

MEASUREMENT AND ERROR

1. Introduction to Error Estimation

The statement "this table is exactly 1.5 m long" cannot be true, or rather can never be demonstrated to be true. It may be that the length differs from 1.5 m by an amount that the speaker does not know how to determine, and perhaps tacitly assumes cannot be determined. More likely, the speaker means merely that any deviation from 1.5 m is insignificant for the purpose at hand.

Why can't the statement be true? One unique object in the whole world can be defined to have some exact length (say, the standard meter) as a matter of convention. All other lengths must be determined by measurement, which is a process of comparison, carried out experimentally. The length of the table cannot be determined by mathematical reasoning (logic). The measurement process is always subject to some degree of uncertainty, no matter how small.

The term "error" in this context does not refer to mistakes, but has come to mean the uncertainty in a quantity. Error is usually appended to a quantity with the \pm sign. Thus the expression (1.500 ± 0.001) m tells us not only that the table is about 1.5 m long, but that its length is known to be most probably between 1.499 m and 1.501 m.

A typical laboratory exercise involves measuring some quantity, and comparing the value obtained either with a calculated value or with another measurement (e.g. a handbook value). The expressed or implied question "Is your result the same as . . .?" CANNOT be answered categorically "yes" or "no" since the quantities compared are subject to uncertainties. If the errors in the quantities overlap, they may be said to be "in agreement" or "mutually consistent", but obviously the strength or weakness of this statement depends a lot on how big the errors are. Similarly, you can't claim that the quantities are different unless their values differ by much more than their errors.

Some of the laboratory exercises in this course explicitly require error analysis. Even those that do not should include estimates of error whenever reasonably possible, and all laboratory work should be done with at least a semi-quantitative awareness of the uncertainties involved.

2. Error in a Single Physical Quantity

2.1 Random vs. Systematic Error

Truly random error is a matter of pure chance. It is common in the microscopic world (e.g., how many atoms in a radioactive sample will decay in a particular minute). Systematic error usually refers to some flaw in an instrument or procedure which will affect the readings consistently (e.g. the end of a meter stick is worn down). Repeating the measurements and averaging the results will reduce random errors but will obviously not affect systematic errors if the same instruments and methods are used repeatedly. All of the "theory of errors" which follows is just an application of the mathematics of probability, and only applies properly to random error.

Suppose that a particular electrical timer goes at a rate which depends slightly, but in certain definite ways, on temperature, humidity, air pressure, the condition of its battery, and so on. In principle one could study all of these systematic effects and apply suitable corrections to each measurement. In practice however we would probably be content to consider these unobserved

variables to affect the timer in a pseudo-random way and assign some conventional error to "cover" the variations. Thus it is that many errors which are fundamentally systematic come to be treated as random.

2.2 Absolute and Relative Error

These terms merely refer to two forms of expression. Any error can be stated either way. "Absolute" error in a quantity simply means the direct amount by which the given quantity may be greater or smaller. The absolute error is the same kind of quantity as the measured value and must be expressed in consistent units: $(1.500 \pm .001)$ m or $1.500 \text{ m} \pm 1 \text{ mm}$ are equivalent acceptable forms. Both establish a range 1.499 to 1.501 m which probably includes the correct value. Note that $1.500 \text{ m} \pm .001$ is not a proper expression.

Relative error is the error expressed as a fraction of the quantity itself, usually in terms of percent (= parts per hundred), sometimes parts per thousand, and in more accurate work parts per million, abbreviated ppm. The measurement $3.00 \text{ m} \pm 1\%$ is identical to (3.00 ± 0.03) m. Relative error itself is dimensionless.

Whether it is better to use the absolute or relative form depends in part on how the measurement was made and in part on how it is to be used; see Propagation of Errors, to follow.

2.3 Accuracy versus Precision

In technical usage, these are quite independent terms. Suppose that you have a digital watch that is guaranteed to be correct to one minute per year. That is a statement of its accuracy, about 2 ppm. Let us assume that the same relative accuracy (2 ppm) describes the watch's ability to keep time over any interval. However, the display on this watch shows only hours and minutes. If your main interest in life is timing 100-m races, this watch is very nearly worthless. Most races will begin and end in the same minute. Sometimes the minute reading will change while the race is in progress. The precision, which is the fineness or smallest distinguishable value, is inadequate to the purpose at hand even though the accuracy of the watch is far better than needed.

Examples where the precision is too great for the accuracy abound. Conversion of units is a common source. A paintbrush is labeled 38.1 mm and a loaf of bread is marked 453.6 grams. See "Significant figures" below.

2.4 Interpolation

Say that you used the same watch as before for a slightly more reasonable purpose, to time a mile race. The watch might read 2:32 at the start and 2:37 at the end. You can figure out that the race may have taken anywhere from just over 4 to just under 6 minutes. You could report the time as (5 ± 1) minutes. The difference between 4 minutes and 6 is obviously of great interest in this case. How can you reduce the error?

You could arrange for the race to start just as the minute changes, and you could also watch each minute advance, with a view toward estimating what fraction, beyond the last whole minute, elapsed before the end of the race. Even with no practice you should be able to say whether it was near to $1/4$, $1/2$, $3/4$ or a whole minute; you could probably learn to estimate tenths. This is called interpolation. The example is somewhat inappropriate, in that a purely

digital device gives no clues for interpolation. Usually interpolation refers to estimating the position of a pointer between marks on a scale, both the marks and the space between being visible. Nevertheless the example illustrates the importance of interpolation; the error can be reduced by about an order of magnitude (± 1 minute to ± 0.1 minute) when the error is limited by precision.

