rainbows of different orders by a suspended drop of water and measure their angular positions. You can observe these rainbows directly with your eye and can measure their angular positions using a telescope and a circular scale.

Begin by determining the least count of the spectrometer.

1. Level the prism table using a spirit level.
2. Fill the syringe with water and mount it on the holder clamped to the prism table.
3. Keep the Petri dish on the prism table to collect the water drops, which may fall down from the nozzle of the syringe.
4. Obtain a steady drop of water at the nozzle of the syringe.
5. Adjust the height of the syringe such that the drop is seen at the centre of the field of view of the telescope.

   Question: The rays of light falling on the drop **must** be parallel. How would you make sure this is the case?

6. Adjust the alignment of the source, the collimator, the suspended drop, and the telescope, such that the drop is fully illuminated by the parallel beam of light. Close the back slit of the collimator partially, so that only the vertical half of the drop is illuminated by white light.

Question: If the rays of light falling on the drop are parallel, does this mean that all the rays striking the drop will be refracted by the same angle? Explain your answer.

7. Keep the telescope, collimator, and the drop in the straight line; you will see white light coming from the central region of the drop. Note the angular position of the telescope at which this happens and treat it as the "direct" – or undeviated – reading for further measurements. As stated earlier, this drop will produce rainbows of different orders, which can be observed at different angular positions around the drop. Thus the angular positions of different order rainbows (for water) will be as shown earlier in Figure (5.2).
8. Observe the bright first-order rainbow to the right of the drop ($K = 1$) first by eye and then by telescope. Adjust the telescope such that the bright red line of the spectrum can be seen. Measure the total angle of deviation ϕ_1.
9. Repeat the above step for the second-order rainbow ($K = 2$) and then for orders up to at least the fifth ($K = 5$). (In this case you may change the intensity of lamp.) Measure the total angle of deviation ϕ_K for each order. (Note: Before taking readings for each order, adjust the telescope to see the red line of the spectrum.)
10. Plot a graph of ϕ_K against K for the different order rainbows ($K = 1, 2, 3, \ldots$) formed due to a water drop.

Part B

In this part, you will obtain the rainbows with different liquids and relate the deviation for the second-order rainbow with the refractive index of the liquids.

1. Replace the syringe and Petri dish containing water by the syringe and Petri dish containing glycerine and adjust the syringe to see the glycerine drop from all the positions of the telescope.
2. Measure ϕ_2 for glycerine, following the same procedure as in **Part A**.
3. Repeat the above steps, for the clove oil and the given unknown liquid A. Determine ϕ_2 for clove oil and the liquid A.
4. Plot a graph of $\cos(\phi_2/6)$ against $1/\mu$ for water, glycerine and clove oil.

Question: When $\mu = 1$ (air), what is the value of ϕ_2? This should be taken as one of the points for plotting the above graph.

Part C (*optional*)

In this part, you will use the data from **Part B** to determine the refractive index of the unknown liquid A.

1. Measure the angle ϕ_2 for the unknown liquid A, following a procedure similar to **Part A**.
2. Use this value of ϕ_2, along with the graph from **Part B**, to determine the refractive index of the liquid A.

References

[1] Gunther P. Können. "Rainbows, Halos, Coronas and Glories: Beautiful Sources of Information". EN. In: *Bulletin of the American Meteorological Society* 98.3 (Mar. 2017). Publisher: American Meteorological Society Section: Bulletin of the American Meteorological Society, pp. 485–494. ISSN: 0003-0007, 1520-0477. DOI: 10.1175/BAMS-D-16-0014.1. URL: https://journals.ametsoc.org/view/journals/bams/98/3/bams-d-16-0014.1.xml (visited on 01/16/2022).

[2] Jearl D. Walker. "Multiple rainbows from single drops of water and other liquids". In: *American Journal of Physics* 44.5 (May 1976). Publisher: American Association of Physics Teachers, pp. 421–433. ISSN: 0002-9505. DOI: 10.1119/1.10172. URL: https://aapt.scitation.org/doi/10.1119/1.10172 (visited on 01/16/2022).

[3] Rajesh B. Khaparde and H. C. Pradhan. *Training in Experimental Physics through Demonstrations and Problems*. English. First edition. Penram International Publishing, Jan. 2008. ISBN: 978-81-87972-34-1.

[4] Kenneth Sassen. *Angular scattering and rainbow formation in pendant drops*. URL: https://www.osapublishing.org/josa/abstract.cfm?uri=josa-69-8-1083 (visited on 01/16/2022).

Experiment 6

The Carey Foster Bridge

Objectives

1. To understand how a Carey Foster bridge works.
2. To find the value of an unknown low resistance using the Carey Foster bridge.

Introduction

Measurement of low resistance

Ohm's law, $V = IR$, states that the current flowing through a resistor with resistance R varies linearly with the potential difference across it. To measure the resistance, one method involves studying the IV characteristics of a resistor, i.e looking at the variation in the current, I, flowing through the resistor for a range of applied potential differences, V and then extracting the value of the resistance, R, from the resulting graph/data. This direct application of Ohm's law usually works well for resistances in the range $1 - 100\Omega$.

For the measurement of low resistances, it is preferable to use methods that work on comparison of resistances (so you would need, in the first place, some precisely known, standard resistances). Such methods are based on the principle of a Wheatstone's bridge. A Wheatstone bridge is shown in Figure (6.1). A current measuring device, the galvanometer, registers current flowing in the arm BD. However, when the nodes B and D are at the same potential, no current flows in the arm BD. This happens when the following relation between the resistances is satisfied

$$\frac{P}{Q} = \frac{R}{S} \tag{6.1}$$

In such a situation, the Wheatstone bridge is said to be *balanced*. To find out the resistance of an

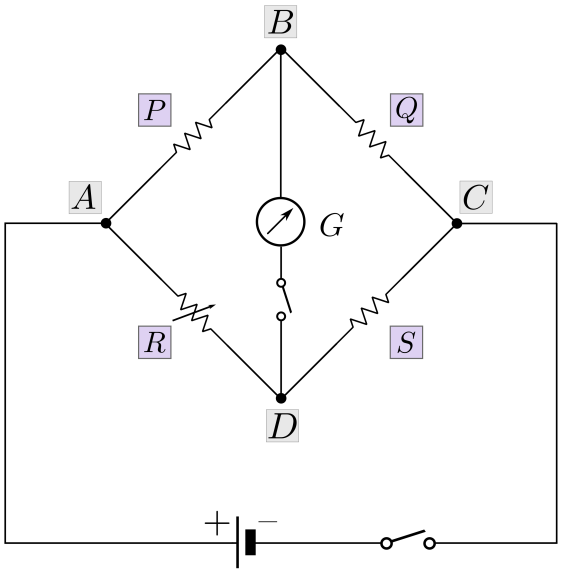


Figure 6.1: Schematic diagram of a Wheatstone bridge.

unknown resistor, two standard resistors are wired in arms AB and BC. The value of the (variable) resistor, R, is varied until the point of zero/null deflection/display is found on the galvanometer. Since, the bridge is now balanced, Equation (6.1) holds good, and we can infer the value of an unknown resistance, S.

We note here that this null method, which hinges on balancing the Wheatstone bridge, is useful because by not measuring a finite value of the current, the galvanometer does not have any role to play in the accuracy of the resistance measurement, our measurement of an unknown low resistances is limited only by the precision of other resistors in the experimental setup. In this experiment, we will use a form of modified Wheatstone bridge called a Carey Foster bridge, which is useful to measure low resistances while eliminating end corrections.

