

V. Precision and Accuracy

Experimental Uncertainties

If you want to get the most reliable value for a measured quantity, even something as simple as the width of this piece of paper, the best thing to do is to make many measurements and average them. All of your measurements will not be exactly the same, but they will cluster closely about an average value. The spread in the individual measurements and the number of measurements you make give an estimate of how well you know the value. For example, if I make one measurement of the width of this page, I have no idea how likely (with what probability) it is that I will get this answer in another measurement. As I make more measurements, my confidence in the results increases and I can get an idea of the probability of getting a particular result. If I divide the number of times I get a particular answer by the total number of trials, I get a probability for each value. The spread in the answers is determined by the *precision* of the experiment. The measure of the precision is called the *uncertainty*. Precise experiments have small uncertainties. When a series of individual readings are made under identical conditions, we report the average value, \bar{x} , and an estimate of the uncertainty called the *standard deviation*, $\Delta\bar{x}$. When n readings of a quantity x are made, the result is given as $x = \bar{x} \pm \Delta\bar{x}$, where the average is

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} \quad (\text{V.1})$$

and the standard deviation is given by

$$\Delta\bar{x} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}} \quad (\text{V.2})$$

In these expressions, $x_1, x_2, x_3, \dots, x_n$, represent the values measured in your experiment. The differences between the individual measurements and the average, $(x_i - \bar{x})$, are called the deviations. Because some of these are positive and some are negative (from the definition of the average), we square them so we have a positive measure of the deviation. The standard deviation is a kind of average of these square deviations. For technical reasons the denominator is the one fewer than the number of measurements. Finally, we take the square root to get the standard deviation, $\Delta\bar{x}$. This will have the same units as the measured quantity.

In general, all quantities are to be measured several times so that the uncertainty can be

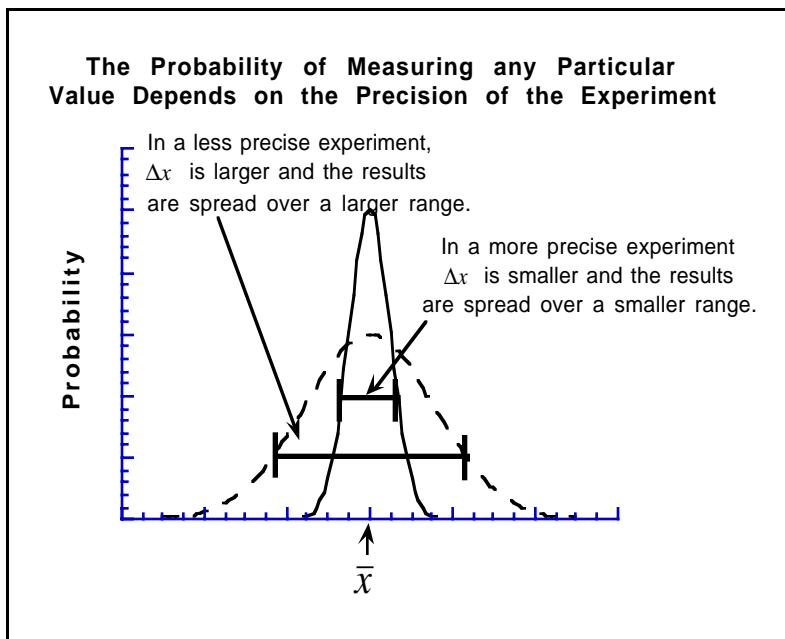


Figure V.1 Schematic representation of measurements randomly distributed about the average.

estimated. In most cases, five measurements should be done and the standard deviation used as this estimate. In some cases more or fewer will be required and will usually be stated in the project description. When in doubt, consult your laboratory instructor.

This method applies to situations where the precision in the measurement is limited by something other than the instrument. For example, if I use a stopwatch to time the oscillations of a pendulum, the precision is not determined by the accuracy of the timer, but by the technique (timing one or many oscillations), the construction of the pendulum (does it not swing long enough to get a good measurement?), and the period (is it too short to measure?) among other things. In this case, repeating measurements can average out some of these uncertainties. This also applies to distance measurements when using a secondary standard or when the meter stick or ruler must be moved. If you measure the distance with a string or a bicycle wheel and then measure that to get the data, repeated measurement is necessary. Also, if you measure a distance by moving the meter stick, the alignment errors are random and repeated measurement will let you average these out.

There are measurements for which repeating the process gives little or no information about the uncertainties. This occurs when using a voltmeter or ammeter and may occur when using a meter stick. When using an analog meter (one with a scale and a pointer), the uncertainty is determined not by the reproducibility of the measurement but by your ability to read the instrument. If, for example, you are using a voltmeter on a scale which has smallest divisions of 0.2 volts, then a reasonable estimate of the uncertainty in your data is ± 0.1 volts or $1/2$ the smallest marked division. This is different from the discussion above about standard deviations because it is the precision with which the meter can be *read* that is the limiting factor, not the ultimate precision of the meter. This situation is often encountered when using a meter stick. It may be that the marks on the stick are very precisely positioned, implying that we'd know the distance quite accurately. However, we can only record the distance as accurately as we can *read* the scale. Thus, when measuring distances shorter than the ruler, the limit on the precision of the measurement is the size of the smallest division on the scale.

When using a digital meter, the uncertainty is given by the number of significant figures displayed. Suppose your meter reads 1.275 volts. Here you are given data with four significant figures. Since we don't know the fourth place to the right of the decimal (the meter doesn't supply it), we can only assume that the uncertainty is no larger than ± 0.0005 volts or ± 5 in the first unspecified decimal place.

In summary then, uncertainties are found as follows:

Table V.1 Methods of estimating uncertainties.

ability to read instrument more precise than method	All quantities are to be measured several times so that the uncertainty can be estimated. In most cases, five measurements should be done and the standard deviation used as this estimate. In some cases more or fewer will be required and will usually be stated in the lab instructions. When in doubt, consult your laboratory instructor.
analog meter or unmoved meter stick (ability to read the meter less precise than the method)	The standard deviation is taken to be equal to $1/2$ of the smallest division on the scale in use.
digital meter	The standard deviation is taken to be ± 5 in the first unspecified decimal place.

In lecture you learn that there are two equally important portions of an answer: the value and the units. When you solve a problem in the text book and report that a force is 3.0, this doesn't tell you the answer. If you report the value and the units so that the answer is, say, 3.0 Newtons, then the reader knows what your answer means. The reported quantity isn't just math anymore, it conveys physical information. If either of these is missing, the answer is wrong. When doing experimental science (in any field!), there is a third, equally important component: the uncertainty. When you report an experimental result, a statement of the precision is required. The measurement of a distance with a string that stretches will be less certain than a distance measured with a steel tape. **These uncertainties are not errors in the usual sense!** A measurement can be made twice with no mistakes and still give different results. These

uncertainties are a measure of the precision of the experiment. Ultimately, these differences arise from random fluctuations in the underlying physical processes. For example, suppose I want to measure the time required for a pendulum to make 10 oscillations. If I do this several times, then on average, I'll record a reasonably accurate time. But any one of my measurements could be different from the average because estimating when the pendulum has completed exactly 10 oscillation is very hard. But if I'm careful and repeat the experiment several times, I can have some assurance that any bias from this effect will average out, giving me an accurate result. Nevertheless, the precision of the experiment is determined by the spread (standard deviation) in the individual answers. Reporting the uncertainty gives others an indication of how well the measured quantity is determined. As will be discussed below, the *precision* of an experiment is different from the *accuracy*.

Absolute and Relative Uncertainties

Any measurement you make has an uncertainty (precision) associated with it. There will be an uncertainty even in a single measurement but we don't usually know what it is. Such an uncertainty can be measured and reduced by repeating measurements and applying the methods of the previous section. The measurement and its uncertainty are stated as $\bar{x} \pm \Delta\bar{x}$. Typical results of measurements might be a length of (6.37 ± 0.02) cm, a temperature of $(47 \pm 1)^\circ\text{C}$, or a mass of (487.2 ± 0.7) g.

If the result of a length measurement of a particular object is reported as (2.08 ± 0.05) m, we mean that the uncertainty in the length is 0.05 m. It is common to refer to this uncertainty as the *absolute uncertainty* in the measurement. The spread in values (± 0.05 m) is the range within which other measurements most probably fall. The uncertainty is absolute in the sense that it is stated in the same physical units as the quantity measured.

The *relative* or *fractional uncertainty* and *percent uncertainty* are also useful ways to express the uncertainties. These are defined and illustrated in table V.2.

Table V.2 Definition of absolute, relative and percent uncertainty.

Quantity	Definition	Example if $x = (2.08 \pm 0.05)$ m, then
Absolute Uncertainty	$\Delta\bar{x}$ of the individual measurements	0.05 m
Relative or Fractional Uncertainty	$\frac{\Delta\bar{x}}{\bar{x}}$	$\frac{0.05\text{m}}{2.08\text{m}} = 0.02$
Percent Uncertainty	$\frac{\Delta\bar{x}}{\bar{x}} \times 100\%$	$\frac{0.05\text{m}}{2.08\text{m}} \times 100\% = 2\%$

Usually when reporting a measured value, we give the absolute uncertainty. As you will see in the next sections, relative uncertainties are important when multiplying or dividing two measured quantities and percent uncertainties are most useful when assessing the accuracy of the experiment.

