Advanced Physics Laboratory

Physics 370

Fall 2014

Instructor:

Name: Dr. Tom Kirkman   Office: PEngel 111   Phone: 363–3811
email: `tkirkman@csbsju.edu`   Informal Office Hours: 7:30 A.M. – 5:30 P.M.

Texts:

- *An Introduction to Error Analysis*
  by John R Taylor (University Science, 1997)

- `http://www.physics.csbsju.edu/370/`
  `http://www.physics.csbsju.edu/stats/`

Grading:

Your grade will be determined by averaging five scores: 3 lab scores, electronics workshop score and poster score. Lab grades are based on what is recorded in your lab notebook. Please be complete and legible! You will probably need at least 2 of these notebooks (which may be "used") as I'll be grading old labs while you'll working on current labs. Assigned work is generally due at the beginning of the following class period. In particular your lab notebook must be turned in before the lab lecture for the following experiment. (If your lab work is incomplete, you may request an improved grade for a completed report, nevertheless turn in what you have!) All work contributing to your grade must be turned in by: Wednesday 10 December.

Questions:

There is no such thing as a dumb question. Questions asked during lecture or lab do not "interrupt", rather they indicate your interests or misunderstandings. The aim of lab is to do things you've never done before; it's no surprise if you've got questions.

Remember: you are almost never alone in your interests, your misunderstandings, or your problems. Please help your classmates and yourself by asking any question vaguely related to physics lab. If you don't want to ask your question during class, that's fine too: I can be found almost any time on the 100-level floor of Engel Science Center. Note: Mondays and Tuesday mornings look for me in Phys 105 lab.

Times/Locations:

Half of this course will be self-scheduled. I hope many of you will still choose to do that work in the scheduled slot, because you can be then sure to find me (i.e., help) at those times and it will help you avoid the crime of procrastination. However, because of limited lab equipment, in fact you cannot all perform the data collection simultaneously. Of course, data analysis (which usually takes much longer than data collection) can be done simultaneously.

Four weeks are scheduled for each lab! Most of the actual data collection and analysis will take place in the suite of labs across from my office.

Half of this course will meet at the scheduled time: lab lectures, workshops and the poster sessions. We will be using the astronomy lab room (PEngel 319) for lab lectures. If you cannot attend at those times, the responsibility of mastering the material falls on you. (An alternative class time—agreed to by all—would also be fine.) Note that lab lectures typically run a bit more than an hour, which leaves plenty of time to bf start the lab immediately following the lab lecture. Groups will also need to schedule a night lab at the observatory for photometry data collection.

"Do I have to do my lab work during the scheduled lab period?"

The answer is "No, but be forewarned:" last year about 10% of the students in this course received Fs or Ws because they did not turn in the required reports at the required time. If you do not procrastinate and actually put in the scheduled four solid hours[1] of lab work per week, I'll work with you to make sure you complete labs on time. Again: the lab is scheduled for 1 P.M. to 5 P.M. once per week, if you fiddle around in lab and leave at 3:30, you are doing half the required work and 50%=F.

Lab Notebook:

Your lab notebook is the primary, graded work-product for this course. It should represent a *detailed* record of what you have done in the laboratory—complete enough so that you could look back after a year or two and reconstruct your work just using your notebook and this manual.

Your notebook should include your preparation for lab, sketches and diagrams to explain the experiment, data collected, comments on difficulties, self-documented spreadsheets, sample calculations, data analysis, fit reports, final graphs, final results, answers to questions asked in the lab manual, and a critique of the lab. A list of suggested sections can be found in the 191 lab manual.

DO NOT collect data on scratch paper and then transfer to your notebook. Your notebook is to be a running record of what you have done, not a formal (all errors eliminated) report. There will be no formal lab reports in this course. Do not delete, erase, or tear out sections of your notebook that you want to change. Instead, indicate in the notebook what you want to change and why (such information can be valuable later on). Then lightly draw a line through the unwanted section and proceed with the new work.

Be Prepared!