Theory

The Carey Foster bridge with a circuit around it is illustrated in Figure (6.2). The bridge consists of a copper strip with four gaps (these act as four arms of a Wheatstone's bridge). Two known resistances P and Q (equal and small) are connected in the inner gaps, VV' and WW'. A null detector (galvanometer) is attached between the terminal B and the jockey at D. A variable resistance box, R, with fractional resistances is placed in the outer gap at UU'. The unknown (low) resistance, S, which is to be determined, is put in the outer gap at XX'. A battery, a key, and a resistance box (if necessary) are connected between the terminals A and C. A one metre long wire MN of uniform resistance per unit length is mounted alongside a metre rod, and is soldered to the two ends of the copper strip. Since the jockey can be slid along the wire MN, the point D can be anywhere between M and N. It represents the position at which the null detector would register a value of 0. To locate

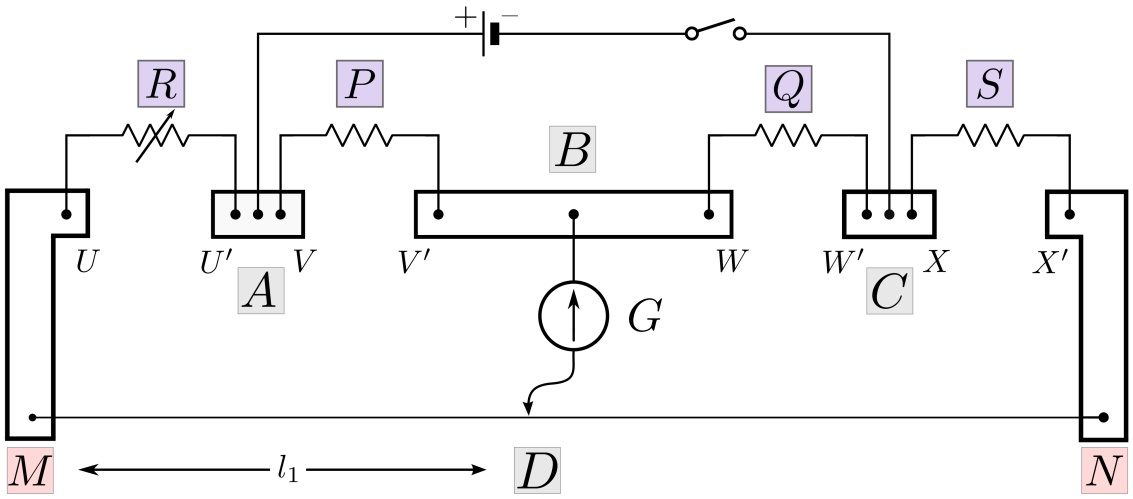


Figure 6.2: Schematic diagram of the Carey Foster bridge.

this point, we tap the jockey over the wire.

The wire, MN, is uniform and therefore, we may assume that it has a constant resistance per unit length, say, r Ωcm^{-1}. When the point D, located at some distance, l_1, from the end M is chosen such that the null detector displays 0, from the condition of a balanced Wheatstone bridge, we have, for the Carey Foster bridge

$$\frac{P}{Q} = \frac{R + \delta_1 + l_1 r}{S + \delta_2 + (100 - l_1)r} \tag{6.2}$$

Here, δ_1 is the value of the resistance at the soldered junction, M, and δ_2 is the value of the resistance at the junction N. These are end/contact resistances, which are unknown. Their presence would introduce an error in our measurement of S if used the CF bridge directly as a Wheatstone bridge. So we would like to eliminate them. Notice that if we interchange R and S (the resistor R is placed in gap XX', while the unknown resistance is now connected in gap UU'), then, because δ_1 and δ_2 correspond to the same physical contacts, the balance equation becomes

$$\frac{P}{Q} = \frac{S + \delta_1 + l_2 r}{R + \delta_2 + (100 - l_2)r} \tag{6.3}$$

Equations (6.2) and (6.3), together, yield the *nice* (independent of end corrections) relation

$$S = R + (l_1 - l_2)r \tag{6.4}$$

Clearly, by measuring l_1 and l_2 for an appropriate range of resistances, R, one can deduce the value of the unknown resistance, S and also the resistance per unit length of the bridge wire.

Question: Can we exchange the position of battery and the galvanometer? Which is the preferred arrangement and why?

Question: Why does the procedure involve interchanging R and S?

Question: Draw the equivalent Wheatstone Bridge.

Question: How is the Carey Foster an improvement over a standard Wheatstone bridge?

Question: What is the largest unknown resistance that can be measured using a Carey Foster bridge?

Question: In the derivation of the Equation (6.4), we eliminated δ_1 and δ_2 in favour of a relationships between R and S. We can also use the two preceding balance equations to eliminate R and S, yielding an equation for $\delta_1 - \delta_2$. Derive this, explain why it might be useful.

Experimental Setup

Apparatus

1. A low voltage DC power supply
2. A Carey Foster bridge
3. Two fixed standard resistances (1 or 2 Ω)
4. A fractional resistance box
5. A one-way key
6. A thick copper strip
7. An unknown low resistance (S)
8. A null detector (galvanometer)
9. Connecting wires.

Precautions

- Clean all contacts thoroughly including the teeth of the alligator clips with sand paper.
- Make sure that enough wire is exposed while making connections, so that insulation does not come in between.

- Check each contact by tugging on the wire and ensuring that it does not loosen or come undone.
- Arrange the elements of the circuit, so that it resembles your circuit diagram.
- While making the circuit think in terms of loops, not in terms of terminals.
- Make sure there is a finite resistance in the circuit, *before* the key is closed.
- Measure all balance lengths from the same side and on same scale.

Procedural Instructions

Before starting observations, *test* and *debug* your circuit. Some common methods of debugging include checking the DC supply potential, checking continuity of your circuit using a multimeter. If there is no potential drop across a passive element, then check for continuity.

Part A

In this part, you will determine the apparent value of a small resistance.

1. Make the connections as shown in the circuit diagram and connect the DC source. S is the unknown resistance.
2. Take out the zero resistance key from the resistance box (R) and note down the balance length (l_1). When determining the balance point, approach it from both sides and determine the range, if any, over which you get a null deflection.
3. Change the resistance in the box and find the new balance length.
4. Interchange the position of the unknown resistance and the resistance box and repeat steps 1 to 3 to determine l_2.
5. Tabulate $l_1 + l_2$.
6. Plot a graph between R and $l_1 - l_2$.
7. Use this graph to determine the the resistance per unit length of the wire, and the apparent value of the unknown resistance.

Part B

In this part, you will determine a small correction to the above value.

1. Make the connections as shown in the circuit diagram and connect the DC source. (S is a copper strip.)
2. Take out the zero resistance plug from the resistance box (R) and note down the balance length (l_1). Approach the balance point from both sides, and if there is a range of readings corresponding to the null deflection, note the entire range.

   Tip

   To make sure you are indeed at the balance point, infinitesimally moving the jockey, *on either side*, should correspond to a small but finite current value on the galvanometer.

3. Vary the resistance in the resistance box and find the new balance length.
4. Interchange the position of the copper strip and the resistance box and repeat steps 1 to 3 to determine l_2.
5. Tabulate $l_1 + l_2$. This is connected to the formula for $\delta_1 - \delta_2$ that you derived earlier.

   Question: Why have you been asked to tabulate $l_1 + l_2$ even though it is not used in the calculations?

6. Plot a graph between R and $l_1 - l_2$ on the same graph sheet as the one on which you plotted the readings in part A.
7. Examine the two lines carefully and check (i) whether they are parallel and (ii) what the apparent resistance of the copper strip is.
8. The unknown resistance in part A is not equal to the intercept but the difference between the intercepts of the lines in parts A and B.

Question: Why do the two lines – from **Parts A** and **B** – need to be parallel?

Question: Explain the correction factor determined in **Part B**.

Question: Why is the graph with the unknown resistance alone enough to determine both the resistance per unit length and the unknown resistance?

References

[1] B. L. Worsnop and H. T. Flint. *Advanced Practical Physics for Students*. English. 9th Revised edition. Littlehampton Book Services Ltd, Dec. 1951. ISBN: 978-0-423-73890-2.

Part III

Appendices

Appendix 1

Writing a Lab Report

1.1 Objective

To learn how to write a lab report.

1.2 Introduction

Writing how to report an experiment you have done is one of the most important skills you will learn in the lab. The lab report starts from your log book, in which you make notes on the various stages of the experiment as you do it: (i) planning and designing the experiment, (ii) making observations, (iii) analysing data and doing relevant calculations, including error analysis (iv) interpreting your results in light of the theoretical background. What you write in your log book is to the lab report what the initial sketches of an artist are to the final work of art that she produces.

You will perhaps also want to remember that your lab report will be graded. Here are some things that you should keep in mind; they are what your instructor and teachings fellows will notice.

Make your lab report **clear** and **easy to follow**. A common mistake students make is to write extremely long and elaborate reports that convey very little. We are not looking for pages and pages of writing; in fact, from experience, we have seen that the best reports are usually **short**. It is your job to decide what is relevant and what isn't. (This is good training for the future.) What you decide to include should be arranged logically and have a certain flow.

Your lab report may be written by hand, on Microsoft Word, or using LaTeX. Of these options, we suggest you begin using LaTeX as soon as possible: it makes your work look professional, and learning to use it is good practice.

Data does not lie. Contrary to popular opinion, grades are not assigned on how close the value obtained agrees with the "correct" one. If you are supposed to verify a law of nature, and you end up disproving it, that's fine, provided that you say that you have disproved it. If however you could not verify it but say that you did, we can only assume that you didn't understand the experiment. If your results are completely different from established values, then you have probably measured or calculated something incorrectly.