As a general rule, record all readings on analogue scales to 0.1 or 0.2 of the smallest division. Not to do so may seriously impair the quality of your results.

2.5 Significant Figures and Rounding Off

In a number, all of the digits are significant from the first non-zero digit to the last non-zero digit. Leading zeroes are never significant, included zeroes always are. Thus the numbers 62085, 60.285, and 0.00060285 all have five significant figures. This has nothing to do with the number of decimal places. However, trailing zeroes are sometimes added to make the decimal point come out right, and this creates ambiguity. If the population of a large city is reported as 6 028 500, it is not clear whether the people have been counted down to the last one and it happened to come out a multiple of 100, or whether the figure is only meant to be precise to the nearest 100. The ambiguity should be avoided by writing 6.0285×10^6 , or 6.0285 million. (Zeroes added on the right past the decimal point are obviously intended to be significant).

The use of significant figures should always be consistent with the precision of a reported measurement. The following rules apply to the statement of a final result. During calculations, more digits may and should be retained.

1. The error should be rounded off to one, or at most two, significant digits.
2. The last or least-significant digit in the value must be in the same decimal location as that in the error.

Example:

$$\begin{aligned}60285 \pm 5\% &= 60285 \pm 3014.25 \\ &\rightarrow 60285 \pm 3000 \\ &\rightarrow 60000 \pm 3000 \\ &= (60 \pm 3) \times 10^3 = 6.0 \times 10^4 \pm 5\%\end{aligned}$$

Note that the zero to the right of the decimal point is significant. The trailing zeroes in "3000" are not.

In rounding off, conventionally, .499 rounds down (to 0) but .500 rounds up (to 1). However, errors should never be rounded down in such a way as to imply significantly better accuracy than achieved. It would be misleading to report an error of 1.49 as "1". This is a case where keeping two digits in the error is justified (1.5) although rounding up to 2 is not bad.

For further handling of significant figures, see "Propagation of Errors in a Single Quantity" in Section 4.

2.6 Estimated Errors

Random, or statistical, errors, can be both determined and reduced at the expense of repeating the measurement many times. This will not work at all with errors which are systematic. In high-

precision measurements, accurate knowledge of systematic errors requires a profound understanding of the instruments and the procedures for using them. In addition, systematic errors can be checked by measuring the same quantity by different methods and with unrelated equipment. It is not unusual for a determination of a fundamental constant to take several years to complete.

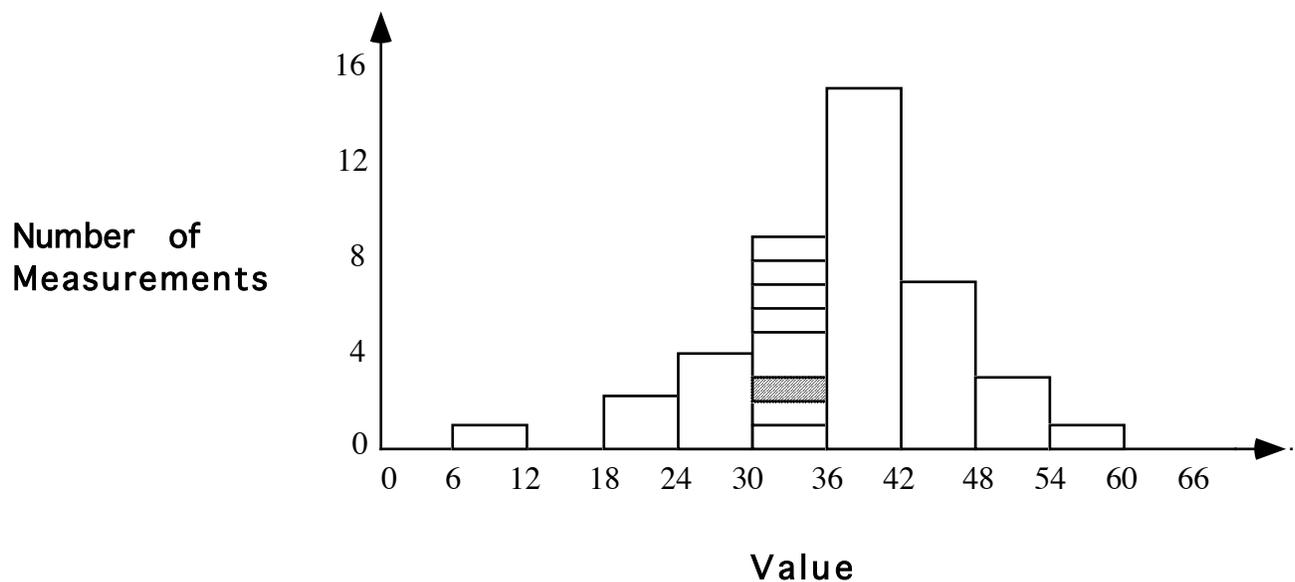
Nevertheless in ordinary lab experiments one of the simplest ways to determine errors is to estimate or "assign" them. This only requires an objective view of the apparatus and how it is used, some common sense and experience. Any difficulty that you experience in making the measurement should be translated into a quantitative estimate of error. Are you having trouble locating the exact end of the meter stick? Are the marks blurred? Can you get different readings by sighting from different directions? Is the object to be measured a blur rather than a point? All these and factors like them should be translated into amounts contributing to the error.