Propagation of Uncertainties

Knowing the precision of a measured quantity is often only the beginning. Usually, the quantity we are after requires that several measured quantities be combined to give the desired result. To understand how to find the uncertainty in a calculated result, we must recall what the uncertainties mean. If we were to simply add up the uncertainties, we would overestimate the total uncertainty since it is unlikely that all measurements are off in the same direction by the maximum amount. So that the uncertainty in the sum has the same meaning as the individual uncertainties, we must combine them in a way that is consistent with the definition of the standard deviation.

Often we must combine data with uncertainties together with numbers that have no uncertainties. For example, if we measure the period of rotation, T , of a mass moving in a circle with radius r , then the linear speed of the object is given by $v = \frac{2\pi r}{T}$. The 2π factor is an exact number with no

uncertainty. We must consider its effect on the uncertainty of the final answer as well. The rules for computing answers using the equations we find in physics are given in table V.3.

Table V.3 Rules for propagating uncertainties through a calculation. Numbers with no uncertainties are denoted by a , and b

Quantity	Value	Uncertainty
Average	$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots x_n}{n}$	$\Delta\bar{x} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots (x_n - \bar{x})^2}{n-1}}$
Sum	$\bar{z} = a\bar{x} + b\bar{y}$	$\Delta\bar{z} = \sqrt{(a\Delta\bar{x})^2 + (b\Delta\bar{y})^2}$
Difference	$\bar{z} = a\bar{x} - b\bar{y}$	
Product	$\bar{z} = a\bar{x}\bar{y}$	$\Delta\bar{z} = \bar{z}\sqrt{\left(\frac{\Delta\bar{x}}{\bar{x}}\right)^2 + \left(\frac{\Delta\bar{y}}{\bar{y}}\right)^2}$
Quotient	$\bar{z} = a\frac{\bar{x}}{\bar{y}}$	
Square root	$\bar{z} = a\sqrt{\bar{x}}$	$\Delta\bar{z} = \frac{1}{2}\bar{z}\left(\frac{\Delta\bar{x}}{\bar{x}}\right)$
Square	$\bar{z} = a\bar{x}^2$	$\Delta\bar{z} = 2\bar{z}\left(\frac{\Delta\bar{x}}{\bar{x}}\right)$

As an example, suppose that you have five measurements of the length, L , and width, W , of a board

Table V.4 Sample data set representing the measurement of the length and width of a board.

L	32.7 cm	33.0 cm	32.5 cm	33.1 cm	32.6 cm
W	15.7 cm	15.8 cm	15.5 cm	15.6 cm	15.7 cm

With this data set, the uncertainties for each type of calculation can be illustrated.

Averages and Uncertainties

We can find the average length of the board as

$$\bar{L} = \frac{32.7cm + 33.0cm + 32.5cm + 33.1cm + 32.6cm}{5} = 32.8cm \quad (\text{V.3})$$

and the uncertainty, $\Delta\bar{L}$, is found to be

$$\Delta\bar{L} = \sqrt{\frac{(32.7cm-32.8cm)^2 + (33.0cm-32.8cm)^2 + (32.5cm-32.8cm)^2 + (33.1cm-32.8cm)^2 + (32.6cm-32.8cm)^2}{4}} = 0.259cm \quad (\text{V.4})$$

When reporting an experimental result, the uncertainty should never be given to more than 1 significant figure so the length of the board would be reported as $L = (32.8 \pm 0.3)$ cm. See section V for a more detailed discussion of the relationship between uncertainties and significant figures.

Similarly, the average width, \bar{W} , of the board is

$$\bar{W} = \frac{15.7cm + 15.8cm + 15.5cm + 15.6cm + 15.7cm}{5} = 15.7cm \quad (\text{V.5})$$

and the uncertainty in the average is

$$\Delta \bar{W} = \sqrt{\frac{(15.7\text{cm}-15.7\text{cm})^2 + (15.8\text{cm}-15.7\text{cm})^2 + (15.5\text{cm}-15.7\text{cm})^2 + (15.6\text{cm}-15.7\text{cm})^2 + (15.7\text{cm}-15.7\text{cm})^2}{4}} = 0.122\text{cm} \quad (\text{V.6})$$

We would report the width of the board as $W = (15.7 \pm 0.1)$ cm.

Sum

To find the sum of these two numbers, we add the average values to get

$$\bar{S} = \bar{L} + \bar{W} = 32.8\text{cm} + 15.7\text{cm} = 48.5\text{cm} \quad (\text{V.7})$$

and the uncertainty in this quantity is

$$\Delta \bar{S} = \sqrt{(\Delta \bar{W})^2 + (\Delta \bar{L})^2} = \sqrt{(0.1\text{cm})^2 + (0.3\text{cm})^2} = 0.3\text{cm} \quad (\text{V.8})$$

giving $S = (48.5 \pm 0.3)$ cm. Note that in this example both a and b in table V.3 are equal to 1.

As an example of computing the uncertainties when a and b are not equal to unity, consider case of computing not simply the sum of the length and width but the perimeter of the board. This is given by

$$\bar{P} = 2\bar{L} + 2\bar{W} = 2 \cdot 32.8\text{cm} + 2 \cdot 15.7\text{cm} = 66.2\text{cm} \quad (\text{V.9})$$

In this case, a and b both equal 2. Now the uncertainty is given by

$$\begin{aligned} \Delta \bar{P} &= \sqrt{(2\Delta \bar{W})^2 + (2\Delta \bar{L})^2} = \sqrt{4(0.1\text{cm})^2 + 4(0.3\text{cm})^2} = \\ &= 2\sqrt{(0.1\text{cm})^2 + (0.3\text{cm})^2} = 0.6\text{cm} \end{aligned} \quad (\text{V.10})$$

Difference

If we wanted to compute the difference, we would get

$$\bar{D} = \bar{L} - \bar{W} = 32.8\text{cm} - 15.7\text{cm} = 17.1\text{cm} \quad (\text{V.11})$$

and for the uncertainty, we get

$$\Delta \bar{D} = \sqrt{(\Delta \bar{W})^2 + (\Delta \bar{L})^2} = \sqrt{(0.1\text{cm})^2 + (0.3\text{cm})^2} = 0.3\text{cm} \quad (\text{V.12})$$

exactly the same as for the sum. Here, the reported value is $D = (17.1 \pm 0.3)$ cm.

Product

To compute the area of the board, we get

$$\bar{A} = \bar{L} \cdot \bar{W} = 32.8\text{cm} \cdot 15.7\text{cm} = 515\text{cm}^2 \quad (\text{V.13})$$

and for the uncertainty we have

$$\Delta\bar{A} = \bar{A} \cdot \sqrt{\left(\frac{\Delta\bar{W}}{\bar{W}}\right)^2 + \left(\frac{\Delta\bar{L}}{\bar{L}}\right)^2} = \bar{A} \cdot \sqrt{\left(\frac{0.1\text{cm}}{15.7\text{cm}}\right)^2 + \left(\frac{0.3\text{cm}}{32.8\text{cm}}\right)^2} \quad (\text{V.14})$$

$$\Delta\bar{A} = \bar{A} \cdot 11 \times 10^{-3} = 515\text{cm}^2 \cdot 11 \times 10^{-3} = 5.7\text{cm}^2$$

So, for the area we have as the final answer $A = (515 \pm 6) \text{cm}^2$.

Quotient

If we wish to compute the ratio, then we have

$$\bar{R} = \frac{\bar{L}}{\bar{W}} = \frac{32.8\text{cm}}{15.7\text{cm}} = 2.09 \quad (\text{V.15})$$

The uncertainty in the ratio is

$$\Delta\bar{R} = \bar{R} \cdot \sqrt{\left(\frac{\Delta\bar{W}}{\bar{W}}\right)^2 + \left(\frac{\Delta\bar{L}}{\bar{L}}\right)^2} = \bar{R} \cdot \sqrt{\left(\frac{0.1\text{cm}}{15.7\text{cm}}\right)^2 + \left(\frac{0.3\text{cm}}{32.8\text{cm}}\right)^2} \quad (\text{V.16})$$

$$\Delta\bar{R} = \bar{R} \cdot 11 \times 10^{-3} = 2.09 \cdot 11 \times 10^{-3} = 0.02$$

In this case the final answer is $R = (2.09 \pm 0.02)$. Note that even though the equations for the relative uncertainties in a product and quotient are exactly the same, the absolute uncertainty depends on the value of the product or quotient. The absolute uncertainty will always have the same units as the value.

Square Root

Suppose we wished to find the length of the side a square that would have the same area as the rectangular board we measured. This is found by computing the square root of the area of the board.

$$\bar{l} = \sqrt{\bar{A}} = \sqrt{515\text{cm}^2} = 22.7\text{cm} \quad (\text{V.17})$$

The uncertainty in this length is

$$\Delta\bar{l} = \frac{1}{2}\bar{l} \frac{\Delta\bar{A}}{\bar{A}} = \frac{22.7\text{cm} \cdot 6\text{cm}^2}{2 \cdot 515\text{cm}^2} = 0.1\text{cm} \quad (\text{V.18})$$

so that the final answer we report is $l = (22.7 \pm 0.1) \text{cm}$.