The lab report should contain **all** the information required for someone else to reproduce your experiment **and** its results.

A non-standard answer with a clear path as to how you got there is worth significantly more than a standard answer that appears unjustified or out of nowhere.

The handouts contain questions about the experiment are scattered around the manual. Make sure you try to answer them as you write your report; these questions are included at critical points to check your comprehension, and the answers often provide ideas of what to include in a good report.

Spend some time thinking about the format of your lab report. We don't expect something publishable in *Physical Review Letters*, but on the other hand, we don't expect four pages of stream-of-consciousness writing either.

1.3 Essential Elements of a Report

As stated above the the lab report should contain all the information needed for someone else to redo the experiment. This means that the following need to be written:

1. A brief introduction to the experiment, including the aim, the approach to be followed, and a statement of the theoretical background, if needed. (In the introductory lab, enough theoretical background is generally provided, but that may not be true in later labs; in that case you may want to include some of the theory needed to understand the experiment.)
2. The equipment used, with details. The details of the instruments don't need to be provided at the beginning but should be provided in the section where they are actually used. For example, if you are using Vernier calipers to measure the diameter of a set of balls, you need to specify the least count of the main scale and of the Vernier scale just before the table with the data on the diameters.
3. The procedure followed. This section should be very clear, and actually state the procedure *you* followed, not what some book says you ought to have followed.
4. The data, clearly tabulated. As stated above, each table should be preceded by the details of the instrument used to collect the data.

5. Error analysis, if needed.
6. Graphs, where needed.
7. The result, clearly stated.
8. A discussion of the difficulties encountered, and any ideas that may have occurred to you for improvement of the experiment, or other related experiments that could be done.

Tables, Figures, and Graphs

Tables, Figures, and Graphs (which we will come to in the next session) have a required standard of presentation, usually much higher than those you might put in your log book. If possible, have any tables and figures at the top and/or bottom of a page, and do not wrap text around figures or tables.

> In general, you may use any software for data analysis. Beginners will usually prefer to use Microsoft Excel or Google Sheets. However, as you progress, it is **very strongly** advised that you use this lab to learn how to do very elementary plotting and data analysis using Python and Jupyter Notebook. We will come to this in the next session.

Figures:

1. The figures you will include in your lab report will usually be descriptions of the apparatus or schematic diagrams.
2. In general, they should be well marked and labelled and you should be able to refer to them through the report to better explain what you've done.
3. All figures should have a label (such as "Figure 1:" or "Fig.1:") at the start of a caption explaining what they describe.
4. Captions should be short and sufficiently informative so that anyone with some knowledge of the experiment will understand what it represents without having to refer to your report.

Graphs:

1. Your graphs should be easy to read and clearly show all key features.
2. Graphs are figures, and should therefore have a figure number at the start of their caption ("Fig.1:", not "Graph 1:").
3. Consider whether various results can be combined into a single graph.
4. Here are some things your graph **must** have:

(a) Labelled axes with units,

(b) Error bars, indicating the uncertainty with which you know the location of the point,

(c) A white background,

(d) Sensible maximum and minimum values so that most of the space is filled by the graph.

5. Your graph **does not** need to have:

(a) A title: all the information about the graph should be in the caption,

(b) Grid-marks or a border,

(c) Legends (unless the graph cannot be understood without one): this information would preferably also go into the caption.

Tables:

1. Like figures, tables should be labelled and have a caption. However, they are **not** figures, but should instead be labelled "Table 1:", etc.
2. All the entries, including the headings, should fit comfortably in the width or height of the columns or rows; long headings should be avoided.
3. The heading of each column should include the **unit** of all the readings in that column. (There is no need to write the unit next to each reading.)
4. Include the uncertainties in every reading (after Session 3).
5. Make sure the rows and columns are the same size throughout the table.

While you will not encounter such situations in this laboratory, there are times when you may have an extremely large amount of data. **In such cases**, try to avoid tables of data that span many pages when the information is adequately given in a graph or by a few words of text; this is redundant and wasteful of space.

However, during this introductory lab, we will ask you to always include both tables and graphs.

1.3.1 Data Interpretation

The last step in your report is an **analysis** of your data. Usually, this is something that you can get from your graphs. Throughout this course you will be asked to plot graphs and to extract the relevant information from them. In general, this is something that is better learnt by doing rather than by reading. Once a regression line has been found, the equation must be interpreted in terms of the context of the situation being analysed.

In general, it is good practice to plot **linear** graphs whenever possible. In certain cases, the relationships between physical quantities are linear, and this is easy: for example, if a ball is dropped from rest from some height, its **velocity** v varies as

$$v = gt$$

Plotting v against t will give you a **linear** graph whose slope is the acceleration due to gravity. However, such a variation is not always guaranteed for all physical quantities. For example, the **position** S of the same ball varies as

$$S = \frac{1}{2}gt^2$$

which is a *quadratic* relationship. In some cases, you might be tempted to plot S as a function of t and fit a quadratic curve. In general, it is better to plot a graph between S and t^2. These two quantities exhibit a linear relationship which is easier to fit, and whose slope gives you $g/2$.

Of course, if you released the same ball with some (downward) initial velocity u, then

$$S = ut + \frac{1}{2}gt^2.$$

In such cases – i.e. cases where the relationship between two quantities contains *both* linear and parabolic terms – a second-order (or polynomial of order two) fit is necessary.

Tip

You will almost never, in an undergraduate laboratory, need to fit polynomials of orders higher than two. If you *must* fit such a polynomial, make sure you have a very compelling reason (say, an equation predicted by some physical model) before attempting such a fit.

Suppose – in the example of the falling ball given above – a higher-order polynomial is used. While this might *appear* to fit data better as it goes through all the points, when *more* data points are taken, the fit may not remain as "good". In such a case, however, a simple second-order polynomial would continue to fit the data just as well.

You will sometimes have to fit different functions like exponentials, logarithms, and power-laws. The above ideas can help here as well, but we will see more on this in Session 2 (*Graphing*).

Consider a quantity that obeys the following relationship:

$$y = A\ln(x) + B.$$

While the relationship between y and x is not linear, the relationship between y and $\ln(x)$ is. Thus, if the graph between y and $\ln(x)$ appears to be linear, then the slope and intercept would give us A and B.

As a slightly more complicated example, consider a physical quantity y that varies exponentially with another quantity x:[1]

$$y = A\, e^{x}$$

In such a case, the equation can be rewritten as

$$\ln(y) = \ln(A) + x$$

and so $\ln(y)$ and x obey a linear relationship from which A can be calculated.

1.4 Conclusion

Writing a coherent lab report is one of the most important skills that you can learn in this laboratory. The skill of communicating your results clearly and concisely is one that you will need to rely on frequently, irrespective of whether you choose to pursue a career in research or otherwise.

A basic report should give the reader the most essential information related to the experiment you are performing, along with details of the procedures, materials, and conditions used in the experiment so that they may reproduce the experiment and its results should they so desire. A report should also include the data, tabulated in such a way that your interpretation of the data comes about almost naturally. Your may also choose to illustrate this by recasting your data using appropriate graphs, tables, or figures. The reader will also need to know how certain you are of your results, meaning that you should have a detailed section on the analysis of errors and uncertainties in your report.

A good report is one that possesses all of these features, and is additionally a pleasure to read. This is by no means an easy task, but one that we hope that you will be able to master by the end of your undergraduate degree.

[1] For example, in a diode, the variation of the current with small positive input voltage behaves in this manner.

Appendix 2

A) Graphing with Python (solved)

Programming goals

1. Getting the packages you need to get started
2. Plotting and formatting your first graph
3. Getting your data into the program
4. Adding and formatting a trendline

Getting the packages you need

Python `packages` are pre-written scripts that you can `call` with your program to help you do whatever it is you want to do. They come with pre-defined functions and values that you can use. For example, imagine you wanted the value of the constant π.

Run the cell below this one. You can run it either by clicking the `Run` button on the top (in the toolbar), or by pressing `Shift + Enter` simultaneously.

[ ]:
```
pi
```

You have now seen your first Python error. Get used to errors; this one will not be your last.

You have asked Python to do something, and it does not understand what it is you have asked of it. As a result, it tells you where *it* thinks there's a problem and what the problem is (it's often right, but not always).

In this case, it says the problem is on line 1 of the cell (`----> 1 pi`) and the error is that it does not understand what you mean by `pi`.

Now compile the cell below:

[ ]:
```
import math
math.pi
```

We have `imported` the `math` package, a collection of standard mathematical values and functions. It contains, among other things, the value of `pi`, which we call by calling `math.pi`.