In this "Advanced Lab" you will typically be combining some fairly advanced physics concepts with equally advanced instruments. The 10 minute pre-lab talk from 191/200 is

---

[1]i.e., hours when I'm immediately available to answer to your questions and not counting time spent on computer games, web browsing, waiting for your lab partner, etc.....

now stretched into an hour "lab lecture"; in a four hour "workshop" you will demonstrate your ability to use the electrical instrumentation you spent a whole semester developing in Physics 200. It will be quite easy to be overwhelmed by the theory and the instrumentation. Your main defense against this tsunami of information is to *read and understand the material before the lecture/lab.* I know that this is difficult: technical readings never seems to make sense the first time through. But frankly, one of the prime skills you should be developing (i.e., the prime skill employers seek) is the ability to read, understand, and act on technical documents. In 191 you were told to: *Read aggressively!* Read with a pencil in hand so you can jot down questions, complete missing steps of algebra, and argue with the author. (In this case you can actually take your complaints, comments, and arguments to the author, rather than imagining how the author would respond.) A significant problem is that readings (in contrast to lectures) generally aim at getting the details right. But details obscure the big picture and misdirect attention. This leads to the suggestion of "skimming" the material... which is OK as long as that's just the first step to understanding. I usually instead start by reading for detail, but bit-by-bit my confusion grows and I switch to skimming. But then I repeat the process from the start. After several repeats, I usually reach a point where I'm not making progress, and I find I must do something more active like: talk to somebody about the material, or try to solve a problem—perhaps one of my own design. The aim is to try to find out why the author thinks his points are the important ones.

Posters: A stitch in time saves nine.

Presentation of a lab project as a poster is the final component of this course. While I know procrastination always seems like the easiest course, in fact, putting together a poster months after you've completed the lab is time consuming. The easy course is actually to start your poster (particularly the figures) soon after you've completed the lab. While you can delay final construction, preparation of poster-quality figures immediately following the lab will save you a lot of time just when you most need it (at the end of the semester). See page 141 for basic poster information and that section's references for much more detailed information. Poster topics will be assigned on a first come first served basis, so there is no reason to delay selecting your poster topic.

Schedule:

Tuesday Section

Class		Date	Topics
1	T	Aug 26	Electrical Measurements Lab[a]
2	T	Sep 2	
3	T	Sep 9	Lab Lecture: Bubble Chamber & Photometry[b]
4	T	Sep 16	
5	T	Sep 23	
6	T	Sep 30	
7	T	Oct 7	Lab Lecture: Thermionic Emission, Fortran, GPIB
8	T	Oct 21	
9	T	Oct 28	
10	T	Nov 4	
11	T	Nov 11	Lab Lecture: Langmuir Probe?[c]
12	T	Nov 18	
13	T	Nov 25	
14	T	Dec 2	
15	T	Dec 9	Poster Conference

Thursday Section

Class		Date	Topics
1	R	Aug 28	Electrical Measurements Lab[a]
2	R	Sep 4	
3	R	Sep 11	Lab Lecture: Bubble Chamber & Photometry[b]
4	R	Sep 18	
5	R	Sep 25	
6	R	Oct 2	
7	R	Oct 9	Lab Lecture: Thermionic Emission, Fortran, GPIB
8	R	Oct 16	
9	R	Oct 23	
10	R	Oct 30	
11	R	Nov 6	Lab Lecture: Langmuir Probe?[c]
12	R	Nov 13	
13	R	Nov 20	
14	R	Dec 4	
15	T	Dec 9	Poster Conference[d]

[a]Read "Electrical Measurement Review" *before* the lab! Meet in room PE116.
[b]ASAP schedule a few nights for data collection at observatory
[c]Only if observatory observations were impossible due to weather.
[d]NOTE: this one conference includes both sections

Contents

0: Systematic Error

Physical scientists...know that measurements are never perfect and thus want to know *how* true a given measurement is. This is a good practice, for it keeps everyone honest and prevents research reports from degenerating into fish stories.

Robert Laughlin (1998 Physics Nobel Laureate) p.10 *A Different Universe*

A hypothesis or theory is clear, decisive, and positive, but it is believed by no one but the man who created it. Experimental findings, on the other hand, are messy, inexact things, which are believed by everyone except the man who did the work.