As suggested in this simple example, it is possible to combine the simple cases of addition, subtraction, multiplication, division and square roots to construct more complicated formulas and propagate the uncertainties along with the values.

Square

Suppose we wished to find the area of a square board that has a side of length L . This is found by computing the square of the length of board.

$$\bar{A} = \bar{L}^2 = (32.8\text{cm})^2 = 1075.84\text{cm}^2 \quad (\text{V.17})$$

The uncertainty in this length is

$$\Delta\bar{A} = 2\bar{A} \frac{\Delta\bar{L}}{\bar{L}} = 2 \cdot 1075.84\text{cm}^2 \frac{0.3\text{cm}}{32.8\text{cm}} = 19.68\text{cm} \quad (\text{V.18})$$

To get the correct number of significant figures we must convert this result to scientific notation. Thus, the final answer we report is $\bar{A} = (1.08 \times 10^3 \pm 2 \times 10^1)\text{cm}^2$. Notice that in the rules for propagating the uncertainties for products, quotients, squares roots, and squares the constants a and b do not appear. This is because they are already included in the calculation of \bar{z} . Also notice that the uncertainty in the square of a number is not the same as the uncertainty you would get by multiplying a number by itself.

Comparing Results

Estimating and carrying the experimental uncertainties through your calculations gives you an estimate of the precision of the experiment. This reflects, among other things, how carefully the measurement was done. Sometimes it is a reflection of the quality of the equipment, although this is not the case in this lab. Sometimes it is dependent on the general method being used. For example, you might expect that measuring the distance between two buildings on campus would be more precise if a two-meter stick were used rather than a 1 cm stick. But in neither case have we learned anything about whether the answer we obtain is correct. Very precise measurement can be very wrong and imprecise measurements can be very accurate (correct). How do we estimate the accuracy of the experiment?

Often the measured quantity is a well known physical constant, such as the acceleration due to gravity or the speed of sound, and you wish to compare your result with this accepted value. Other times you have made two independent sets of measurements of a particular quantity and you wish to compare them. In the first case, one answer, the accepted value, is believed to be correct, whereas in the other, either may be correct. To compare your measured values with the accepted values, compute the *percent error*. This is given by

$$\text{Percent Error} = \frac{|\text{Measured} - \text{Actual}|}{\text{Actual}} \times 100\% \quad (\text{V.19})$$

The vertical bars indicate that you take the absolute value of the difference and ignore the algebraic sign. If your experiment measures the value of the acceleration of gravity, a comparison with the accepted value would give information about the *accuracy*. The uncertainty that you compute by repeating the measurement several times, gives the *precision* in the value. **When an experiment is done properly, the percent uncertainty in the measurement (the precision) is usually larger than or equal to the percent error (accuracy).** Precise measurements should have good accuracy and imprecise measurements may be more in error. If your percent error is larger than the percent uncertainty, this could suggest that your instrument is calibrated incorrectly, you made a calculation mistake, used a poor method, or have not been entirely objective about estimating your uncertainties.

For example, suppose that I measure the time it takes the moon to make one revolution around the earth. After several such measurements, I get (27.1 ± 0.3) days. This is, however, a well known quantity, having a value of 27.32118 days. We can assess the accuracy of our experiment by computing the percent error and the percent uncertainty and comparing them. The percent uncertainty is

$$\frac{0.3days}{27.1days} \times 100\% = 0.01 \times 100\% = 1\% \quad (V.20)$$

and the percent error is

$$\begin{aligned} \frac{|27.1days - 27.32118days|}{27.32118days} \times 100\% &= \frac{|-0.221days|}{27.32118days} \times 100\% \\ &= \frac{0.221}{27.32118} \times 100\% = 0.008 \times 100\% = 0.8\% \end{aligned} \quad (V.21)$$

Here we see that the percent error is smaller than the percent uncertainty and our confidence in the measurement is justified.

In the case where you want to compare two different measurements of the same quantity, use the *percent difference*. This is found by computing the difference between the two values and dividing by the average.

$$\text{Percent Difference} = \frac{|1^{st} \text{ measurement} - 2^{nd} \text{ measurement}|}{\left(\frac{1^{st} \text{ measurement} + 2^{nd} \text{ measurement}}{2}\right)} \times 100\% \quad (V.22)$$

Suppose that I make two sets of measurements to determine the diameter of a wheel. In the first set I find the average diameter directly with a ruler and get $d=(3.7 \pm 0.1)$ cm. For the second measurement I roll the wheel, measure the distance it travels in one revolution and use the relationship between circumference and diameter to compute the diameter of the wheel. This gives $d=(3.8 \pm 0.2)$ cm. To assess the accuracy of the measurements I compute the percent uncertainties of both measurements

$$\frac{0.1cm}{3.7cm} \times 100\% = 3\% \qquad \frac{0.2cm}{3.7cm} \times 100\% = 5\% \quad (V.23)$$

Then I find the percent difference

$$\frac{|3.8cm - 3.7cm|}{\left(\frac{3.8cm + 3.7cm}{2}\right)} \times 100\% = \frac{0.1cm}{3.75cm} \times 100\% = 3\% \quad (V.24)$$

In this case, we compare the percent difference to the larger of the two percent uncertainties since this will be the limiting factor in the accuracy. Thus, we see that the estimates of the uncertainty are commensurate with the accuracy of the experiment. By measuring the same quantity in two (presumably independent) ways, the overall accuracy of the experiment is assessed by comparing the percent difference to the percent uncertainties. If the percent difference is larger than the percent uncertainty, you cannot tell from this one number which measurement is wrong or if they are both wrong. Usually the only recourse is to repeat the measurements, check your calculations for errors and scrutinize the uncertainties to be sure that they are reasonable. The percent error and/or percent difference should be computed, as appropriate, for each measured quantity.

VI. Significant Figures

In mathematics, when we give a numerical value for a result, it is exact. If the result of a particular calculation is 10.2, then a mathematician would say that we were free to add zeros to the right as we like without fear of changing the meaning of the answer. In physics, this is not the case. Here, numerical quantities have three parts: the value, the units, and the precision with which it is known. The value and the units are clear when we give a result such as 10.2 cm. The precision of the result is expressed in the number of digits quoted. By giving the result as 10.2 cm, we mean a length that is closer to 10.2 cm than to either 10.1 cm or 10.3 cm. If we had used a ruler with finer division, we might have said the length was 10.23 cm. This means a length closer to 10.23 cm than to 10.22 cm or 10.24 cm, and so on. Thus, we see that the number of digits quoted in the result carries information about how precisely we have made the measurement or calculation. Unlike the case of a numerical result in mathematics, the result of a physical measurement is stated with a number of digits that gives information about how the number came to be known. The number of digits reported in a measurement, irrespective of the location of the decimal place, is called the number of *significant figures*. In classical (as opposed to quantum) physics, distances are infinitely divisible. This means that to specify a distance "exactly" would require an infinite number of significant figures which we can never have. Therefore, all numbers that represent measurements necessarily carry with them information about how well that number is known. **Every value in physics is ultimately the result of a measurement or derived from one, hence every value has a finite precision.**

These measured values with finite precision are often used to calculate other quantities. For example, suppose that we measure the length of the table top to be 155.6 inches and we want the answer in centimeters. From the definition of centimeters and feet, we know that there are *exactly* 2.54 cm/in (This is not a measured value but a defined one and can be exact.). If we blindly enter these numbers into our calculator and do the math we get

$$L = \frac{155.6 \text{ inches}}{2.54 \text{ cm/in}} = 61.25984252 \text{ cm} \quad (\text{VI.1})$$

However, we are doing physics, not math and to a physicist this is not a meaningful answer. The distance cannot be known in centimeters with a greater precision than it was originally determined in inches. In this case, the precision is indicated by the fact that the length is given with four significant figures. So that the computed answer is given with a precision that reflects the precision of the data on which it is based, we round off the final answer to four significant figures:

$$L = 61.26 \text{ cm} \quad (\text{VI.2})$$

Note that the last figure reported has been rounded up. The rule for rounding the number is

- **If the first digit beyond the last significant figure is 5 or greater, the last significant figure is to be increased by one. The remaining figures beyond the last significant figure are discarded. If the digit beyond the last significant figure is less than 5, the last significant figure is not changed.**

Thus 3.05 rounds off to 3.1, but 3.04 rounds off to 3.0. Most pocket calculators that round off automatically use this rule.

Another numerical example will further illustrate the point. Let us calculate the area of our table top. In addition to the length from above, let us suppose that it has a width of $W = 97.33 \text{ cm}$. To find the area A of a rectangle we multiply the length, L , times the width, W .

$$A = L \times W \quad (\text{VI.3})$$

The arithmetic product of 61.26 and 97.29 is 5959.9854. However, it is not correct to give the area of the table top as 5959.9854 cm². This would imply that the area is known to be between 5959.9855 cm² and 5959.9853 cm², a precision that is unwarranted since the lengths of the individual sides are only known to four significant figures.