Some packages have horribly long names, and it would be quite cumbersome to have to type them out everytime, so we give them 'nicknames'. See if you can understand the following code snippet:

[ ]:
```
import math as m
m.pi
```

The NumPy package

NumPy (pronounced "Num Pie", not "Num Pee") is the fundamental package for scientific computing with Python. It contains almost all scientific functions that you require, and is optimised to be significantly faster than the usual Python functions for scientific operations.

If we wish to call any functions it has, we will be using the `np` nickname. This is not essential, but it is common. Try the following:

```
[ ]: import numpy as np
     np.pi
```

The MatPlotLib package

From the Matplotlib website:

"*Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code*".

We will be using it to plot all our data.

You will notice a strange line underneath this: `%matplotlib inline`

This is magic, don't worry about it. It's a command that needs to be there so that figures can be printed "inline" in Jupyter notebook and be stored in the document.

The SciPy package

SciPy ("Sigh Pie") is *another* library of open-source software for mathematics, science, and engineering. We will be using it for **curve fitting**.

You can now go ahead and import all the above packages using the cell below.

```
[ ]: import numpy as np                    # Importing the NumPy package

     import matplotlib.pyplot as plt       # Importing the Matplotlib package for plotting
                                           # "Magic" to display images inline
     %matplotlib inline

     import scipy as scp                   # Importing the SciPy package
     from scipy.optimize import curve_fit # Importing the curve fitting module from SciPy
```

Plotting and formatting your first graph

Lists

Let's start off immediately with a simple plot. The basic syntax for creating line plots is `plt.plot(x,y)`, where x and y are lists ***of the same length*** that specify the (x, y) pairs that form the line.

We can define a list by simply enclosing its elements within square brackets `[]`, like so:

```
listName = [element1, element2, element3]
```

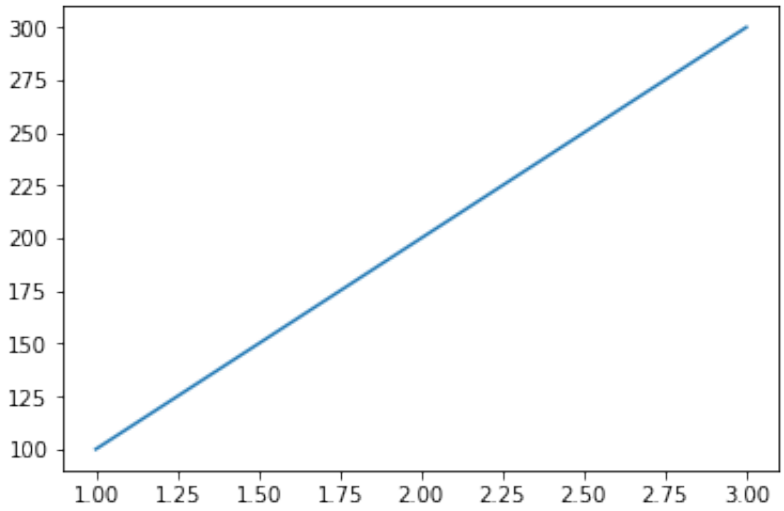
Exercise 1: *** In the following empty cell, define two lists, `xtest` and `ytest`, where:

`xtest` takes values 1, 2, and 3 `ytest` takes values 100, 200, and 300

```
[ ]: # SOLUTION:

     xtest = [1,2,3]           # Insert a three-element list of x-values
     ytest = [100,200,300]     # Insert a three-element list of y-values
```

Your first plot

Run the following code snippet: it will create a line plot with the above x and y values you have defined. You should get a graph like this:

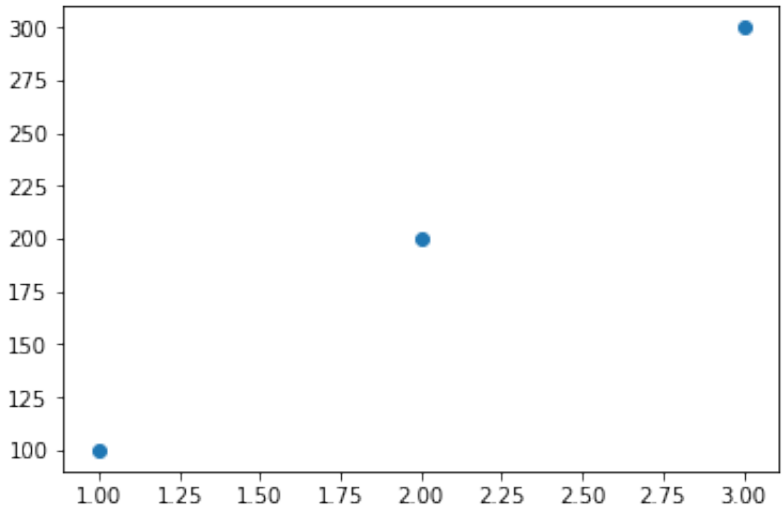


[ ]:
```
plt.plot(xtest,ytest)     # Plotting a 'line' plot of ytest vs. xtest
plt.show()                # This function simply "shows" the graph.
                          # it is not essential to understand how or why it works.
```

There you go, your first graph!

Of course, as mentioned in the previous section on plotting with MS Excel, **you will never plot a line from your data**. Instead, you will more often use the `plt.scatter(x,y)` command.

Exercise 2: *** In the following empty cell, insert one line to plot a `scatter` plot with the same data. After running it, you should get something like this:



[ ]:
```
# SOLUTION:

# Insert ONE line here to plot a 'scatter' plot of ytest vs. xtest
plt.scatter(xtest,ytest)

plt.show()                # This function simply "shows" the graph.
                          # it is not essential to understand how or why it works,
```

Formatting your graph

You will notice that it is still incomplete, you will need to label the axes. This can be done by adding modifying the `plt.xlabel` and `plt.ylabel` parameters, and re-plotting as shown below:

[ ]:
```
# SOLUTION:

plt.xlabel('x values (unit)')
plt.ylabel('y values (unit)')

plt.scatter(xtest,ytest) # Insert ONE line here to plot a 'scatter' plot of ytest␣
 ↪vs. xtest

plt.show()               # This function simply "shows" the graph.
                         # it is not essential to understand how or why it works.
```

You will often have to use scientific symbols in your graphs like λ and θ, and so it is helpful to learn how to do this.

Just type out `plt.xlabel(r'$\lambda$ ($\mu$ m)')` and `plt.ylabel(r'$\theta$')` in the cell below and compile it.

(The `r` present before the single quotes indicates that it is a "raw" string, and the greek letters are called using their LaTeX names. You can find out more **here**.)

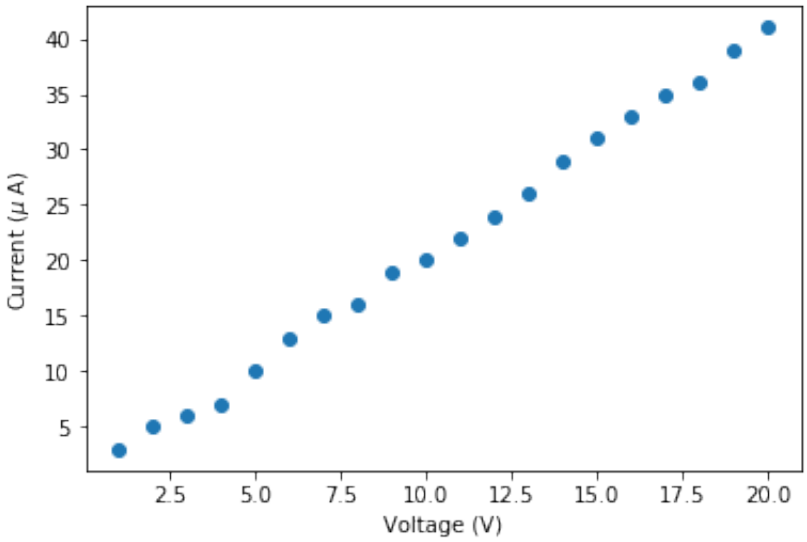
[ ]:
```
# SOLUTION:

# Insert one line of code to to change the x label of the plot
plt.xlabel(r'$\lambda$ ($\mu$ m)')
# Insert one line of code to to change the y label of the plot
plt.ylabel(r'$\theta$')

# Insert ONE line here to plot a 'scatter' plot of ytest vs. xtest
plt.scatter(xtest,ytest)
plt.show()
```

What if you had two sets of data to plot for the same x values? Try out the snippet given below:

[ ]:
```
xtest  = [1,2,3]          # A three-element list of x-values
y1test = [100,200,300]    # A three-element list of y-values
y2test = [50,150,250]
plt.scatter(xtest,y1test) # Plotting a 'scatter' plot of y1 vs. x
plt.scatter(xtest,y2test) # Plotting a 'scatter' plot of y2 vs. x
plt.show()
```

Plotting your data

Exercise 3: *** Let's now create two simple lists of (x, y) data given in this week's handout, and plot a scatter plot of it. You should get something that looks like this:



[ ]:
```
# SOLUTION:

xval = [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20]
yval = [3,5,6,7,10,13,15,16,19,20,22,24,26,29,31,33,35,36,39,41]

plt.scatter(xval,yval)
plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')
plt.show()
```

NumPy arrays

The "lists" that you've seen earlier (such as `[a,b,c]`) are those defined by default in Python. It turns out that the numpy package has its *own* version of lists, known as **arrays**. We will be using these arrays instead of the standard lists, as they have been optimised for scientific operations.