It was explained previously that errors which are fundamentally systematic may behave as random errors, at least in their variations. Thus in some experiments you will be asked to compare your estimated errors with the standard deviation, as if they were random.

3. Histograms

When you measure the same quantity many times, you do not always get the same value (unless the precision is inadequate). If the errors are purely random, the deviations should follow definite statistical laws. In any case, it is informative to prepare a histogram. Values or readings are grouped into "bins", and the number of readings which fall into each "bin" is plotted against the bin value. The histogram in Figure (3-1) is for about 40 measurements. Each small rectangle represents one individual measurement. Note that if the bins are too fine, not enough duplications occur to give the histogram a meaningful shape, but if the bins are too wide, there isn't enough differentiation (precision again). If there were a very large number of readings, the bins could be made quite small and the histogram would approach a continuous curve.

Figure (3-1): Histogram



3.1 Mean Value and Standard Deviation

Say that the same quantity x has been measured N times; the values are x_1, x_2, \dots, x_N . The mean \bar{x} is then

$$\bar{x} = \sum_{i=1}^N x_i / N,$$

that is, add all the values and divide by their number. Their deviations are the differences between each original value x_i and the mean \bar{x} :

$$\Delta x_i = x_i - \bar{x}.$$

Some deviations must be positive and some negative, and indeed by construction the mean of the deviations is automatically zero. This gives us no clue as to how the original readings were distributed. An empirical "cure" for negative signs is to square and take the square root. This is an heuristic justification for the definition of standard deviation σ (sigma):

$$\sigma^2(x) = \sum_{i=1}^N (\Delta x_i)^2 / (N-1)$$

or
$$\sigma(x) = \sqrt{[(\Delta x_1)^2 + (\Delta x_2)^2 + \dots + (\Delta x_N)^2] / (N-1)}$$

This is very close to averaging the squares of the deviations and then taking the square root. The standard deviation σ is an indication of how much the individual readings typically defer from their own mean. The denominator $N-1$ rather than N can be rationalized by noting that there is no mean and no deviation if only one measurement is taken.

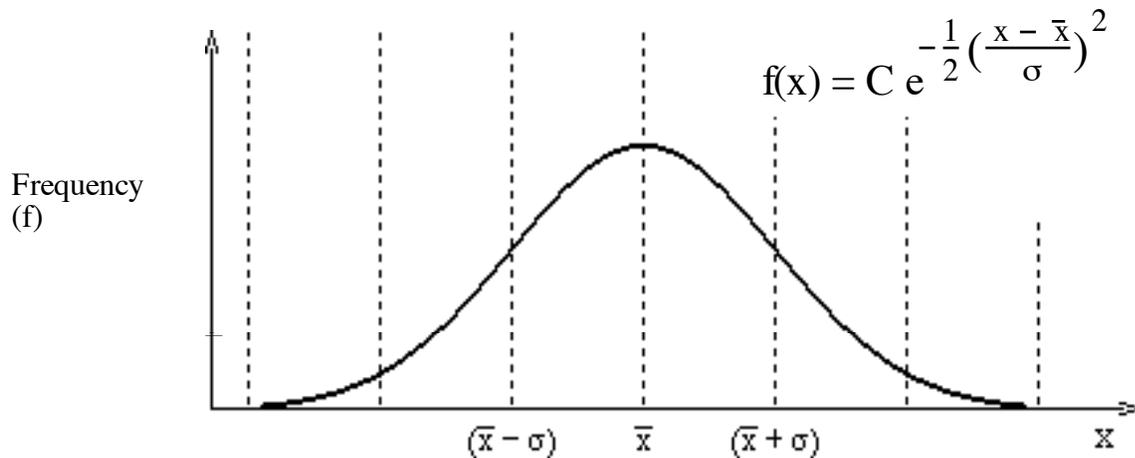
Note that mean and standard deviation are simply definitions which do not imply any assumptions about the kind of errors or their distribution.

3.2 Normal or Gaussian Distribution

If errors are purely random (chance) variations, then it is possible to predict the most probable histogram, i.e. calculate the "odds" of any given reading, for a given total number of readings. Of course the actual histogram need not be the most probable one in any given case.

If the number of measurements is very large ($N \rightarrow \infty$) then the calculated distribution approaches a continuous function. This is called the normal, or Gaussian, distribution function. The Gaussian distribution is graphed in Figure (3-2). The mean \bar{x} and standard deviation σ are parameters; that is, the position of the peak and its width are adjustable. Its mathematical form is also given in Figure (3-2). The Gauss function is useful as it is easy to derive definite results from a single continuous function, and these are reasonable guides to what to expect in practical cases.

Figure (3-2): Gaussian Distribution



In the Gaussian distribution, the fraction of all readings which fall within one standard deviation ($\pm \sigma$) of the mean is approximately 0.68, while more than 0.95 of all readings fall within two standard deviations of the mean. These are important guidelines in interpreting the error assigned to any measurement. That is, the meaning which we attach to any error is conditioned by our familiarity with the standard deviation of the Gaussian distribution.

Some of the rules for propagation of error described in Section 4 below are based on explicit properties of the Gaussian function.