Harlow Shapley, *Through Rugged Ways to the Stars.* 1969

Perhaps the dullest possible presentation of progress[1] in physics is displayed in Figure 1: the march of improved experimental precision with time. The expected behavior is displayed in Figure 1d: improved apparatus and better statistics (more measurements to average) results in steady uncertainty reduction with apparent convergence to a value consistent with any earlier measurement. However frequently (Figs. 1a–1c) the behavior shows a 'final' value inconsistent with the early measurements. Setting aside the possibility of experimental blunders, *systematic error* is almost certainly behind this 'odd' behavior. Uncertainties that produce different results on repeated measurement (sometimes called *random errors*) are easy to detect (just repeat the measurement) and can perhaps be eliminated (the standard deviation of the mean $\propto 1/N^{1/2}$ which as $N \to \infty$, gets arbitrarily small). But systematic errors do not telegraph their existence by producing varying results. Without any tell-tale signs, systematic errors can go undetected, much to the future embarrassment of the experimenter. This semester you will be completing labs which display many of the problems of non-random errors.

Experiment: Measuring Resistance I

Consider the case of the digital multimeter (DMM). Typically repeated measurement with a DMM produces exactly the same value—its random error is quite small. Of course, the absence of random error does not imply a perfect measurement; Calibration errors are

---

[1]Great advancements is physics (Newton, Maxwell, Einstein) were not much influenced by the quest for more sigfigs. Nevertheless, the ability to precisely control experiments is a measure of science's reach and history clearly shows that discrepant experiments are a goad for improved theory.
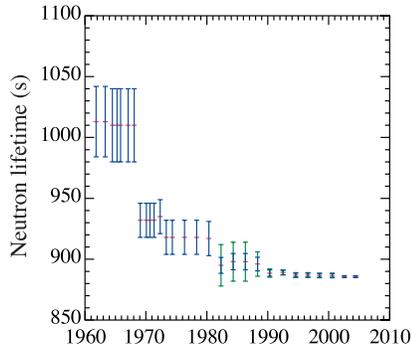
(a) Neutron lifetime vs. Publication Date
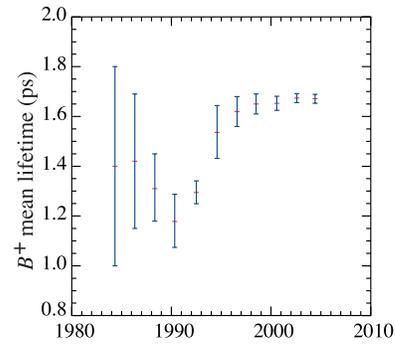
(b) B^+ lifetime vs. Publication Date
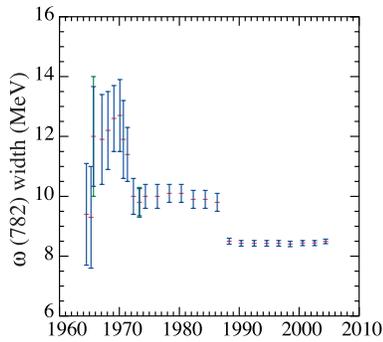
(c) ω width vs. Publication Date
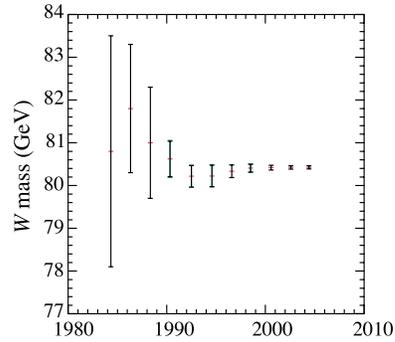
(d) W mass vs. Publication Date

Figure 1: Measured values of particle properties 'improve' with time, but 'progress' is often irregular. The error bars (δx) are intended to be '$\pm 1\sigma$': the actual value should be in the range $x \pm \delta x$ 68.3% of the time (if the distribution were normal) and in the range $x \pm 2\delta x$ 95.4% of the time. These figures are from the Particle Data Group, pdg.lbl.gov.