To see what this means, notice that the length is between 61.25 cm and 61.27 cm, and the width is between 97.28 cm and 97.30 cm. The product of the smaller of these numbers is 5958.4 and the product of the greater is 5961.571. We see, then, that even the last place before the decimal is uncertain, so it would be a mistake to give an answer with more than four figures. Thus, we round off the computed area to four significant figures and report 5960 cm².

- **The general rule is that your answer must have no more significant figures than is consistent with the least precise of your values (that is, the value with the fewest significant figures).**

Sometimes it may not be clear whether the final zeros in a number are significant figures or are merely needed to locate the decimal point. In the preceding example, the result of 5960 cm² was correct to four figures so that the zero was a significant figure. To avoid any confusion we can express a number in scientific notation. Our answer of 5960 cm² becomes 5.960 x 10³ cm². The presence of the final zero indicates that the number is known to four significant figures and it is now clear that the final zero is not simply a place holder.

- **We assume integers to be exact.**

If you count the number of floor tiles along the length of the hallway or the number of oscillations of a pendulum, this is an exactly known number. If we find the length of the hallway by multiplying the number of tiles by the width of each tile, then the final answer is limited by our knowledge of a tile width. If we find the period of one oscillation by timing ten of them and dividing by ten, the precision of the answer is limited by the precision of the time measurement.

Hopefully, it is clear by now that the number of significant figures reported is closely related to the precision of the experiment. A precise measurement should be reported with more significant figures than an imprecise one. If we have made a measurement with a 10% uncertainty, it would be misleading to report the result to more than two significant figures. The number of significant figures you report should be commensurate with the uncertainties assigned to the value.

For example, suppose that you make five measurements of a time and get

2.45 s 2.378 s 2.514 s 2.47 s 2.4098 s

Using the methods outlined in section V, we can compute the average to be

$$\bar{t} = \frac{2.45s + 2.378s + 2.514s + 2.47s + 2.4098s}{5} = \frac{12.2218s}{5} = 2.44436s \quad (\text{VI.4})$$

and the standard deviation (our estimate of the uncertainty) is

$$\begin{aligned} \Delta\bar{t} &= \sqrt{\frac{(2.45s-2.44436s)^2 + (2.378s-2.44436s)^2 + (2.514s-2.44436s)^2 + (2.47s-2.44436s)^2 + (2.4098s-2.44436s)^2}{4}} \\ &= \sqrt{\frac{3.18 \times 10^{-5} s^2 + 3.290 \times 10^{-3} s^2 + 4.849 \times 10^{-5} s^2 + 6.97 \times 10^{-4} s^2 + 1.1943 \times 10^{-3} s^2}{4}} = \sqrt{\frac{5.26 \times 10^{-3} s^2}{4}} = 0.0362s \quad (\text{VI.5}) \end{aligned}$$

and our estimated percent uncertainty is

$$\frac{0.0362}{2.44436} \times 100\% = 1.48\% \quad (\text{VI.6})$$

If we say that we know something to $\approx 1\%$, we mean that we know the value to three significant figures (1 part in 100). We can compare our uncertainty with our tabulated data and see that the values given with four or five significant figures don't really give us any useful information because of the inherent precision of the experiment.

- **When reporting a result, the number of significant figures should reflect the number of significant figures in the least well determined datum or the uncertainty, whichever is the least precise.** We should thus report the answer as (2.44 ± 0.03) seconds. In this case this happens to be the same number of significant figures you would assign from an inspection of the data (2.45s and 2.47s are both only known to three significant figures). This may not always be the case. Significant differences between the two measures of experimental precision indicate that either the uncertainties are estimated incorrectly and/or the data is being recorded with too many (or too few) significant figures. In the textbook, significant figures are used to indicate how well a quantity is known. Rather than give a set of data that must be averaged each time, you are simply given the value with the correct number of significant figures. If all you were told was 2.44 s, you would know that the error is of the order of 1% (smaller than 10% and bigger than 0.1%).

So we see that the experimental precision is connected with both the uncertainty and the number of significant figures reported in the answer. This is why significant figures are emphasized in lecture and uncertainties are emphasized in lab.

The rules governing significant figures and uncertainties can be summarized as follows

Table VI.1 Rules for maintaining the correct number of significant figures.

1	If the first digit beyond the last significant figure is 5 or greater, the last significant figure is to be increased by unity (one). All other figures beyond the last significant figure are dropped. If the digit beyond the last significant figure is less than 5, the last significant figure remains unchanged.
2	Use scientific notation to avoid using zeros as place holders. A trailing zero is to be included only if it is significant.
3	When reporting a result, the number of significant figures should reflect the number of significant figures in the least well determined datum or the uncertainty, whichever is the least precise.

VII. Graphing Data

Graphing your data is a very efficient way to present the information gathered in your experiment. The reader can then easily see the scatter in the data (the precision) and any trends indicated (is it a straight line? what is the slope?...). However, as in photography, perspective is important. Just as two different camera angles can emphasize different aspects, choosing axes differently can hide or enhance the important aspects of your data. Presented here are guidelines for making a good graph. As with many things in life, there will be times when these rules should be broken. The guidance in this matter should come from your lab instructor.

Consider an experiment where you are asked to measure the length of a board during different phases of the moon. By plotting the length of the board as a function of the lunar phase, you can determine if there is any correlation between the two. Over the next month you determine the length of the board four times, once during each phase and get the data in the table.

Table VII.1. The observed correlation between the phase of the moon and the length of my board.

Lunar Phase	Length of the Board	Uncertainty in Board Length
1 st quarter	12.1 cm	0.5 cm
2 nd quarter (full)	13.0 cm	0.8 cm
3 rd quarter	11.8 cm	0.5 cm
4 th quarter (new)	12.6 cm	0.5 cm

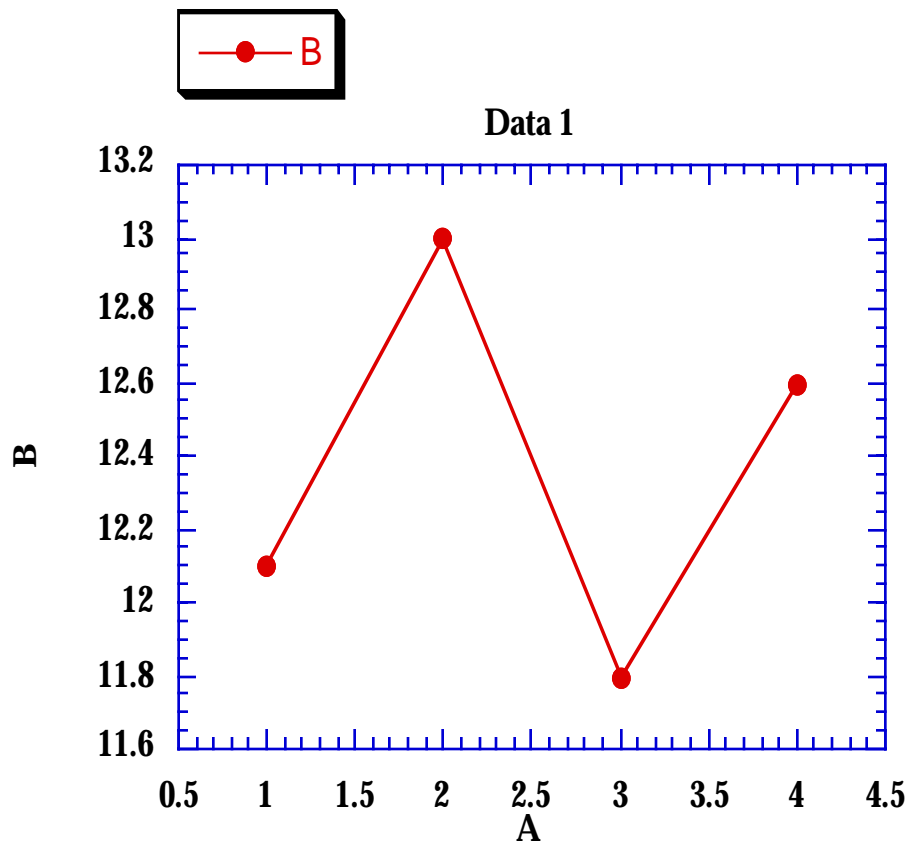


Figure VII.1 The first (wrong) attempt at plotting the data.

The usual arrangement of a graph is to plot the independent variable (here, the lunar phase) as the abscissa (along the horizontal axis) and the dependent variable (the length of the board) as the ordinate

(along the vertical axis). If I put this into a plotting program on the computer and plot length (ordinate) versus lunar phase (abscissa) I get something like the figure VII.1 shown here. There are numerous things wrong with the plot as shown. First, every graph should have a meaningful title and the axes should be labeled to indicate what is being plotted. Instead of **Data 1** we should put something like **Correlation between the length of my board and the phase of the Moon**. Now the reader knows what to look for in the plot. The horizontal axis should be labeled **Lunar Phase (quarters)** and the vertical axis should be labeled **Length (cm)**. But there is still much to be done to make this into an acceptable plot.