Numpy arrays can simply be created by wrapping the earlier lists with the `np.array` command, i.e. `np.array([list])`.

Exercise 4: *** In the next cell, define `xval` and `yval` as NumPy arrays, with the same values as above. Nothing should change with your output.

[ ]:
```
# SOLUTION:

# The same thing as the previous cell, only this time with NumPy arrays.

xval = np.array([1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20])
yval = np.array([3,5,6,7,10,13,15,16,19,20,22,24,26,29,31,33,35,36,39,41])


plt.scatter(xval,yval)
plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')
plt.show()
```

Importing data

Let's say you've got your data in MS Excel, and you want to use it in Python. If you have very little data, the easiest way to do this is to type it out yourself (as we've done above). However, if you have a large amount of data, it may be better to import a `.csv` file. CSV (or "Comma Separated Value") files

can very easily be created from spreadsheets like Excel (go to `File`⟶`Save As`, and select it from the drop-down list).

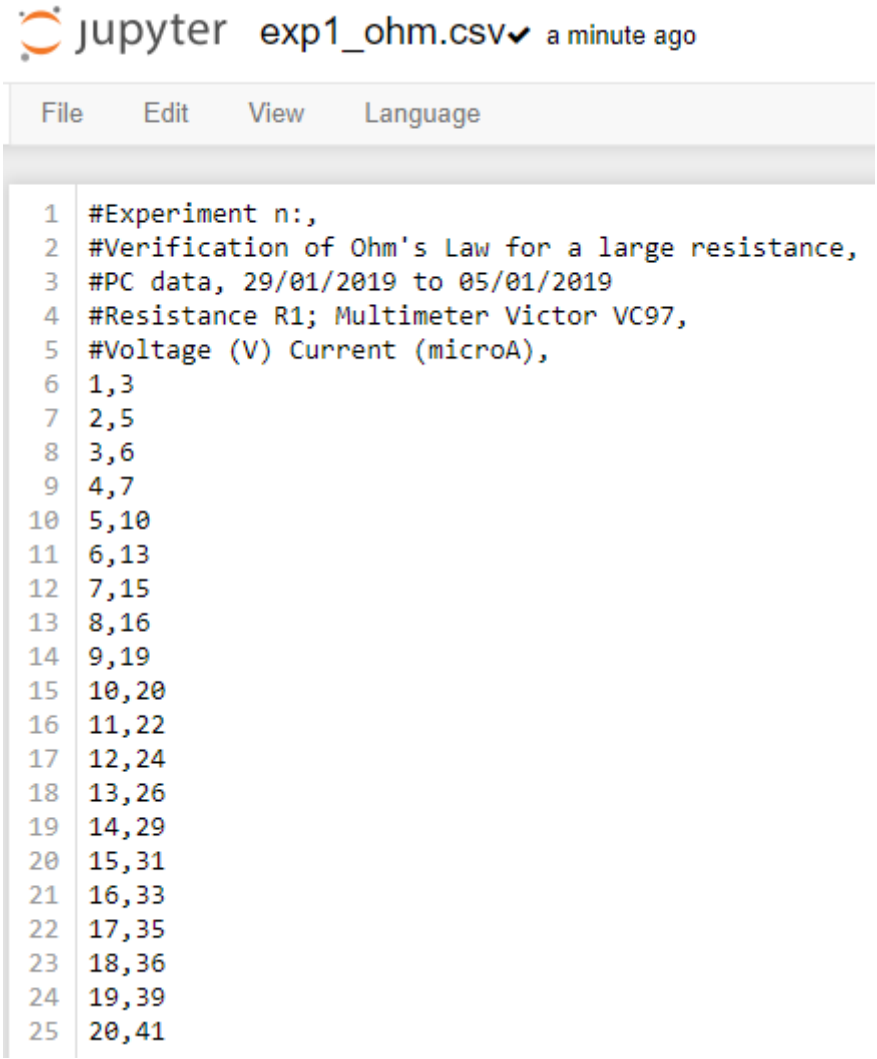
If you are asked to select a delimiter, choose `,`. This does not matter, but let us be consistent. Save the CSV file in the same folder as this Jupyter notebook, and remove all the lines that aren't the actual data.

Tip:

If creating CSV files is too difficult, here's a quick workaround:

1. On Excel, create a column with just your data values, separated by a comma. This can be done easily using the `Concatenate` function in Excel. For example: type `=CONCATENATE(A1,",",B1)` into the cell to merge them, separated by a comma.
2. Now copy the range you're interested in, and paste it in a `Something.txt` or `Something.csv` file and save it in the "data" folder that came along with this Notebook. ***

Sample data file In the "data" folder, you will see there is already a file called `exp1_ohm.csv` with sample data. You can access it in the `data` folder. In Jupyter, go to the `File` tab, and click `Open`. Navigate to the `data` folder and open `exp1_ohm.csv`. You will see something that looks like this:



You will notice that in each row, elements of the two columns are separated by a comma, and that the first rows with the names of these columns and other details have a # symbol before them. This is very important! When you import the data into Python, this # at the beginning tells the function that's importing the data to ignore these line. (Otherwise, these lines would be a string of letters, and other

lines would be numbers, and this wouldn't work). Put the # symbol in front of any lines you'd like to ignore while importing.

You are now ready to import this data using a simple `loadtxt` function that's present in the `numpy` package. We're going to learn how to do this in a slightly roundabout way:

1. Call the `loadtxt` function of the `numpy` package. You will see that it throws an error, since you have not given it the name of the file to load.
2. Now, you will start writing out the name as a string. Write out the following first `"./"` and then press the `TAB` button on your keyboard. You will see that Jupyter will show you the files in the current working directory (where your code is stored)! Thus, the "current working directory" is called `..`
3. You need to get to the `data` folder which is one folder up. The standard way to get to the folder above it by using `../`. Type this out, and press the `TAB` button.
4. Select the `data` folder, and then press the `TAB` button again. Now select `exp1_ohm.csv`.
5. Running this will *still* cause an error, since you haven't told Python that the different columns are separated by `,`s. Add a comma after the string, and type out `delimiter = ","`. This tells Python that the data in each row is separated ("delimited") by commas.

Tip:

Using the `TAB` key to autocomplete commands is a fantastic trick, useful not only here but also on most UNIX-type machine terminals (like Macs). ***

Exercise 5: ***

Complete the steps detailed above and run them in the empty cell below, you should get an output that looks like this:

```
array([[ 1.,  3.],        [ 2.,  5.],        [ 3.,  6.],        [ 4.,  7.],        [
5., 10.],        [ 6., 13.],        [ 7., 15.],        [ 8., 16.],        [ 9., 19.],
[10., 20.],        [11., 22.],        [12., 24.],        [13., 26.],        [14.,
29.],        [15., 31.],        [16., 33.],        [17., 35.],        [18., 36.],
[19., 39.],        [20., 41.]])
```

This is a two-dimensional array (very much like a matrix). It is useful to learn how to manipulate them, and I've added a short optional section about them at the bottom. But for now, we won't really use them since there's a much simpler way to "unpack" your data directly in a format you can use.

[ ]:
```
# SOLUTION:


# Insert ONE line here to load the file exp1_ohm.csv from the data folder
np.loadtxt("../data/exp1_ohm.csv",delimiter=",")
```

Unpacking data Wouldn't it be nice if Python understood that the first column was one variable's data and the second another variable's data? It turns out that the `NumPy` package allows for just this, using another option called `unpack`.

Using this option, you can equate **any number** of variables to this data, and Python will automatically send the first column to the first variable, the second to the second, and so on. Thus, if you did:

```
x1,x2,x3,...,xn = np.loadtxt(...,unpack=True)
```

to a file that contained `n` columns, the first column would be stored in `x1`, the second in `x2` and so on!

Exercise 6: *** In the cell below, write down the code to take the first column as `xpoints`, and the second column as `ypoints`.