3.3 Standard Deviation of the Mean

The standard deviation as defined previously indicates the scatter or distribution of the individual readings used to determine the mean, and may be taken as the error in each reading. It seems wise to anticipate a result at this time: the error associated with the mean value itself is less than that of the individual readings, by a factor of $1/\sqrt{N}$. Thus when many readings are taken, the mean value, \bar{x} , and the standard deviation of the readings, σ , are calculated. The standard deviation, (or error) of the mean is σ/\sqrt{N} , or, combining the steps,

$$\sigma(\bar{x}) = \sqrt{\frac{\sum_1^N (x_i - \bar{x})^2}{(N)(N-1)}}$$

The final result of a set of repeated measurements is then stated as

$$\bar{x} \pm \sigma(\bar{x}).$$

4. Propagation of Error

Let us assume that some quantity, say A or x , has been measured and that the corresponding error δA or δx has been determined. For this discussion it makes no difference how the error was obtained. As a matter of notation the symbol δ (lower case Δ , delta) is usually used to designate the error in a quantity, as distinct from d (differential) and Δ (difference or change). Very rarely is the measured quantity itself the end result of the experiment. Usually the data have to be manipulated in some way, i.e. the final result is some function of one or more measured quantities. At each step in the calculation, the error in the result is related in some definite way to the error in the data. If we know the rules for each kind of operation, we can carry the errors along from step to step, up to the final result. This procedure is called "propagation of error".

Recall that the error in a quantity A may be expressed either as absolute error (δA) or relative error ($\delta A/A$). The rules for propagation of error are easy to remember and apply if you are prepared to switch back and forth between the two forms of expression.

Many, but not all, of the following rules are approximations which are valid when the relative error is small ($\delta A/A \ll 1$). The reasoning behind such rules is that the error in a quantity can be identified with a kind of change, which is then approximated by the differential. In shorthand notation, $\delta A \rightarrow \Delta A \approx dA$. Since the error in a quantity is usually not sharply defined nor precisely known, we can tolerate poorer approximations in calculating the errors than in the quantities.

4.1 Operations on a Single Quantity

For this section, a "constant" means any quantity whose error is negligible.

- **Multiplying or Dividing by a Constant**

$$c(A \pm \delta A) = cA \pm c\delta A.$$

Just multiply or divide the absolute error by the same constant as the quantity. The relative error is unchanged.

Application: to find the thickness of a sheet of paper you measure, with a ruler, a new pack of 500 sheets, and divide both the answer and the error by 500. A reasonable relative error for the whole would be 5%, so you have attained a 5% error in measuring, say, 4×10^{-3} cm, for an absolute error of 2×10^{-4} cm ($2\mu\text{m}$).

- **Adding or Subtracting a Constant**

$$c + (A \pm \delta A) = (c + A) \pm \delta A$$

The absolute error is unchanged. The relative error must be recalculated.

Experimental procedures which require subtraction should be avoided whenever possible since diminishing a quantity without diminishing the error makes the relative error worse. Say that you place a book on a table, measure the top of the book from the floor as 102.4 ± 0.3 cm, then subtract the height of the table (100 cm, assumed error free). The result is (2.4 ± 0.3) cm, or 13% error even though your measurement had only 0.3% error. Far better to measure the book directly, where you could use a caliper.

This is closely related to round-off errors in calculation. Try $\pi + 10^{10} - 10^{10}$ on your calculator, for example.

- **Exponentiation (Powers and Roots)**

If a quantity is raised to the power n , the relative (fractional) error is multiplied by n .

$$B = A^n, \delta B/B = n \delta A/A.$$

Remember that roots are just fractional powers ($\alpha A = A^{1/2}$).

Example: to find the area of a circle, you measure its diameter to be $D = 2.0 \pm 0.1$ cm. The area is calculated by $A = \pi D^2/4$. The fractional or relative error in D is $0.1/2.0 = 5\%$. Squaring D doubles the relative error, and multiplying by $\pi/4$ does not affect the relative error, so the relative error in area A is 10%. The area is then expressed as $(3.142 \pm 0.314) \rightarrow (3.1 \pm 0.3) \text{ cm}^2$.

- **The General Rule for a Single Quantity**

The rules given above cover the majority of steps in everyday calculation. They are special cases of a general rule whose rationale was mentioned earlier.

If A is a quantity whose measured value is A_0 with error δA , then the error in some calculated quantity B which can be expressed as a function $f(A)$ is approximated as

$$\delta B = \left. \frac{df}{dA} \right|_{A_0} \delta A,$$

except that negative errors are meaningless so we want the absolute value of $\frac{df}{dA}$. One way to write that is

$$\delta B = \sqrt{\left(\left. \frac{df}{dA} \right|_{A_0} \right)^2 \delta A^2}.$$

Remember that if the "function" is actually a series of arithmetic operations, it may be more convenient to apply the rules for each operation in succession, i.e. to literally propagate the error, than to use the general rule.

- **Caution**

Sometimes even physicists, who should know better, apply the rules of error propagation without giving much thought to what they are doing. For an extreme example, suppose that you measure an angle to be $89^\circ \pm 2^\circ$. There's certainly nothing wrong with that. But if you need the tangent of that angle you're in big trouble. If you don't immediately see why, try finding the tangent of each whole number of degrees from $89 - 2$ to $89 + 2$, on your calculator. That range of values can never be approximated by a differential.