Never play 'connect-the-dots' with your data. Lines should represent a physical interpretation or calculation relevant to understanding the experiment.

On the horizontal axis, we are plotting the phase of the moon which we know as an integer. It is not useful and can even be misleading to show such a finely divided scale on a plot with integer data (like trial number, revolution number,...) The horizontal axis should have the range and the number of tick marks changed so that we only label the scale with integers. It will also help the reader find all the data on the plot if there is some space before the first point and after the last one so that no data point is on an axis line. This rule is commonly broken to avoid another problem. When the least sensible value of the independent variable is zero (when measuring time for example), the plot is stopped at zero even if there is a data point at that position. This avoids labeling the strange looking negative value at the position of the vertical axis. This is not a problem here, however.

What about the vertical axis? Certainly, we don't want the second and third data points to be on an horizontal axis (either top or bottom), but where do we want them? A good rule of thumb is that you should include zero on the vertical axis if zero could, under some circumstance, be a sensible value. Although my board is not 0 cm long, very short boards are possible, so I will include zero on the vertical axis. **In most of the graphs you will produce in this lab, zero should be included on the vertical axis.**

On the plot as shown above in figure VII.1, it appears that data are uncorrelated and random. This is because the variations at which we are looking represent the random fluctuations inherent in the measurement process which give rise to uncertainties. If I were to attach a vertical line to each data point that represents ± 1 RMS deviation, as is illustrated in figure VII.2, I could show the precision of the experiment graphically. Including these so-called *error bars* along with the other changes discussed gives the corrected graph, figure VII.3.

Since some experiments generate many data points and others only a few, the propagation of the uncertainties (section V) to put error bars on *every* data point can be a time consuming chore. We can reduce this amount of work to do without sacrificing the information in our graph by observing that the size of the error bars are often similar and they are most similar near the middle of the range of the data. If there were data points with larger uncertainties, they are most likely near the ends of the measurement

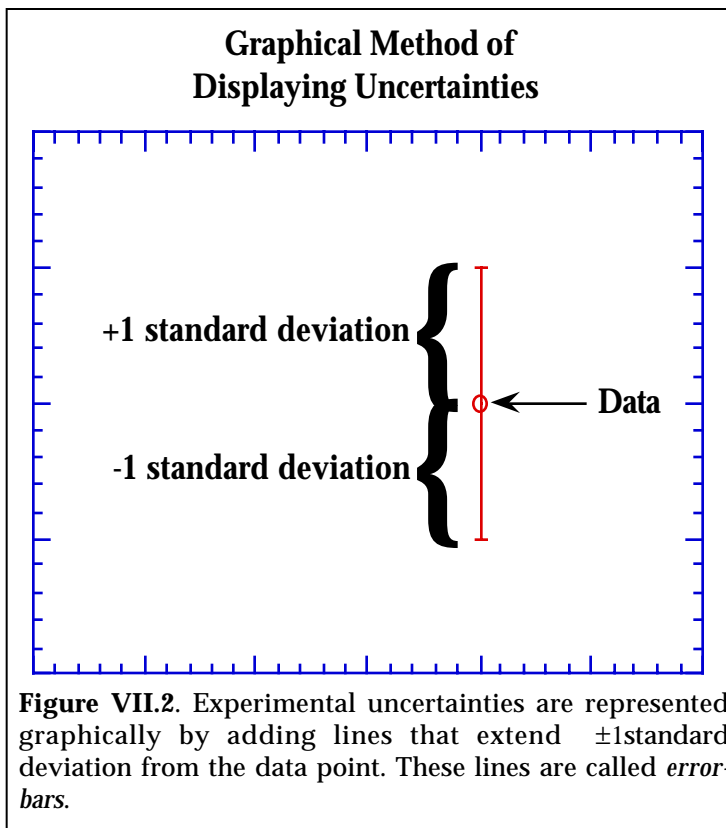


Figure VII.2. Experimental uncertainties are represented graphically by adding lines that extend ± 1 standard deviation from the data point. These lines are called *error bars*.

range. So, if we were to compute the RMS deviations for the first two data points and the last two data points, it is likely that we'll have a good, conservative, indication of the uncertainties in the measurements. Thus, the rule we'll use for making graphs in this course is the following: **always compute the standard deviations for the first two and last two data points and plot errors for these data on your graph.** This means that all the data will have error bars when there is four or fewer data points, but the work load never gets any larger than this, regardless of the size of the data set. In section VIII, the methods for using error bars on graphs to interpret the experimental results are discussed.

There is a very useful side effect to plotting the data with the vertical axis starting at zero. The distance from $y=0$ up to the value is proportional to the value. Thus, the variations in this distance give a visual measure of the differences among the data points. On the corrected graph, you can now see that the variations in the measured lengths are not significantly larger than the uncertainties in the data points. In other words, the length of my board is the same, independent of the phase of the moon, just as I expected. This is why it can be so useful to include the uncertainties on the graph. **If all (or nearly all) of the data points are within \pm the uncertainty (± 1 standard deviation) of the same value, we have no basis for claiming that they are different.** Now we see that the right answer is to simply average all the various measurements of the length of the board to get the best estimate.

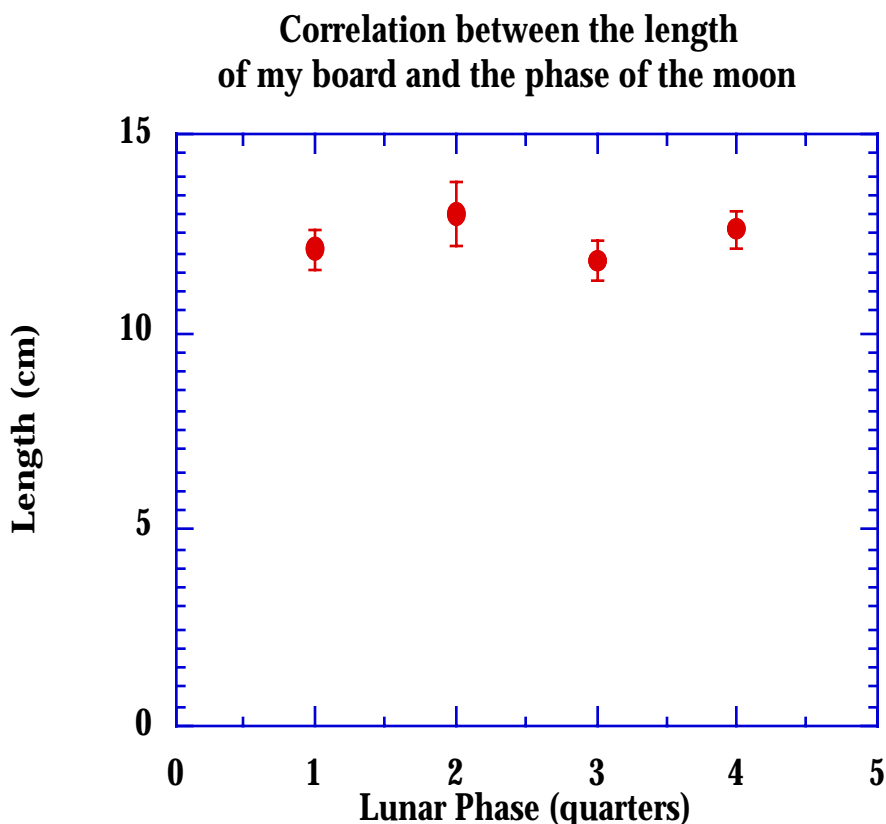


Figure VII.3. The graph in figure VII.1 after applying the corrections discussed in the text.

In your plots, always be mindful of the uncertainties in the data. The way the graph was first (figure VII.1) made might lead you to believe that some strange correlation existed. But when plotted sensibly, taking into account the uncertainties on the data, this odd behavior disappears. The rules of thumb for producing graphs are summarized in table VII.1.

Table VII.1 Summary of rules for producing a graph.

0	When asked to plot "Thing-1 versus Thing-2", this means that "Thing 1" is the ordinate, plotted on the vertical axis and "Thing-2" is the abscissa, plotted on the horizontal. That is, "Thing-1" is the <i>dependent variable</i> and "Thing-2" is the <i>independent variable</i>
1	Never draw lines from point to point on your graph. Any lines should represent a physical interpretation of the data. Use what you know about the physical situation as a guide.
2	Every plot should have a descriptive title. This title should not be just a statement of 'thing A versus thing B', explain what the graph means.
3	Every plot should have the axes labeled with the quantity being plotted and its units. Remember that when you compute a slope, or find an intercept, these quantities must have the correct units.
4	The horizontal axis should extend beyond the range of the data in both directions to prevent having a data point right on an axis. The exception to this is when negative values of the independent variable are unphysical or inappropriate. In this case stop at zero and allow a datum on this axis.
5	The vertical axis should include zero and extend beyond the largest data point. Again, no data points should be obscured by axes. By including zero, the distance up from the horizontal axis is proportional to the value and the fluctuations give a visual indication of the uncertainties in the data.
6	Display the uncertainties graphically by plotting ± 1 standard deviation on four data point. For small data sets, this will be all the points plotted. For large data sets, error bars should be plotted for the first two data points and the last two data points. This provides a visual display of the uncertainties in the data set without calculations that add little to the usefulness of the graph as an analysis tool. See section VIII for more information on the graph as an analysis tool.