[ ]:
```
# SOLUTION:
```

```
# Load the file exp1_ohm.csv from the data folder, unpacking by column
xpoints,ypoints = np.loadtxt("../data/exp1_ohm.csv",delimiter=",",unpack=True)
```

Adding a Trendline

Adding a trendline in Python is not as obvious for the simple linear graphs as it was in Microsoft Excel, but provides much more functionality. The idea we will use is the following:

1. We will define a function (say) `f` which accepts some parameters (`x,a,b,c...`) (how many depends on what type of function we're fitting) and `returns` a value.
2. We will then call `scipy.optimize`'s `curve_fit` function with `f` and our data (`xpoints` and `ypoints`) which will automatically vary the parameters to give us the best values for `a,b,c...`

This is shown in the following snippet:

[ ]:
```
def f(x, a, b):           # Define a function `f` which `returns` a value of a*x+b
    return a*x + b        # This line makes sure that the function returns the above␣
 ↪value

par, covariance = curve_fit(f, xpoints, ypoints)

print("Variable par is this array:",par)
print("")
print("The slope is:",par[0], " and the intercept is: ",par[1])

m = np.round(par[0],3)    # We use the `numpy.round` function to round to 3 decimal␣
 ↪places
c = np.round(par[1],3)
```

The line `par, cov = curve_fit(f, xpoints, ypoints)` might require some explanation: the `curve_fit` function returns **two values** by default, an array with the *parameters*, which we have called `par` here, and an array with the *covariance* which is a statistical concept that you do not need to interest yourself with now.

All you need to know is that `par[0]` is the value of `a` (the slope of your line) and `par[1]` is the value of `b` (its intercept).

Let's use this now to create an array of "theoretical" y values, using the formula $y(x) = mx + c$ – a straight line.

[ ]:
```
ytrend = m*xpoints+c      # Creates an array of y values corresponding to the␣
 ↪xpoints,
                          # which satisfy the trendline given by the parameters in␣
 ↪[par]

plt.plot(xpoints, ytrend, '--',color="red") # Plot a red dashed line of ytrend vs.␣
 ↪xpoints
plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')
plt.show()
```

Now, we can simply *combine* the above two graphs, as shown below.

You can also add an equation using the `plt.text` command, which allows you to add a string of text (here, the variable `eqn`) at specified coordinates (in our case, $x = 4, y = 14$), with font size 12pt. If you're plotting a straight line, you don't need to change `eqn`, it takes the value of the slope and intercept defined earlier, converts them into strings and adds them to the graph. (Placing `{c:+}` simply states that if `c` is a positive number, it is made into a string `"+c"` instead of just `"c"`(see **here** for more).

```
[ ]: plt.scatter(xpoints,ypoints)                          # Plotting the data-points
     plt.plot(xpoints, ytrend, '--',color="red")           # Plotting the trend-line

     slope_string = round(m,3)                             # Rounding off the slope to 3␣
      ↪digits
     intercept_string= round(c,3)                          # Rounding off the intercept to 3␣
      ↪digits

     eqn = 'y(x) = '+f'{slope_string}'+'x'+f'{intercept_string:+}'  # The equation string

     plt.text(1.5, 36,eqn,fontsize=12)                       # Displaying the above string


     plt.xlabel(r'Voltage (V)')                            # Formatting the axes
     plt.ylabel(r'Current ($\mu$ A)')
     plt.show();
```

FIN

Slightly more advanced topics

Two-dimensional arrays *(optional)* Sometimes you might want to include your data as an array of (x, y) points, for example something that looks like:

```
z = [[ 1.  3.],[ 2.  5.],[ 3.  6.],[ 4.  7.],[ 5. 10.],[ 6. 13.],[ 7. 15.],[ 8. 16.],[
9. 19.],[10. 20.],[11. 22.],[12. 24.],[13. 26.],[14. 29.],[15. 31.],[16. 33.],[17.
35.],[18. 36.],[19. 39.],[20. 41.]]
```

In this case, to plot it, you would first have to take all the xs, and then all the ys. We do this using the highly useful : symbol as follows:

```
[ ]: z = np.loadtxt("../data/exp1_ohm.csv",delimiter=",")

     print("The 2D array z:")
     print(z)

     xpoints = z[:,0]

     print()
     print("The slice of x values:")
     print(xpoints)
```

The : symbol here acts like a **wildcard**. Suppose our array z looks as follows:

```
[[ 1.  3.]  [ 2.  5.]  [ 3.  6.]  [ 4.  7.]  [ 5. 10.]  [ 6. 13.]  [ 7. 15.]  [ 8.
16.]  [ 9. 19.]  [10. 20.]  [11. 22.]  [12. 24.]  [13. 26.]  [14. 29.]  [15. 31.]
[16. 33.]  [17. 35.]  [18. 36.]  [19. 39.]  [20. 41.]]
```

The : (called the *slicing* operator) gives us a 'slice' of the data: **all values in the column on the left**. (This is Column "0". Column "1" would be the one on the right.)

Exercise 7: ***

Complete the following cell: Write down a line of code to collect the y points and create an array `ypoints`.

```
[ ]: # SOLUTION:

     # Write down a line of code to collect the y points and create an array `ypoints`
```

```
z = np.loadtxt("../data/exp1_ohm.csv",delimiter=",")

xpoints = z[:,0]
ypoints = z[:,1] # Insert ONE line of code here to get the ypoints data
```

For those interested in more advanced fits *(optional)* In order to plot a polynomial trendline, we will evaluate this (polynomial) function at the different `xpoints` and plot a dotted line over the earlier graph.

In order to do this, we will be using the `numpy.polyval` function which accepts two variables – an array of coefficients ($[a_n, a_{n-1}, ..., a_2, a_1, a_0]$) and a value of x – and prints out $y(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_2 x^2 + a_1 x + a_0$.

Thus, calling `np.polyval([a,b],x)` would give you $y(x) = ax + b$. If we used an *array* of x points, we'd get an *array* of y values!

We'll try to plot the line $y(x) = -x + 4$ to see how it works. The parameters will be `[-1,4]`. The resulting array will contain the values of y at these points.

[ ]:
```
xvalues = [0,1,2,3,4,5]                # A sample array of x values
yvalues = np.polyval([-1,4],xvalues) # y values corresponding to -1*x + 4

print(yvalues)                          # Printing out the yvalue array

plt.plot(xvalues,yvalues,'--')          # Plotting yvalues vs. xvalues in a line plot,
                                        # with a dashed line (given by '--')
plt.show();
```

Using np.genfromtxt *(very optional)* Apart from the simple `loadtxt` function, `NumPy` also has a slightly more powerful `genfromtxt` function which allows you to deal with `.csv` files that have missing values and so on. However, at your level, I think there is no difference at all between the two, and `loadtxt` is easier to remember!

You can use the `unpack` function here as well, but for illustrative purposes, I have used the slicing operator `:`.

[ ]:
```
data = np.genfromtxt("../data/exp1_ohm.csv", delimiter=",") # Looks in the current␣
↪directory
                                                            # for the csv (or txt) file and
                                                            # imports it, with commas being
                                                            # treated as delimiters.


xpoints = data[:,0]                                         # Get the first column of data,
                                                            # saving it to xpoints


ypoints = data[:,1]                                         # Idem for second column and␣
↪ypoints.

plt.scatter(xpoints,ypoints)
plt.xlabel(r'Voltage (V)')
plt.ylabel(r'Current ($\mu$ A)')
plt.show()
```

Appendix 2

B) Linear Regression Code

Experiment n - NAME

Date:

Data taken by:

Equipment used:

1. Instrument 1 (LC: 0.1mm)
2. Instrument 2 (LC: 0.2V)
3. ...

[ ]:
```
################### INPUT YOUR DATA HERE ####################

data_file = "../data/exp_n_data_1.csv"                # Add the name of the data␣
 ↪file (csv or txt).

x_label = r'x-axis (unit - e.g. $\mu$ A)'             # X axis labels
y_label = r'y-axis (unit)'                            # Y axis labels

round_slope = 3                                       # Number of digits to␣
 ↪round-off slope
round_intercept = 3                                   # Number of digits to␣
 ↪round-off intercept

#Once this is done, run the rest of the code: it should print a well formatted␣
 ↪graph.

#############################################################
```

[ ]:
```
import numpy as np                                    # Importing the NumPy package

import matplotlib.pyplot as plt                       # Importing the Matplotlib␣
 ↪package for plotting
                                                      # "Magic" to display images␣
 ↪inline
%matplotlib inline

import scipy as scp                                   # Importing the SciPy package
from scipy.optimize import curve_fit                  # Importing the curve fitting␣
 ↪module from SciPy
```