4.2 Operations on Two or More Measured Quantities

For several variables (measured quantities) the same general reasoning applies. When appropriate, derivatives become partial derivatives. In order to formulate rules, a new assumption is required, namely that the different errors are independent, or unrelated.

Two purely random errors are automatically independent, as are one random and one systematic error. When two or more systematic errors occur, the rules do not apply if they are related. If

you measure the length and width of a rectangle with the same worn-down meter stick, both values will be too large and you cannot treat them as independent. But when you find a runner's speed, the errors in the distance and in the time, even if they are both systematic, are almost surely independent.

- **The Basic Principle**

Independent Errors Add Quadratically (or Add in Quadrature)

The phrase "add quadratically" expresses a relationship which occurs often in physics. It means always that the square of the result is the sum of the squares of all the other variables. For example you could say that two sides of a right triangle "add quadratically" to give the hypotenuse. Just how this general rule applies varies according to the operation, as before.

- **Adding and Subtracting**

$$X = A \pm B \pm C,$$

$$\delta X = \sqrt{\delta A^2 + \delta B^2 + \delta C^2},$$

or absolute errors add quadratically. Note that the errors increase just the same irrespective of whether the data are added or subtracted. Example: If a race started at 2:32:00 \pm :06 and ended at 2:37:30 \pm :06, then the time was 5:30 and the error was $\sqrt{(6 \text{ sec})^2 + (6 \text{ sec})^2} = 6\sqrt{2} = 8$ or 9 sec if the errors were unrelated.

Application: Averaging. Simple averaging consists of adding up N values and dividing by N. Consider the case where the errors are all the same, say δA , as frequently is the case. Then there are N equal errors, and the error in the sum is $\propto N \delta A$. When the sum is divided by N, so also is the error. Hence the error in the average $\delta \bar{A} = \delta A / \sqrt{N}$. Thus repeating measurements does reduce the error, if the errors are independent, but only in proportion to the square root of the number of measurements.

- **Multiplying and Dividing**

$$X = A \times B \div (C \times D),$$

$$\delta X / X = \sqrt{(\delta A / A)^2 + (\delta B / B)^2 + (\delta C / C)^2 + (\delta D / D)^2}$$

or relative errors add quadratically.

Don't fail to notice that the result of this "quadratic sum" is itself a relative, not absolute, error. Also, it makes no difference whether a particular quantity is a multiplier or divisor.

Example: To estimate how long it will take to paint a large building, you paint a sample wall section, and measure the length L, width W and time T. Your "areal velocity" is then $\Omega = LW/T$. If all three data were measured with 10% error each then

$$\Delta \Omega / \Omega = \sqrt{3 \times (10\%)^2} = 17\%$$

Common sense will tell you something that you can now prove mathematically. Suppose that you wanted to reduce your error, in this example, by improving one of the measurements, say the width of the wall. If the error in width is reduced to 2%, the others remaining the same, the overall error becomes 14%. Heartened by this gain, you work even harder on the width

measurement and eventually reduce its error to virtually zero. Nevertheless your overall error is still 14%. Moral: there is no use improving one factor in an experiment beyond the point where its contribution to the overall error is significant. Because of the quadratic rule, an error which is several times (say 5 x) smaller than another is insignificant. In this case, that applies to relative errors; for adding and subtracting the same reasoning would apply to absolute errors.

- **The General Rule for Several Measured Quantities**

Remember that you can propagate errors through a calculation by applying the particular rule which is applicable at each step. Sometimes, however, it is desirable to come up with a single overall prescription for the error in a derived quantity. The general rule for doing so follows. Here A,B,C, etc. are measured quantities with respective errors δA , δB , δC etc. If X is a result calculated from the measurements, then X is a function of those several variables. With the same assumptions as before, an extension of the same reasoning gives

$$\delta X = \sqrt{(\partial f/\partial A)^2 \delta A^2 + (\partial f/\partial B)^2 \delta B^2 + \dots}$$

Each partial derivative means that the function $f(A,B,C, \dots)$ is differentiated regarding all but the indicated variable temporarily as constants. As in the rule for a single variable, the derivatives must be evaluated at the measured values A_0 , B_0 , C_0 , etc.

You can verify that all of the rules given for either one or several variables are special cases of this rule.