VIII. Data Reduction

Fitting a Straight line

In the lecture portion of this course, physics is presented logically, beginning with the general law followed by its application to specific instances. In the lab you are presented with a specific instance and are asked to verify/derive/propose and understand the general case. These two portions of the course require you to do entirely different things. So the question naturally arises: How is the material presented in lecture connected to the experiments done in lab?

Part of the reason this connection is muddled is that a general law and a single measurement have a one-way relationship. If the general law is known, then it is possible to predict the outcome for any particular instance. For example, if I know Newton's laws and the initial speed and angle, I can predict where a thrown baseball will land. If, however, all that is known is one particular instance, then it is impossible to predict the general law. This is represented in figure VIII.1 by the upper arrow pointing from the general law to the single measurement. To proceed from experiment to theory, we must have many measurements so that we can find the pattern in the data. If I throw the baseball many times with different angles and initial speeds, then I can, with care and patience (and a good bit of insight), infer Newton's laws. It is only by inspecting the pattern made by the data that we can possibly recover the underlying physics. In our example, the final location of the baseball as a function of throwing angle and speed provides the pattern to infer the general law we seek. This is represented by the lower arrow which points both directions.

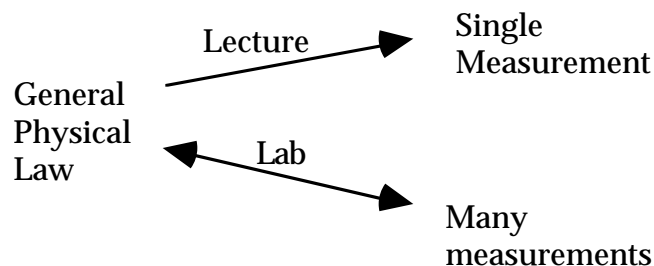


Figure VIII.1. The relationship between experiment and theory.

Therefore, in lab you are asked to do something quite different from what is done in the lecture portion of the course. In lab we are concerned not just with the various values of a measured quantity, but also with how these measurements relate to one another. The relationships among the various data are determined by physical laws. The primary tool used to examine our data for these patterns is the graph. This allows us to visualize the relationship among the various measurements. But how do we reduce the collected set of experimental data to results which can be directly compared with theory? How does the graph contribute to this data reduction process?

One of the most powerful analysis tools we will use in this course is that of finding the best straight line to represent a data set. When we plot our data and find that the data lies approximately along a line, we can draw the straight line which best represents, or fits the data. This line is then used for interpolations and extrapolations. More importantly, this best fit line is used in place of the data since it quantifies the pattern we seek.

Suppose that we plot a set of data, y versus x , and from the graph see that a straight line will represent our data. Describing this straight line requires two parameters: the slope, m , and y -intercept, b . Once these two numbers are known, we can compute the value of y that corresponds to any x from

$$y = mx + b \quad \text{(VIII.1)}$$

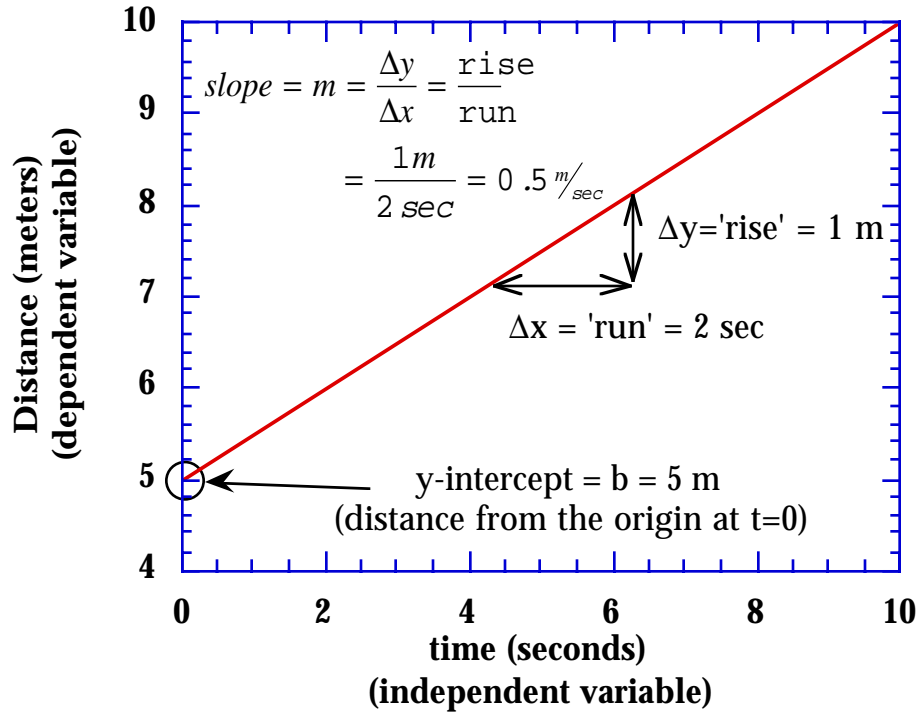


Figure VIII.2. Graph of distance traveled versus time obeying a straight line relationship. The meaning of the slope and y-intercept parameters, m and b are given in the figure.

More importantly, it is precisely the slope and intercept that play a direct role in the theoretical calculations. For example, if I plot position versus time, the intercept is the position at time=0 and the slope is the speed. This is illustrated in figure VIII.2.

Thus, the problem becomes one of finding the slope and y-intercept of the line that represents our data, along with estimates of the uncertainties in these quantities. Because the arithmetic is cumbersome without a computer, we will use an approximate graphical technique. To illustrate the method, let us suppose that we have the data given in table VIII.1. Uncertainties have only been computed for the first two and last points as described in section VII.

Table VIII.1. Sample data for the example described in the text.

Time (sec)	Distance, y (cm)	Distance Uncertainty, Δy (cm)
1.0	20	1
2.5	59	1
4.7	64	(Not computed)
8.0	98	2
10	129	3

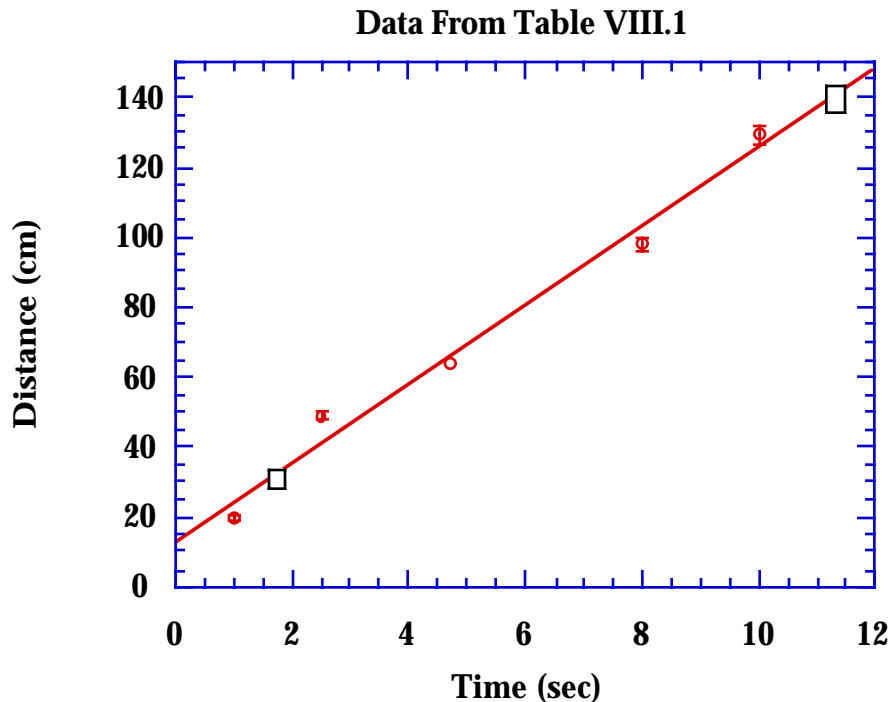


Figure VIII.3. Data from table VIII.1 with a 'best fit' line drawn in. The circles represent the data and the boxes indicate the points on the line used to compute the slope. Notice that the error-bars illustrating the uncertainties in this data set are about the same size as the plotting symbol. Following the rules outlined in section VII, errors are computed and shown only for the first two and last two data points.

The first step is to plot the data as shown in figure VIII.3. Using a ruler, we can draw the line that comes the closest to all the data points. This line we call the 'best fit' line. If all the uncertainties on the data points are small or about the same size, then this is pretty easy. If, however, some data have much larger or smaller uncertainties than the rest, then this is a bit more difficult. In this case you must try to give more importance to the data with smaller uncertainties. Large uncertainties means that those measurements provides less information and thus, should be given less importance.