[ ]:
```
x,y= np.loadtxt(data_file, delimiter=",",unpack=True) # Imports the data located in␣
↪`data_file`,
                                                      # unpacking it such that the␣
↪first column is
                                                      # stored in the variable␣
↪`xpoints`, and the
                                                      # second in the variable␣
↪`ypoints`.
```

[ ]:
```
def f(x, a, b):                                       # Define a function `f` which␣
↪`returns` a value of a*x+b
    return a*x + b                                    # This line makes sure that␣
↪the function returns the above value

par, covariance = curve_fit(f, x, y)

m = np.round(par[0],round_slope)                      # For simplicity, we assign␣
↪the values of par[] to variables m
c = np.round(par[1],round_intercept)                  # and c, rounded off␣
↪approrpiately using the np.round() function.
```

[ ]:
```
ytrend = m*x+c                                        # Create an array of y values␣
↪corresponding to the xpoints,
                                                      # which satisfy the trendline␣
↪with the given slope (m) and intercept (c)


##########  Displaying a trendline on the graph area  ############

eqn = 'y(x) = '+f'{m}'+'x'+f'{c:+}'                   # Equation of trendline as a␣
↪string; we'll print this on the graph

xmin = np.min(x)*1.1                                  # Coordinates to place the␣
↪trendline: I have chosen to place it near the
ymax = np.max(y)*0.9                                  # minimum x-coordinate and␣
↪the maximum y-coordinate. You could use any two
                                                      # points like (14,2), for␣
↪example.

plt.text(xmin, ymax,eqn,fontsize=12)                  # Adding the equation string␣
↪to the graph area, at xmin and ymax

#############################################################

plt.scatter(x,y)                                      # Plotting the data-points
plt.plot(x, ytrend, '--',color="darkorange")          # Plotting an orange␣
↪trend-line

plt.xlabel(x_label)                                   # Formatting the axes
plt.ylabel(y_label)
plt.show();
```

Appendix 3

More Advanced Error Analysis

3.1 Objectives

1. To understand the concept of the random error.
2. To study the propagation of errors from measured to derived quantities.

3.2 Introduction

In the first chapter on errors, in the section on errors on time measurement, we found that the error on time measurement in our lab is almost certainly dominated by the reaction time of the observer. And, to determine this error, one method we suggested was to determine the time period many times and find the spread in the readings, using a measure called the standard deviation. The standard deviation is used in statistics to measure the characteristic spread of a random variable. The question then arises: What is a random variable? We will investigate this question in this chapter.

The other matter we will address ourselves to in this chapter is that of propagation of errors. You will have noticed that in the earlier chapter on errors, we spoke at length on the errors on the observations themselves, e.g. the length l, or obtained by a simple operation from the observations, e.g. the time period T, but we said nothing about how this may affect quantity derived from these observations, e.g. the acceleration due to gravity g. It is clear that how certain we are of the time period and the length will affect how certain we are of g. In other words the errors on l and T will *propagate* to g. How error propagates from observed quantities to derived quantities is the second subject of this chapter.

3.3 Random Errors

To understand this, let us imagine trying to make a large number of tiles, by a manually operated machine. It is clear that though the tiles will be similar to each other, no two will be identical. The differences will be of various kinds, appearance, mass, edge length, etc. Since we want to look at the tiles through the eye of a physicist, let us concentrate on measurable quantities; we could choose edge length or mass; let us choose mass.

The mass of a tile is what we call a *random variable*, because the mass of one tile is different from that of another in a random manner (in spite of the template which tends of make them similar). If we want to find the characteristic mass of the tiles, we can weigh a number of them and find the average. However, what we are after is how different the tiles are from each other; so the mean won't do. What we could do, to begin with, is weigh a large number of them, and note them down. If we look carefully at the numbers, we will see that there is a spread. We see that numbers close to each other don't indicate much of a difference, and we'd like to think of them together. A powerful way to represent neighbouring readings together while highlighting the difference between distant readings is to draw a histogram.

3.3.1 Histogram

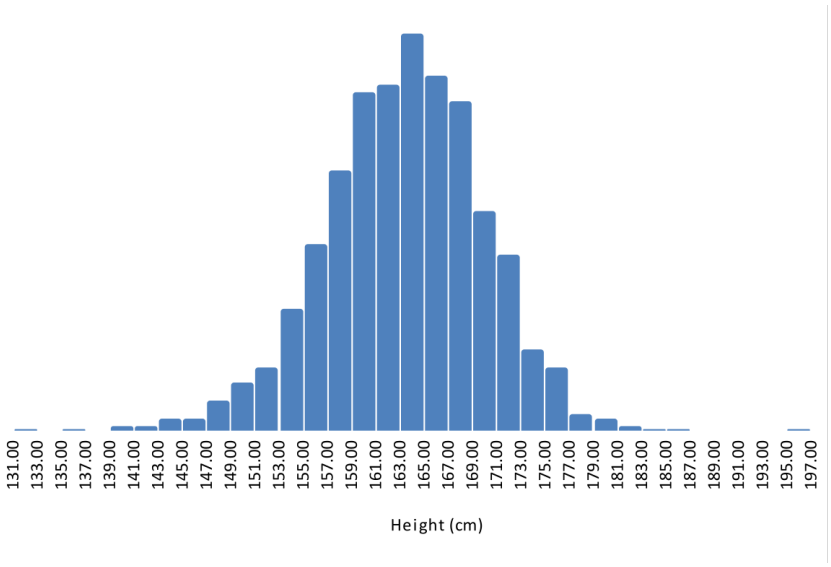


Figure 3.1: A typical histogram, for 1200 readings of the heights of individuals. The bell-like shape of the pattern is characteristic. It is described by a famous mathematical function called the Gaussian.

In a histogram, we put the readings in a number of bins, i.e. we put all the readings in a certain range in the same bin, and we have a number of such bins to cover all the readings. There is no formula for the number of bins, but what we want a large-enough number of bins to see how the population differs from that in the others as we move across the readings, i.e. we want to see the

pattern in the variation of the readings.

What we find is that whenever the variation from one reading to another is caused by a large number of factors unconnected with each other, this shape arises naturally. This is true in our example of the masses of a collection of tiles made with the same template. It is also true of the masses of group of human beings belonging to the same population, e.g. urban Punjabis.

It is clear from the way we have arrived at the histogram that its width is determined by the spread of the readings. The mean does not contain information on this width; in fact what we are looking for is precisely how far the readings are, on the average, from the mean. If we imagine two artisans making the tiles, one a master who is able to achieve greater consistency and the other an apprentice, it ought to be clear that the histogram of the masses of the master's tiles will be narrower than that of the apprentice's tiles.

One way to characterise the spread of the masses might be to find the difference in mass between the heaviest tile and lightest tile. But there is something unsatisfactory about this because we are allowing the outliers to characterise and entire. What we would like to do is use all the readings to characterise the spread. This is done by the *standard deviation* σ, which is given by the formula

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}, \tag{3.1}$$

where x_i is a reading, $\overline{x}$ is the mean, and n is the number of readings.

For a Gaussian distribution of readings, i.e. a distribution the envelope of whose histogram is a Gaussian, about 68% of the readings are found between $x = \overline{x} - \sigma$ and $x = \overline{x} + \sigma$. Thus an arbitrary reading has a 68% likelihood that an individual measurement will fall within one standard deviation ($\pm\sigma_x$) of the mean.[1]

Question: In the last paragraph we have gone from the fraction of readings in a certain range to a probability of an arbitrary reading's having a certain value. Go through this carefully and understand the argument.

Question: Do some research to find out what fraction of the readings are found in the range $\pm 2\sigma$ and $\pm 3\sigma$ from the mean.

When the significant source of error is considered to be random and the distribution is a Gaussian, random error associated with the reading is some factor times σ, the factor depending on how much confidence we wish to associate with our reading. If we choose σ as our error – this is a common choice – the confidence associated with the reading is 68%.

[1] Keep in mind, however, the example of superluminal neutrinos we gave earlier: in that case, the error was *assumed* to be random, and their precise data led them to a false positive. The true error turned out to be systematic, leading to *different (and larger) error bars*!

A note on statistics: Consider two sets of measurements of the acceleration due to gravity:

S. No.	$\boldsymbol{g}$ (ms^{-2})
1	9.80 ± 0.01
2	0.70 ± 0.01
3	18.90 ± 0.01
4	15.60 ± 0.01
5	4.10 ± 0.01
Mean	9.82 ± 0.01

Table 3.1: Measurement of g (Set 1)

S. No.	$\boldsymbol{g}$ (ms^{-2})
1	9.83 ± 0.01
2	9.80 ± 0.01
3	9.82 ± 0.01
4	9.84 ± 0.01
5	9.83 ± 0.01
Mean	9.82 ± 0.01

Table 3.2: Measurement of g (Set 2)

It would obviously be wrong to go simply by the mean and say that both these sets of data were equally reliable. In fact, stating the mean of the first set as the acceleration due to gravity **doesn't make sense**.