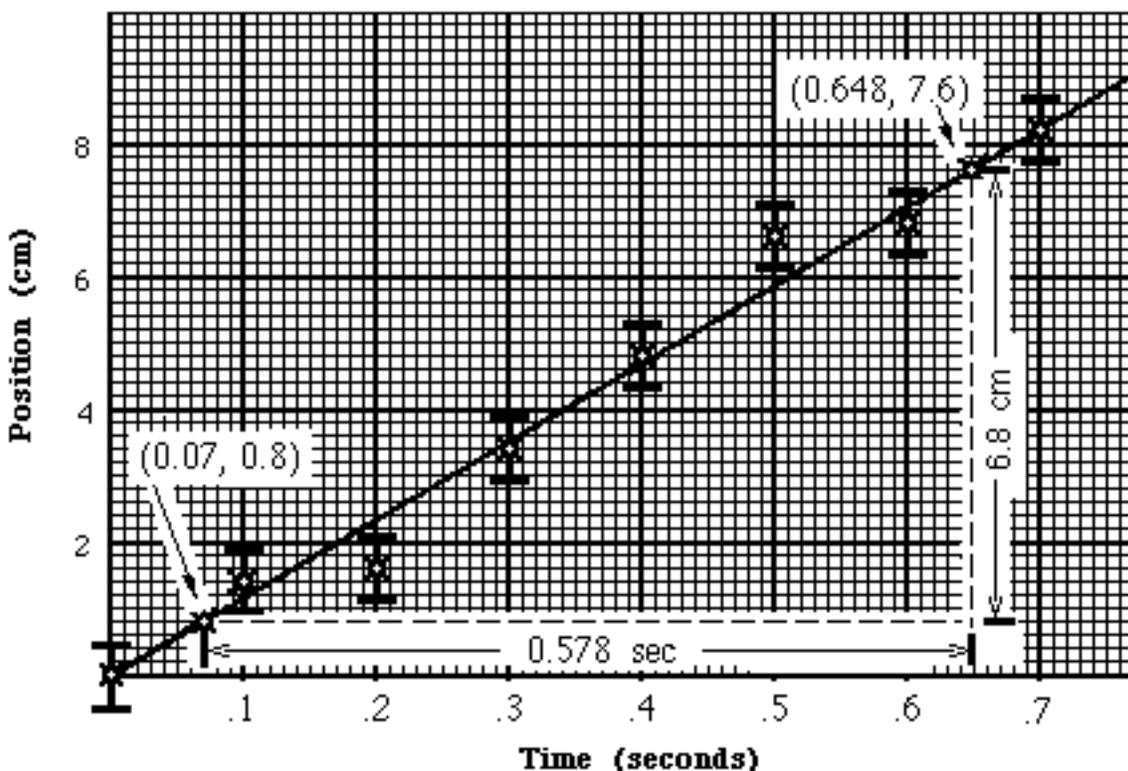
5. Graphical Analysis to Verify Algebraic Equations

A graph is a diagram which, by means of points and lines, allows mathematical relationships to be visualized and consequently more clearly understood. Graphical analysis, the drawing and interpreting of graphs, is essential to physicists as well as to experts in many other fields, both to allow this easy visualization, and to help in determining the form of the mathematical equations which underlie these mathematical relationships.

In this course you will be required to make a number of graphs. This discussion is intended to help you to understand why you are asked to make them, and how you are expected to go about making them.

First we will discuss how to make a graph. In order to illustrate the various elements of a graph and the techniques used in making one, the following example will be used. Imagine a ball rolling along the floor at a constant velocity (no frictional forces are present). If some position in the ball's path is defined to be zero, and if we set the time, t , equal to zero at this point, the position of the ball could be recorded at various times thereafter and this information could be plotted on a graph which would look something like this:

Figure (5-1)
Graph of Position vs. Time for a Ball Rolling along a Floor



5.1 Procedure

- 1) Graph Paper: It is best to make graphs on graph paper. Only in this way can you make an accurate analysis of your data.

- 2) Data Sheet: Before making your graph, you will want to put your data in some clear systematic form, showing the proper units for each quantity and the uncertainties for each measurement (see the pages on error analysis for help in estimating your uncertainties). This way there will be no confusion in your mind about what point you are graphing. Here is an example of a data table for the graph in Figure (5-1).

time (sec ± 0)	distance (cm ± 0.5 cm)	time (sec ± 0)	distance (cm ± 0.5 cm)	time (sec ± 0)	distance (cm ± 0.5 cm)
-					
0.1	1.4	0.4	4.8	0.6	6.8
0.2	1.6	0.5	6.6	0.7	8.2
0.3	3.4				

- 3) Choosing Axes: the first thing you need to decide is what variable to put on each axis. This choice of variables can be very important; often one particular pair of variables will make the relationships you are looking for obvious, while other pairs will leave it obscure. In most experiments, you will be told what variable goes on which axis in the following manner: if you are asked in your instructions to graph a vs. b, the variable before the vs. (here it is a) goes on the

vertical axis and the variable after the vs. (here it is b) goes on the horizontal axis. In Figure (5-1) we have graphed position vs. time; position is on the vertical axis and time is on the horizontal axis. Always label both your axes (showing units) and the entire graph (at the top) for clarity.

4) Choosing Scales: When making a graph, choose your scales carefully. The scales should be chosen so that you can easily see any meaningful variation in the data, but so that random errors are not magnified out of all proportion to their significance.

It is not always essential that each of your axes begin at zero; however, you should always consider carefully whether they should. You will often find, in a given experiment, that the point (0,0) could logically be expected to be one of your data points (even if you didn't actually measure it), in which case it should be included. Since the point (0,0) was defined to be part of our data, it has been included on the graph in Figure (5-1).

5) Error Bars: When plotting data on a graph, it is important, wherever possible, to indicate the uncertainty of each point by using error bars. An error bar is a line passing through the data point and extending from the smallest value which that point could have up to the largest value it could have. An error bar parallel to the vertical axis shows the uncertainty in the variable on that axis, and an error bar parallel to the horizontal axis shows the uncertainty in the variable on that axis. In some cases you will have error bars in only one direction, and error bars can be omitted parallel to the axis for the other variable. Your instructor will tell you when error bars can be omitted in a variable.