V	δV	I	δI
(V)	(V)	(mA)	(mA)
0.9964	.005	0.2002	.001
2.984	.007	0.6005	.003
4.973	.009	1.0007	.005
6.963	.011	1.4010	.007
8.953	.013	1.8009	.009
10.942	.015	2.211	.012
−0.9962	.005	−0.1996	.001
−2.980	.007	−0.6000	.003
−4.969	.009	−1.0002	.005
−6.959	.011	−1.4004	.007
−8.948	.013	−1.8001	.009
−10.938	.015	−2.206	.012
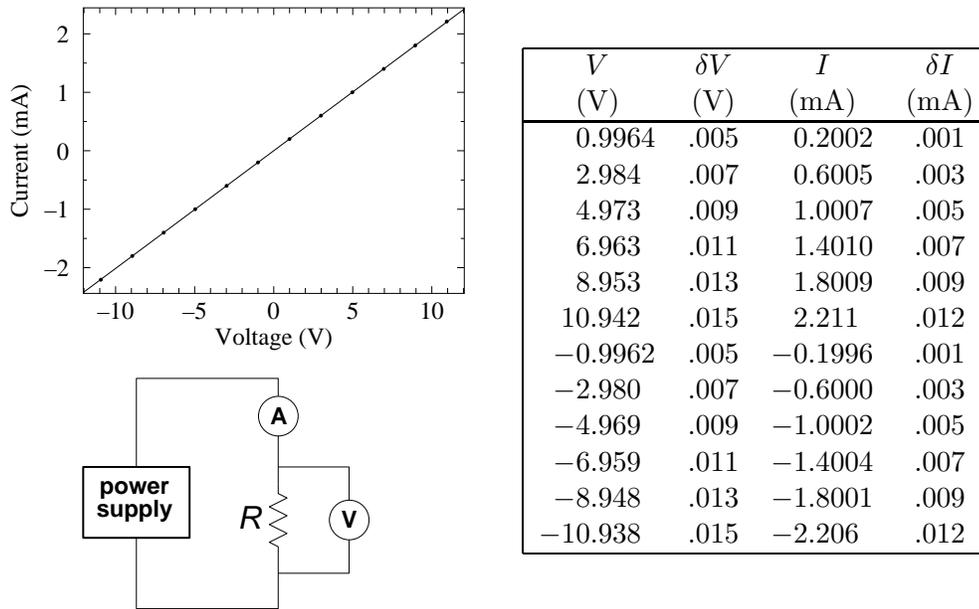
Figure 2: A pair DM-441B DMMs were used to measure the voltage across (V) and the current through (I) a resistor stated to be "4.99 kΩ"

expected and reported in the device's specifications. Using a pair of DM-441B multimeters, I measured the current through and the voltage across a resistor. (The circuit and results are displayed in Figure 2.) Fitting the expected linear relationship ($I = V/R$), Linfit reported $R = 4.9696 \pm .0016$ kΩ (i.e., a relative error of 0.03%) with a reduced χ^2 of .11. (A graphical display showing all the following resistance measurements appears in Figure 3. It looks quite similar to the results reported in Figs. 1.)

This result is wrong and/or misleading. The small reduced χ^2 correctly flags the fact that the observed deviation of the data from the fit is much less than what should have resulted from the supplied uncertainties in V and I (which were calculated from the manufacturer's specifications). Apparently the deviation between the actual voltage and the measured voltage does not fluctuate irregularly, rather there is a high degree of consistency of the form:

$$V_{\text{actual}} = a + bV_{\text{measured}} \tag{1}$$

where a is small and $b \approx 1$. This is exactly the sort of behavior expected with calibration errors. Using the manufacturer's specifications (essentially $\delta V/V \approx .001$ and $\delta I/I \approx .005$) we would expect any resistance calculated by V/I to have a relative error of $\sqrt{.1^2 + .5^2} = .51\%$ (i.e., an absolute error of $\pm.025$ kΩ for this resistor) whereas Linfit reported an error 17 times smaller. (If the errors were unbiased and random, Linfit could properly report some error reduction due to "averaging:" using all $N = 12$ data points—perhaps an error reduction by a factor of $N^{1/2} \approx 3.5$—but not by a factor of 17.) Linfit has ignored the systematic error that was entered and is basing its error estimate just on the deviation between data and fit. (Do notice that Linfit warned of this problem when it noted the small reduced χ^2.)
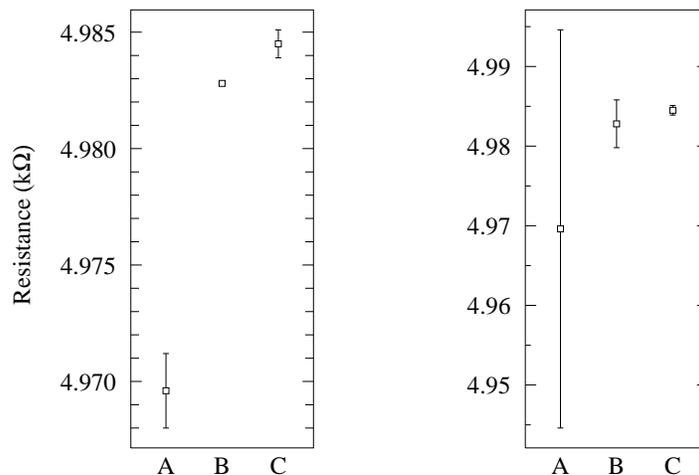
Figure 3: Three different experiments are used to determine resistance: (A) a pair of DM-441B: V/I, (B) a pair of Keithley 6-digit DMM: V/I, (C) a Keithley 6-digit DMM direct R. The left plot displays the results with error bars determined from Linfit; the right plot displays errors calculated using each device's specifications. Note that according to Linfit errors the measurements are inconsistent whereas they are consistent using the error directly calculated using each device's specifications.

When the experiment was repeated with 6-digit meters, the result was $R = 4.9828 \pm .0001$ kΩ with a reduced χ^2 of .03. (So calibration errors were again a problem and the two measurements of R are inconsistent.) Direct application of the manufacturer's specifications to a V/I calculation produced a 30× larger error: $\pm.003$ kΩ

A direct measurement of R with a third 6-digit DMM, resulted in $R = 4.9845 \pm .0006$ kΩ.

Notice that if Linfit errors are reported as accurate I will be embarrassed by future measurements which will point out the inconsistency. On the other hand direct use of calibration errors produces no inconsistency. (The graphical display in Figure 3 of these numerical results is clearly the best way to appreciate the problem.) How can we know in advance which errors to report? Reduced χ^2 much greater or much less than one is always a signal that there is a problem with the fit (and particularly with any reported error).

Lesson: Fitting programs are designed with random error in mind and hence do not properly include systematic errors. When systematic errors dominate random errors, computer reported 'errors' are some sort of nonsense.

Comment: If a high precision resistance measurement is required there is no substitute for making sure that when the DMM reads 1.00000 V the actual voltage is also 1.00000 V. Calibration services exist to periodically (typically annually) check that the meters read true. (However, our SJU DMMs are not calibrated periodically.)

Warning: Statistics seems to suggest that arbitrarily small uncertainties can be obtained simply by taking more data. (Parameter uncertainties, like the standard deviation of the

mean, will approach zero in proportion to the inverse square-root of the number of data points.) This promise of asymptotic perfection is based on the assumption that errors are exactly unbiased — so that with a large number of data points the errors will cancel and the underlying actual mean behavior will be revealed. However, in real experiments the errors are almost never unbiased; systematic errors cannot generally be removed by averaging. Care is always required in interpreting computer reported uncertainties. You must always use your judgment to decide if your equipment really has the ability to determine the parameters to accuracy suggested by computer analysis. You should particularly be on your guard when large datasets have resulted in errors much smaller than those reported for the individual data points.

Measure Twice: Systematic Error's Bane

In the thermionic emission lab you will measure how various properties of a hot tungsten wire are affected by its temperature. The presence of some problem with the thermionic lab measurements is revealed by the odd reduced χ^2 in fits, but how can we determine which measurements are the source of the problem? Systematic errors are most commonly found by measuring the same quantity using two different methods and not getting the same result. (And this will be the approach in this course: you will often be asked to measure a quantity (e.g., path length, temperature, plasma number density) using two different methods, and find different answers.) Under these circumstances we can use the deviation between the two different measurements as an estimate for the systematic error. (Of course, the error could also be even larger than this estimate!)

Problem of Definition

Often experiments require judgment[2]. The required judgments often seem insignificant: Is this the peak of the resonance curve? Is A now lined up with B? Is the image now best in focus? Is this the start and end of one fringe? While it may seem that anyone would make the same judgments, history has shown that often such judgments contain small observer biases. "Problem of definition errors" are errors associated with such judgments.