The y-intercept of the line is simply read from the graph. In this case it is about 13 cm. Remember everything has units! To find the slope, we use the standard 'rise over run' method (see figure VIII.2)

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (\text{VIII.2})$$

where y_1 , y_2 , x_1 , and x_2 are points on the line, not data points. **This is important:** once you have drawn the best fit line through your data, it is the slope and y-intercept of the line in which we are interested, not that of the data. Notice that in this case none of the data points lie exactly on the line. We use the data to find the line and we use the line to find the parameters to compare with theory. In this case we get

$$m = \frac{32\text{cm} - 140\text{cm}}{1.7\text{sec} - 11.3\text{sec}} = 11.25\text{cm/sec} \rightarrow 11\text{cm/sec} \quad (\text{VIII.3})$$

The last step in equation VIII.3 is included because we can only keep two significant figures in the answer since the data in Table VII.1 are only known to two significant figures. Finding the slope of a line by reading a graph might appear to be inaccurate because of the additional uncertainty in reading points from the graph. In practice, however, this is not a problem. By choosing points far apart, $y_2 - y_1$ and

$x_2 - x_1$ are large compared with the uncertainty in reading the graph and we can ignore this tiny effect.

How do we find the uncertainties in the slope and intercept? The method we use is a variation of one first described by Prof. John L. Safko¹. We begin by drawing a parallelogram that encloses all the data and the uncertainties. This *parallelogram of uncertainty* is shown in figure VIII.4.

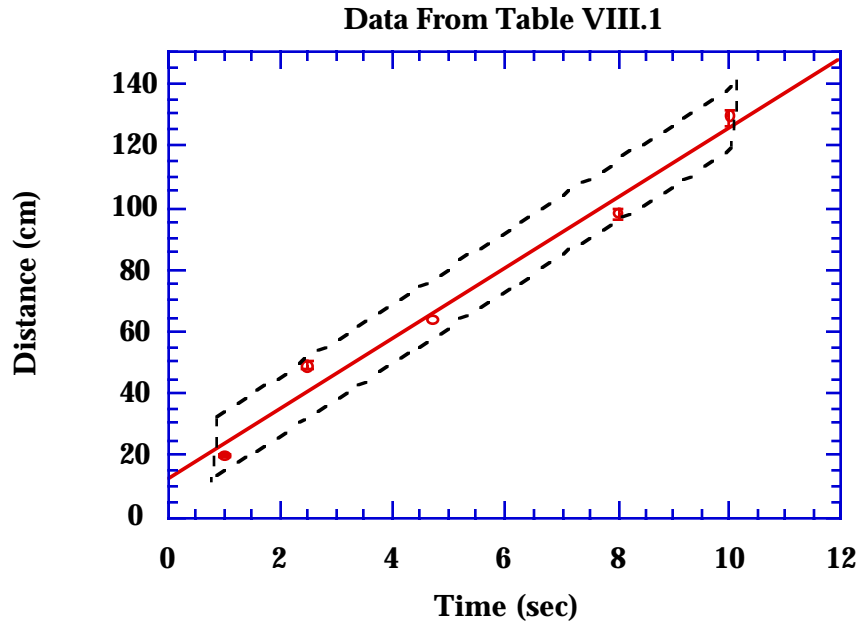


Figure VIII.4. Same as figure VIII.2 but with the parallelogram of uncertainty added.

The upper and lower lines of this parallelogram are drawn parallel to the best fit line and the ends are drawn parallel to the y-axis. The parallelogram of uncertainty is drawn so that it encloses all data points and error bars. Any estimate of the slope and y-intercept we make must result in a line that passes through the ends of the parallelogram without ever crossing the top or bottom. We can thus construct the lines with the maximum and minimum slopes by drawing lines through opposite corners. This is shown in figure VIII.5.

The line labeled A has the maximum slope and the one labeled B has the minimum slope consistent with the parallelogram. Computing the slopes of these two lines as we did for the best fit line gives

$$Slope_{\max} = m_A = \frac{10\text{cm} - 140\text{cm}}{0.86\text{sec} - 10\text{sec}} = 14 \text{ cm/sec} \quad (\text{VIII.4})$$

$$Slope_{\min} = m_B = \frac{31\text{cm} - 122\text{cm}}{0.86\text{sec} - 10\text{sec}} = 10 \text{ cm/sec} \quad (\text{VIII.5})$$

The difference between these two extreme values for the slope gives

$$m_A - m_B = 4 \text{ cm/sec} \quad (\text{VIII.6})$$

Because we want to express our uncertainties symmetrically, as plus or minus some value, not as a total range, we divide this difference by 2 to get a measure of the uncertainty in the slope for this experiment

to be $\frac{m_A - m_B}{2}$.

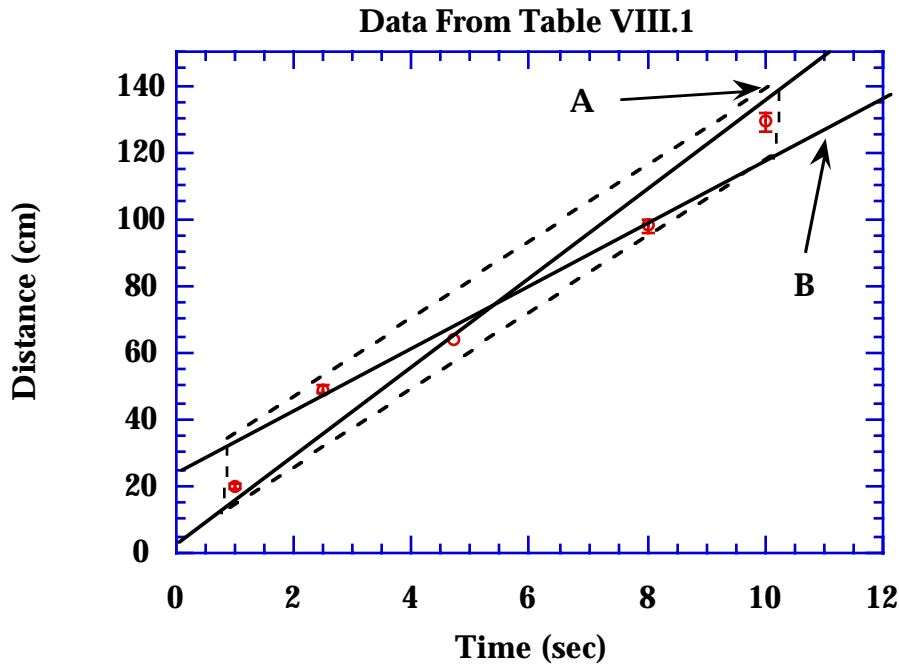


Figure VIII.5. The lines A and B represent the lines of maximum and minimum slope, respectively. They are drawn by connecting the diagonals of the parallelogram of uncertainty.

Comparing the data to the two lines we've drawn, we see that it is unlikely that we'd ever choose either line A or B as the best fit line. Thus, we have overestimated the uncertainty. A more careful analysis of this method shows that we can get a reasonable estimate of the uncertainty in the slope by dividing by a factor related to the number of data points, namely

$$\Delta m = \frac{m_A - m_B}{2} \frac{1}{\sqrt{N-1}} \quad (\text{VIII.7})$$

where N is the number of data points. Equation VIII.7 gives an estimate of the most probable uncertainty that we wish to report. To understand the denominator in equation VIII.7, consider the following. If N=1, $\Delta m = \infty$ since a line with any slope can be drawn through one point. For N=2, a line can be drawn exactly through two points. In this case the parallelogram of uncertainty arises only from the individual uncertainties on the two data points. Hence, it is reasonable that $\Delta m = \frac{m_A - m_B}{2}$. For $N \gg 2$, the measure of the uncertainty decreases like the square root of N. This is exactly the behavior we see when computing the standard deviation to find the uncertainty in an average. Although this is not an exact solution to the problem, it is a reasonable one. So, in the example we are considering here, we have

$$\Delta m = \frac{4 \text{ cm/sec}}{2\sqrt{5-1}} = \frac{4 \text{ cm/sec}}{4} = 1 \text{ cm/sec} \quad (\text{VIII.8})$$

Finding the uncertainty of the y-intercept proceeds in a similar fashion. In this case we simply read the values from the graph to find the upper and lower limits of 24 cm and 0.8 cm. The difference between these two values gives a measure of the uncertainty. The uncertainty we report is related to this difference just as in equation VIII.7

$$\Delta b = \frac{24\text{cm} - 0.8\text{cm}}{2\sqrt{5-1}} = \frac{23.2}{4}\text{cm} = 6\text{cm} \quad (\text{VIII.9})$$

Thus, the final answer is expressed as

$$m = (11 \pm 1)\text{cm}/\text{sec} \quad (\text{VIII.10})$$

$$b = (13 \pm 6)\text{cm} \quad (\text{VIII.11})$$

Notice that the uncertainty in the slope is about the same size as the uncertainties on the data we are fitting. However, the uncertainty in the y-intercept is very large, about 50%. Why is this intercept so poorly determined? These observations about the relative uncertainties are typical of this type of analysis. In the case of the slope, we are finding a kind of generalized average so it makes sense that the uncertainties are commensurate with (or even smaller than) that of the data. For the y-intercept, however, we are making an extrapolation to values of time for which we have no data. The farther we extrapolate from the region where we have data, the more uncertain the values get. At some point (not very far from the data!), we will have over 100% uncertainty in the position of the object and we can say nothing about it from our experiment. You may argue that we know that the object moved at a constant speed and we *know* that it continued for times before and after the measurement interval. This may be so, but we have no data to substantiate this claim. To reduce the uncertainties, we must collect more data.