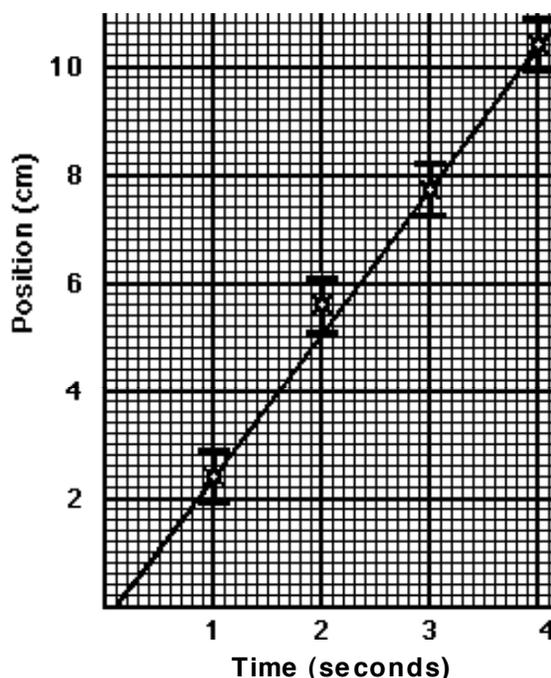
For example, in the data given above, time has been measured using instruments which are very accurate, and for which the uncertainties are smaller than your data points. Thus in Figure (5-1) there are no error bars parallel to the time axis. On the other hand, the uncertainty in the position of the ball is 0.5 cm. This uncertainty is indicated in Figure (5-1) by bars which run parallel to the vertical axis, and have lengths of 1 cm.

Before continuing, it should be emphasized that not every graph you draw will be a straight line like the one in Figure (5-1). Many other curves are possible (parabolas, exponentials, etc.), some of which you will see this term. Because of its simplicity, however, there are several useful pieces of information which can be obtained easily from the graph of a line. We now describe the procedures for obtaining this information. (Note: These procedures should be followed only when your graph is linear.)

6) Finding the Best Straight Line Through a Set of Data Points: Notice in Figure (5-1) that while the data points fall into a pattern which approximates a straight line, there is some random scatter. It is thus not immediately obvious just what line to draw through the data. The usual procedure is to define the best straight line through a set of data points as that line for which the sum of the squares of the deviations is a minimum (the deviation for a particular point is defined as the vertical distance from the point to the line). This is called the "least squares method".

In those experiments where a computer or hand calculator analyzes the data, it will use the least squares method to find the best straight line. It follows a standard mathematical procedure in doing this. The effect of this procedure is to determine a line which is "not too far from any single data point". Since this procedure involves a substantial amount of mathematical calculations, you will not be asked to use it yourself. Rather, in those experiments where you are asked to determine a best line, you will find it by eye, keeping in mind the definition of the "least squares method". There is no set method for such "eyeball fitting"; you must use your judgment to find a line which is not too far from any of the data points. A clear plastic ruler helps. Figures (5-2) and (5-3) show two examples of best straight lines through sets of data.

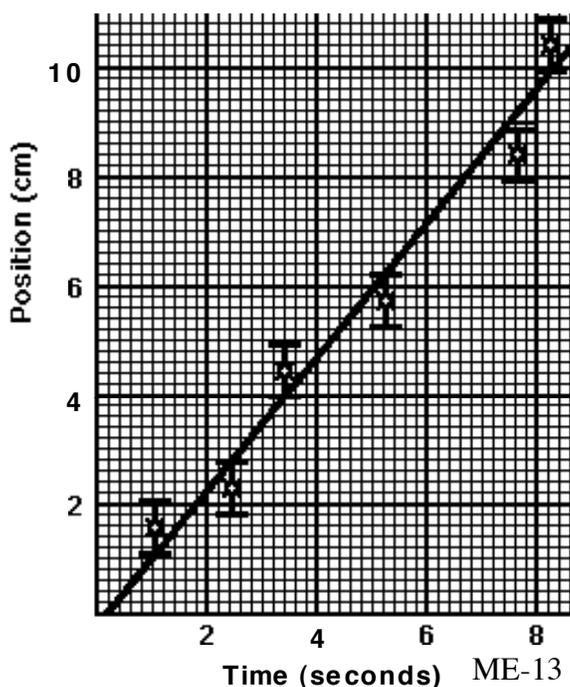
Figure (5-2): Graph of Position vs Time



In Figure (5-2), three points are nearly on a line, while one is not. The best straight line will weight most heavily those points which line up, but it does take into account to a lesser extent the fourth point. The best straight line is slightly above the centers of the three points in a row.

In Figure (5-3), the points lie "approximately" along one line. Here the best straight line represents the average situation. Since the best straight line is an "average" over the data, it need not actually touch any points. Notice that in Figure (5-1) the line only passes directly through two points. Remember also that even if you have estimated your uncertainties correctly, and have not ignored any source of uncertainty, your best line may still pass outside some of the error bars.

Figure (5-3): Graph of Position vs. Time



Once you have determined a best straight line, there are two properties of the line which are of interest: 1) its slope, and 2) its intercept with the vertical coordinate axis. We will now discuss how to find each of these.

7) Finding a Slope: To find the slope of a line, choose two arbitrary points on the line. Do not choose your

data points, since these will seldom actually lie on the line. The farther apart your two points are, the less percent uncertainty you will have in measuring the slope, so choose two points which are not too close together. In Figure (5-1) the points (0.07, 0.8) and (0.648, 7.6) are our two arbitrary points.