Historical Aside: The "personal equation" and the standard deviation of the mean.

Historically the first attempts at precision measurement were in astrometry (accurate measurement of positions in the sky) and geodesy (accurate measurement of positions on Earth). In both cases the simplest possible measurement was required: lining up an object of interest with a crosshair and recording the data point. By repeatedly making these measurements, the mean position was very accurately determined. (The standard deviation of the mean is the standard deviation of the measurements divided by the square root of the number of measurements. So averaging 100 measurements allowed the error to be reduced by a factor of 10.) It was slowly (and painfully: people were fired for being 'poor' observers) determined that even as simple an observation as lining up A and B was seen differently by different people. Astronomers call this the "personal equation": an extra adjustment to be

---

[2]In his book *The Signal and The Noise* (2012) the economist/blogger Nate Silver writes: "Wherever there is human judgment there is the potential for bias".

made to an observer's measurements to be consistent with other observers' measurements. This small bias would never have been noticed without the error-reduction produced by averaging. Do notice that in this case the mean value was not the 'correct' value: the personal equation was needed to remove unconscious biases. Any time you use the standard deviation of the mean to substantially reduce error, you must be sure that the random component you seek to remove is exactly unbiased, that is the mean answer is the correct answer.

In the bubble chamber lab, you will make path-length measurements from which you will determine a particle's mass. Length measurements (like any measurement) are subject to error, say 0.1 mm. A computer will actually calculate the distance, but you have to judge (and mark) the beginning and end of the paths. The resulting error is a combination of instrument errors and judgment errors (problem of definition errors). Both of these errors have a random component and a systematic component (calibration errors for the machine, unconscious bias in your judgments). A relatively unsophisticated statistical treatment of these length measurements produces a rather large uncertainty in the average path length (and hence in the particle's mass calculated from this length). However, a more sophisticated treatment of the same length data produces an incredibly small estimated length error much less than 0.1 mm. Of course it's the aim of fancy methods to give 'more bang for the buck' (i.e., smaller errors for the same inputs), however no amount of statistical manipulation can remove built in biases, which act just like systematic (non-fluctuating) calibration errors. Personal choices about the exact location of path-beginning and path-end will bias length measurements, so while random length errors can be reduced by averaging (or fancy statistical methods), the silent systematic errors will remain.

Experiment: Measuring Resistance II

If the maximum applied voltage in the resistance experiment is increased from ± 10 V to ± 40 V a new problem arises. The reduced χ^2 for a linear fit balloons by a factor of about 50. The problem here is that our simple model for the resistor $I = V/R$ (where R is a constant) ignores the dependence of resistance on temperature. At the extremes of voltage (± 40 V) about $\frac{1}{3}$ W of heat is being dumped into the resistor: it will not remain at room temperature. If we modify the model of a resistor to include power's influence on temperature and hence on resistance, say:

$$I = \frac{V}{k_1(1 + k_2 V^2)} \tag{2}$$

(where fitting constant k_1 represents the room temperature resistance and k_2 is a factor allowing the electrical power dissipated in the resistor to influence that resistance), we return to the (too small) value of reduced χ^2 seen with linear fits to lower voltage data. However even with this fix it is found that the fit parameters depend on the order the data is taken. Because of 'thermal inertia' the temperature (and hence the resistance) of the resistor will lag the $t \to \infty$ temperature: T will be a bit low if the resistor is heating up during data collection or a bit high if the resistor is cooling down. The amount of this lag will depend on the amount of time the resistor is allowed to equilibrate to a new applied voltage. Dependence of data on history (order of data collection) is called *hysteresis*.

You might guess that the solution to this 'problem' is to always use the most accurate model of the system under study. However it is known that that resistance of resistors depends on

pressure, magnetic field, ambient radiation, and its history of exposure to these quantities. Very commonly we simply don't care about things at this level of detail and seek the fewest possible parameters to 'adequately' describe the system. A resistor subjected to extremes of voltage does not actually have <u>a</u> resistance. Nevertheless that single number does go a long way in describing the resistor. *With luck*, the fit parameters of a too-simple model have some resemblance to reality. In the case of our Ohm's law resistance experiment, the resulting value is something of an average of the high and low temperature resistances. However, it is unlikely that the computer-reported error in a fit parameter has any significant connection to reality (like the difference between the high and low temperature resistances) since the error will depend on the number of data points used.