Nevertheless, there will *always* be uncertainty in the y-intercept, even if it is in the middle of the data set. Just because you have a data point on the y-axis, do not assume that it completely determines the y-intercept. In the example above, the best fit line does not pass directly through any of the data points and this may be the case for your data as well. Construct the parallelogram of uncertainty, find the range for the slope and y-intercept and compute the uncertainties.

This technique is so useful that with some thought we can apply it to a variety of other situations that don't appear to be linear relationships. For example, plotting the position versus time squared for a falling object, plotting pressure versus 1/volume for an ideal gas, plotting power versus the square of the current through a resistor and plotting the period versus the square root of the length for a simple pendulum all give straight line graphs. Once you have a graph that is a straight line, the graphical technique can be applied.

Using Semi-log paper

A slightly more complicated, but common case, is the exponential function. This situation is often confusing because we use semi-log graph paper to simplify plotting the data. Suppose we measure the activity of a radionuclide at several times and find the data given in table VIII.2.

Table VIII.2. Activity as a function of time for our radioactive source.

Time (min)	Activity, A (counts/sec)	Activity Uncertainty, ΔA (counts/sec)
0.0	1010	31
1.0	740	28
2.0	585	24
3.0	435	21
4.0	325	18

Since we know that radioactive decay follows an exponential decay law, we expect to find that the activity as a function of time is given by

$$A(t) = A_0 e^{-\lambda t} \quad (\text{VIII.12})$$

where A_0 is the activity at time = 0 and λ is the decay constant. To use our technique to find the decay constant and the initial activity, we must first rewrite equation VIII.12 so we have a linear relationship to plot. This we do by taking the natural logarithm of both sides of the equation and using the basic properties of the exponential and logarithm functions to get:

$$\ln[A(t)] = \ln[A_0 e^{-\lambda t}] = \ln[e^{-\lambda t}] + \ln[A_0] = -\lambda t + \ln[A_0] \quad (\text{VIII.13})$$

Dependant Variable
Independent variable

Slope
Y-intercept

By plotting the natural logarithm of the activity, $\ln[A(t)]$, versus time we should get a line with a y-intercept of $\ln[A_0]$ and a slope of $-\lambda$. We can either compute the logarithms of each of the activities and use regular graph paper or we can use semi-log paper. It is always preferable to use semi-log paper. On semi-log paper, the distance along the vertical axis is proportional to the logarithm of the number plotted. This gives you the advantage that while you plot your data directly, with no extra calculations, the vertical distances are proportional to the desired logarithms. In this way, data that follows an exponential decay curve will produce a straight line as shown in figure VIII.6. **Once we have a straight line relationship on our graph paper, we can apply our parallelogram of uncertainty method exactly as before.**

However, we must be careful. When the data was plotted, the logarithm was never computed. Just because the graph paper has translated the picture for us, does not mean that the logarithms have

been mysteriously computed for us. **When finding the slopes on semi-log graph paper, remember to compute the logarithms!** In this case, the slope of the best fit line is found from the points on the line indicated by the boxes in figure VIII.6 as

$$\begin{aligned} \text{Slope} &= \frac{\ln(860 / \text{sec}) - \ln(282 / \text{sec})}{0.5\text{min} - 4.5\text{min}} \\ &= \frac{\ln\left(\frac{860}{282}\right)}{-4.0\text{min}} = \frac{1.1}{-4.0\text{min}} = -0.27 / \text{min} = -\lambda \end{aligned} \quad (\text{VIII.14})$$

and we find the decay constant to be $\lambda = 0.27 / \text{min}$. The y-intercept, on the other hand, can be read directly from the graph to be 1000/sec or, in the same units as the slope, we have

$$A_0 = \frac{1.00 \times 10^3}{\text{sec}} \times \frac{60\text{sec}}{1\text{min}} = 6.00 \times 10^4 / \text{min} \quad (\text{VIII.15})$$

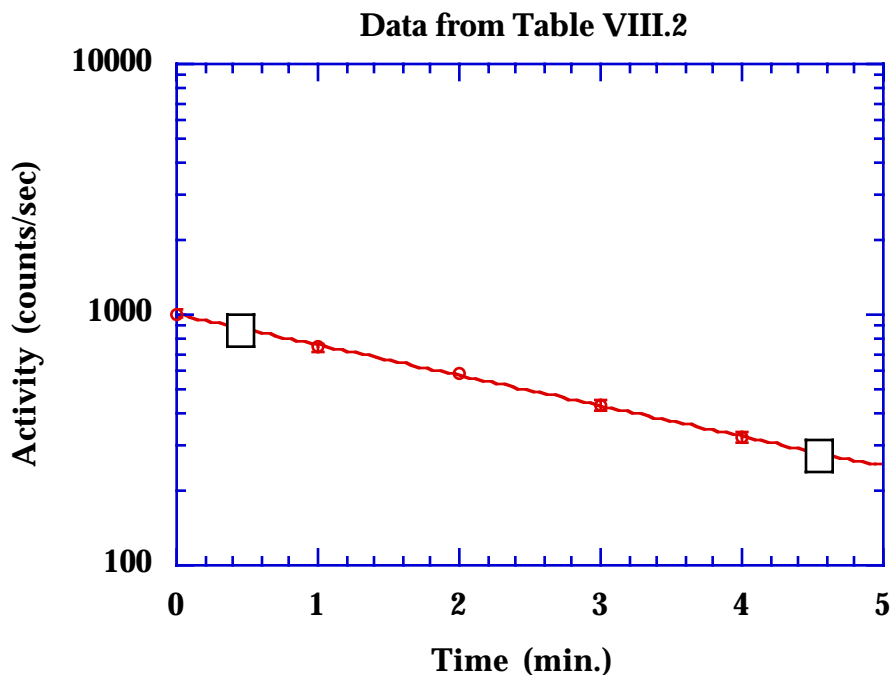


Figure VIII.6. Plot of radioactive decay data from table VIII.2 on semi-log graph paper. Error bars are plotted for only the first two and last two data points as described in section VII.

So by using semi-log paper, we can get a linear relationship to graph and only have to compute one logarithm to find the equation of the line.

The uncertainties in the slope and y-intercept are estimated from this plot in exactly the same way as was done on a plot on regular graph paper. The parallelogram of uncertainty is constructed, the lines of maximum and minimum slope are drawn. The extreme slopes are computed using the same method illustrated in equation VIII.14 and the extreme y-intercepts are read from the graph. The uncertainties are then computed just as before. Thus, using semi-log graph paper follows the same procedures as for a plot on regular paper, with the only difference being the way the slope is calculated.

The process of estimating the best fit straight line through a data set is central to the connection

between physics and experiment. With the methods described here for finding the line and estimating the uncertainties in its slope and y-intercept, you have the tools you need to complete the connection between lecture and lab while carrying along an estimate of the experimental uncertainties. The method described here relies on graphical techniques rather than tedious mathematics that, while rigorously correct, only obscure the important ideas. Regardless of the field of science you choose to pursue, these are invaluable tools for analyzing data. The method is summarized below for ready reference while working in the lab.

Table VIII.3 Summary of data analysis procedures.

1 Plot the data following the guidelines in section VII. If the data does not lie on a straight line, try plotting various (theoretically inspired) combinations of squares, square roots, exponentials or logarithms. If in doubt, consult your lab instructor. **2** Once you have a graph that can be represented by a straight line, use a ruler to draw the line which best represents the data. **3** Compute the slope and intercept of the *line* (not the data). These are the experimentally determined values. **4** Construct the *parallelogram of uncertainty* that completely encloses all the data points including the uncertainties. **5** Draw the straight lines connecting opposite corners of this parallelogram. The slopes and intercepts of these two lines give the maximum and minimum values for the slope and intercept. **6** Compute the most probable symmetric uncertainty on the slope and intercept as

$$\Delta m = \frac{Slope_{\max} - Slope_{\min}}{2\sqrt{N-1}} \quad \text{and} \quad \Delta b = \frac{Intercept_{\max} - Intercept_{\min}}{2\sqrt{N-1}},$$

where N is the number of data points on the graph. **7** Don't forget that when plotting data on semi-log paper, you must still compute the

logarithm when finding the slope: $slope = \frac{\ln(y_2) - \ln(y_1)}{x_2 - x_1} = \frac{\ln\left(\frac{y_2}{y_1}\right)}{x_2 - x_1}$

1) John L. Safko, "Error Analysis of Straight Line Plots in the Undergraduate Physics Laboratory", Am. Jour. Phys. **33**, 379 (1965).