The slope you are trying to find is the vertical distance from one point to another (in Figure (5-1) this is the distance from 0.8 to 7.6) divided by the horizontal distance (0.07 to 0.648). This quantity can be calculated using the equation:

$$\text{Slope} = (Y_2 - Y_1) / (X_2 - X_1)$$

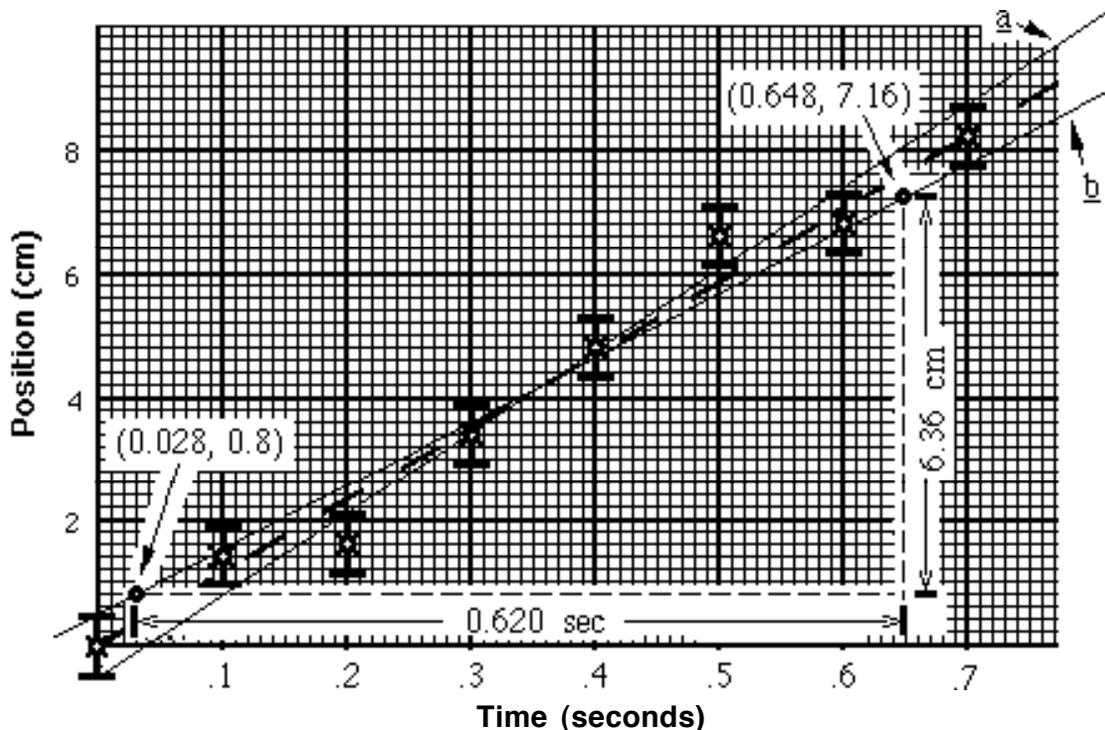
where (X_1, Y_1) and (X_2, Y_2) are the two arbitrary points on your line.

$$(7.6 - 0.8) / (0.648 - 0.07) = 6.8 / 0.578 = 11.8$$

The units of the slope will be the units of the variable on the y-axis divided by the units of the variable on the x-axis: in Figure (5-1) the units of the slope are cm/sec (which, incidentally, are the units of velocity; the graph of position vs. time has a slope which equals the velocity of the ball).

8) Finding the Uncertainty of a Slope: Besides knowing the uncertainties of the variables on your graph, you need some method of estimating, at least roughly, the uncertainty in the slope of your line. Here is one such method. Look at the graph in Figure (5-4) below:

Figure (5-4): Graph of Position vs. Time for a Ball Rolling along a Floor



This is a graph of the same data shown in Figure (5-1), with the dashed line representing the best straight line whose slope we already calculated. In order to find the uncertainty in this slope, you want to estimate the greatest possible variation it could reasonably have. To do this, you should find either the line with the greatest possible slope or the line with the least possible slope which would reasonably fit your data. You may find either of these lines.

In Figure (5-4), line a represents the line with the greatest possible slope and line b represents the line with the least possible slope.

If the best straight line is drawn properly, in most cases when plotting on linear graph paper either line can be used to find the uncertainty in the slope. We will use line b for our example. To find the uncertainty, find the slope of your new line (see topic 7), then find the following difference:

$$\pm [\text{uncertainty in slope}] = \pm ([\text{the slope of the new line}] - [\text{the slope of the best straight line}])$$

In Figure (5-4), the slope of the new line b is 10.3 cm/sec and the slope of the best straight line is 11.8 cm/sec. The uncertainty in slope is then:

$$\pm (10.3 - 11.8) = \pm 1.5 \text{ cm/sec.}$$

In summary, the slope of the best straight line in Figure (5-1) is:

$$(11.8 \pm 1.5) \text{ cm/sec.}$$

9) Finding the Intercept: An intercept is defined as the point where the line on a graph crosses one of the axes. To find the vertical intercept, follow your best fit and extreme lines back until they cross that axis, and record the coordinates of the crossing points. In the example of Figure (5-4) the vertical intercept is (0.0 ± 0.4) cm.