The quote often attributed[3] to Einstein: "things should be made as simple as possible, but not simpler" I hope makes clear that part of art of physics is to recognize the fruitful simplifications.

Lesson: We are always fitting less-than-perfect theories to less-than-perfect data. The meaning of of the resulting parameters (and certainly the error in those parameters) is never immediately clear: judgment is almost always required.

The Spherical Cow

> I conceive that the chief aim of the physicist in discussing a theoretical problem is to obtain 'insight' — to see which of the numerous factors are particularly concerned in any effect and how they work together to give it. For this purpose a legitimate approximation is not just an unavoidable evil; it is a discernment that certain factors — certain complications of the problem — do not contribute appreciably to the result. We satisfy ourselves that they may be left aside; and the mechanism stands out more clearly freed from these irrelevancies. This discernment is only a continuation of a task begun by the physicist before the mathematical premises of the problem could even be stated; for in any natural problem the actual conditions are of extreme complexity and the first step is to select those which have an essential influence on the result — in short, to get hold of the right end of the stick.
>
> A. S. Eddington, *The Internal Constitution of the Stars*, 1926, pp 101–2

As Eddington states above, the real world is filled with an infinity of details which a priori might affect an experimental outcome (e.g., the phase of the Moon). If the infinity of details are all equally important, science cannot proceed. Science's hope is that a beginning may be made by striping out as much of that detail as possible ('simple as possible'). If the resulting model behaves —at least a little bit— like the real world, we *may* have a hold on the right end of the stick.

The short hand name for a too-simple model is a "spherical cow" (yes there is even a book with that title: Clemens QH541.15.M34 1985). The name comes from a joke that every physicist is required to learn:

---

[3] "The supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience" p.9 *On the Method of Theoretical Physics* is an actual Einstein quote, if not as pithy—or simple.

> Ever lower milk prices force a Wisconsin dairy farmer to try desperate—even crazy—methods to improve milk production. At the end of his rope, he drives to Madison to consult with the greatest seer available: a theoretical physicist. The physicist listens to him, asks a few questions, and then says he'll take the assignment, and that it will take only a few hours to solve the problem. A few weeks later, the physicist phones the farmer, and says "I've got the answer. The solution turned out to be a bit more complicated than I thought and I'm presenting it at this afternoon's Theory Seminar". At the seminar the farmer finds a handful of people drinking tea and munching on cookies—none of whom looks like a farmer. As the talk begins the physicist approaches the blackboard and draws a big circle. "First, we assume a spherical cow..." (Yes that is the punch line)

One hopes (as in the spherical cow story) that approximations are clearly reported in derivations. Indeed, many of the 'problems' you'll face this semester stem from using high accuracy test equipment to test an approximate theory. (It may be helpful to recall the 191 lab on measuring the kinetic coefficient of friction in which you found that accurate measurement invalidated $F = \mu_k N$ where μ_k was a constant. Nevertheless 'coefficient of friction' is a useful approximation.)

For example, in the Langmuir's probe lab we assume that the plasma is in thermal equilibrium, i.e., that the electrons follow the Maxwell-Boltzmann speed distribution and make a host of additional approximations that, when tested, turn out to be not exactly true. In that lab, you will find an explicit discussion of the error (20% !) in the *theoretical* equation Eq. 7.53.

$$J_i \approx \frac{1}{2}\, en_\infty \sqrt{\frac{kT_e}{M_i}} \tag{3}$$

Again this 'error' is not a result of a measurement, but simply a report that if the theory is done with slightly different simplifications, different equations result. Only rarely are errors reported in theoretical results, but they almost always have them! (Use of flawed or approximate parameters is actually quite common, particularly in engineering and process control—where consistent conditions rather than fundamental parameters are the main concern.)

What can be done when the model seems to produce a useful, but statistically invalid fit to the data?