Suppose we compared the mean of a data set to its *dispersion* about the mean. If this is small, we could then say that there is some "true" value and that all the different values we measured occurred due to random fluctuations about this "true" value. Since the fluctuations are random, we could assume that they average out to zero, leaving us with a closer estimate to the "true" value than any individual reading.

Question: In above sets of data (with standard deviations of 7.6 ms^{-2} and 0.02 ms^{-2} respectively), does the quoted error of ± 0.01 ms^{-2} make sense

1. For Set 1?
2. For Set 2?
3. For Both?
4. For Neither?

Justify your answer quantitatively.

The two errors in the lab you will encounter that you cannot remove are **least count errors** and **random errors**. Remember to always compare them and ***take the larger value***. In Set 1 it makes no sense to say g is specified to $(9.82 \pm 0.01)\, ms^{-2}$, since all of the values are much farther away than that!

1. **Question:** Multiple measurements with the same instrument increases the
   (a) Accuracy
   (b) Precision
   (c) Both
2. **Question:** Consider the following data-table

S. No.	1	2	3	4	5
Time Period (s)	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1	1.2 ± 0.1

   The standard deviation is 0.0. Would it be right to say that $T_{\text{avg}} = 1.2 \pm 0.0$?

However, keep in mind there are many cases where the mean **does not** represent some true value.[2] There are also cases where the standard deviation contains physical information. Such cases are usually a result of statistical phenomena.

Tip

1. In a coin-toss experiment with a large number of tosses, σ gives you a measure of the bias of the coin.
2. In a random walk, σ can be a measure of the diffusion coefficient D.
3. In shot-noise – the statistical fluctuations of current due to the actual number of electrons flowing in the conductor per unit time – σ gives you a measure of Boltzmann's constant k_B.

3.3.2 Reporting Errors

Significant figures

The significant figures of a number are the digits in its representation that contribute to the precision of the number. In practice, we assume that all digits used to write a number are significant (except leading zeroes.[3])

Results of simple calculations should not increase the number of significant digits. Calculations transform our knowledge; they do not increase it! The rounding should be performed at the final step of a calculation to prevent rounding errors at intermediate steps from propagating through your work but **only** one or two extra digits suffice to prevent this.

[2] For example, even if the average number of siblings every student has is 1.574, there is no student who has a non-integer number of siblings.

[3] Non-leading zeros are considered to be significant. If you write a number as 1,200, we assume there are four significant digits. If you only mean to have two or three, then it is best to use scientific notation: 1.2×10^3 or 1.20×10^3. Leading zeros are not considered significant: 0.55 and 0.023 have just two significant figures.

If you measure a value on a two-digit digital meter to be 1.0 and another value to be 3.0, it is incorrect to say that the ratio of these measurements is 0.3333333. The two values are not exact numbers with infinite precision. Since they each have two significant digits, the correct number to write down is 0.33.[4]

Do not write significant figures beyond the first digit of the error on the quantity. Giving more precision than this to a value is not only irrelevant, *it is misleading.*

If you're told you're using FAR too many digits, please do not try to use the excuse, "That's what the computer gave me." **You** are in charge of presenting your results, not the computer!

Error propagation

Experiments worth carrying out rarely measure only one quantity. Typically we measure two or more quantities and then 'fold' them together in some equation or equations to determine some other quantity that we believe to depend on them. It is thus imperative that we understand how uncertainties in certain measured quantities "propagate" into other derived quantities.

For our analysis, let us assume that

- x and y are measured quantities with uncertainties δx and δy respectively. These errors are considered to be **uncorrelated**, meaning that δx and δy are independent.[5]
- c is a constant known to known absolutely precisely (or with negligible uncertainty).
- z is a quantity *derived* from x and y and possessing a "propagated" uncertainty δz.

The formulae used to compute errors are usually not completely understood: a variety of different formulae are used in different cases, and the reasons **why** are usually lost on students. It turns out that there is only **one** way to add (uncorrelated) errors, which one can manipulate to get the rest:

If $z = x + y$, their (uncorrelated) errors add **in quadrature**:

$$\delta z = \sqrt{(\delta x)^2 + (\delta y)^2} \tag{3.2}$$

The reason for errors adding in quadrature is one that comes from statistics; it is essential that the errors be **independent** of each other (uncorrelated).[6]

[4] If this is an intermediate result, then 0.333 or 0.3333 are preferred, but the final result must have two significant digits.

[5] For example, the precision of measuring the length of your simple pendulum has no effect on the precision of measuring time.

[6] It also makes sense: errors could be positive and some negative, simply adding them could conceivably give a smaller number (or even zero!). Thus, the next best thing is to add their *squares.*

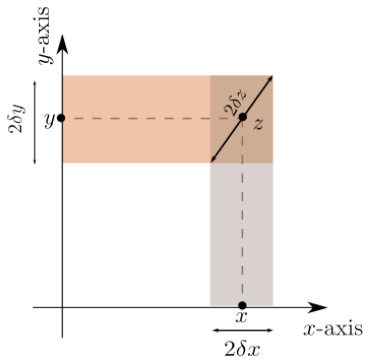


Figure 3.2: Two points $(x, 0)$ and $(0, y)$, with uncertainties δx and δy respectively, can be summed to get a point $z = (x, y)$ with uncertainty δz.

Here is another motivation:[7] you could imagine two uncorrelated measurements to represent two axes (x and y) that are orthogonal to each other (see Figure (3.2). Imagine that you want to specify a point $z = (x, y)$: given that there is an uncertainty in both the coordinates x and y, the point z is uncertain **at most** by $\delta z = \sqrt{(\delta x)^2 + (\delta y)^2}$.

This formula can then be used to get the uncertainties of more complicated relations. For example, consider $z = xy$. In this case, the two quantities are **multiplied**, and so we can't use the above formula as is. We could, however, take the log, and get $\log z = \log x + \log y$. This is of the form $u = v + w$. We could then apply Equation (3.2), and find that

$$\delta(\log z) = \sqrt{(\delta(\log x))^2 + (\delta(\log y))^2}$$

The uncertainty in $\log z$ can easily be related by taking the derivative.[8]

$$\mathrm{d}(\log z) = \frac{\mathrm{d}z}{z} \quad \Longrightarrow \quad \delta(\log z) = \frac{\delta z}{z}$$

Thus,

$$\frac{\delta z}{z} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2} \tag{3.3}$$

[7] This is **not** an explanation of why this is true!

[8] A differential is by definition the variation of function when its parameter changes by a small amount. We want to find how much $\log z$ changes when z changes by δz. We use δ instead of d, since the variation is not truly *infinitesimal*.

Tip

The following rules should exhaust most of the common cases that you will be exposed to in your undergraduate labs. Everything here can be generalised simply to a a *set* of measurements x_i with uncertainties δx_i.

1. **Addition or subtraction by a constant:** If $z = c \pm x$, then
$$\delta z = \delta x \tag{3.4}$$
2. **Multiplication by a constant:** If $z = cx$, then
$$\delta z = c\delta x \tag{3.5}$$
3. **Addition or subtraction of two measured quantities:** If $z = x \pm y$, then
$$\delta z = \sqrt{(\delta x)^2 + (\delta y)^2} \tag{3.6}$$
4. **Multiplication or division of two measured quantities:** If $z = xy$ or $z = \frac{x}{y}$, then
$$\frac{\delta z}{z} = \sqrt{\left(\frac{\delta x}{x}\right)^2 + \left(\frac{\delta y}{y}\right)^2} \tag{3.7}$$
5. **A measured quantity raised to a power:** If $z = x^c$, then
$$\frac{\delta z}{z} = c\frac{\delta x}{x} \tag{3.8}$$

Question: Prove Equation (3.8).

Question: If $z = x^2 = x \times x$, we get different answers if we use the Power Rule (Equation (3.8)) or the Product Rule (Equation (3.7)). Which of the two is correct? Why?

Question: Calculate the uncertainty in z if

1. $z = \frac{1}{x}$
2. $z = \frac{x}{1+x}$
3. $z = \frac{x}{x+y}$

Hint: The last one is slightly hard. You will first write it as $z = \frac{1}{1+u}$, where $u = y/x$. Then, calculate δz in terms of δu, and only then δu in terms of δx and δy.