0. **Use it!** Perhaps the deviations are insignificant for the engineering problem at hand, in which case you may not care to explore the reasons for the 'small' (compared to what matters) deviations, and instead use the model as a 'good enough' approximation to reality.

1. **Find a model that works.**[4] This obvious solution is always the best solution, but often (as in these labs) not a practical solution, given the constraints.

2. **Monte Carlo simulation of the experiment.** If you fully understand the processes going on in the experiment, you can perhaps simulate the entire process on a computer:

---

[4]*Design Report: National Accelerator Laboratory* [a.k.a. Fermilab] (1968, p. 2-2): "That nature is more complex than first expected is a challenge rather than a disappointment".

the computer simulates the experimental apparatus, producing simulated data sets which can be analyzed using the flawed model. One can detect differences (biases and/or random fluctuation) between the fit parameters and the 'actual' values (which are known because they are set inside the computer program).

3. **Repeat the experiment and report the fluctuation of the fit parameters.** In some sense the reporting of parameter errors is damage control: you can only be labeled a fraud and a cheat if, when reproducing your work, folks find results outside of the ballpark you specify. You can play it safe by redoing the experiment yourself and finding the likely range (standard deviation) of variation in fit parameters. In this case one wants to be careful to state that *parameter values* are being reported not physical parameters (e.g., 'indicated temperature' rather than actual temperature). Again, since systematic errors do not result in fluctuation, the likely deviation between the physical parameters and the fit parameters is not known. This was the approach used in the 191 μ_k experiment.

4. **Use bootstrapping[5] to simulate multiple actual experiments.** Bootstrapping 'resamples' (i.e., takes subsets) from the one in-hand data set, and subjects these subsets to the same fitting procedure. Variation in the fit parameters can then be reported as bootstrap estimates of parameter variation. The program `fit` can bootstrap. (Again: report that an unknown amount of systematic error is likely to be present.)

5. **Fudge the data.**

> In dire circumstances, you might try scaling all your x and y error bars by a constant factor until the probability is acceptable (0.5, say), to get plausible values for σ_A and σ_B.
>
> *Numerical Recipes* by Press, et al., 3rd ed. p. 787

Increase the size of your error bars so you get reduced $\chi^2 = 1$, and then calculate errors as in the usual approach. Clearly this is the least legitimate procedure (but it is what *LINFIT* does). One must warn readers of the dicey nature of the resulting error estimates. The program `fit` can fudge.

Special Problem: Temperature

Measuring temperature is a particular problem. (This semester you'll do two labs that involve measuring temperatures above 1000 K in situations a bit removed from the experimenter.) You may remember from 211 that while temperature is a common part of human experience, it has a strikingly abstruse definition:

$$\frac{1}{kT} \equiv \frac{\partial \ln \Omega}{\partial E} \tag{4}$$

While the usual properties of Newtonian physics (mass, position, velocity, etc.) exist at any time, temperature is a property that exists contingent on a situation: 'thermal equilibrium'. And thermal equilibrium is an idealization only approximately achieved—never exactly

---

[5]wiki Bootstrapping (statistics)

achieved—in real life. Furthermore in these experiments, thermal equilibrium is not even closely approximated, so the resulting temperatures have somewhat restricted meanings.

In the photometry lab 'the temperature of stars' is measured. In fact stars do not have <u>a</u> temperature and are not in thermal equilibrium. Nevertheless, astronomers find it useful to define an 'effective temperature' which is really just a fit parameter that is adjusted for the best match between the light produced by the star and the light predicted by the model.

Special Problem: Assuming Away Variation

In the 191 μ_k lab, you assumed the sliding motion was characterized by <u>one</u> value of μ_k, whereas a little experimentation finds usually slippery and sticky locations (handprints?). In the thermionic emission lab you will measure how various properties of a hot wire depend on temperature, however the hot wire does not actually have <u>a</u> temperature: near the supports the wire is cooled by those supports and hence is at a lower temperature. Our spherical cow models have simplified away actual variation. The hope is that the fit model will thread between the extremes and find something like the typical value. Of course, real variations will result in deviations-from-fit which will be detected if sufficiently accurate measurements are made.

Special Problem: Derive in Idealized Geometry, Measure in Real Geometry

Often results are derived in simplified geometry: perfect spheres, infinite cylinders, flat planes, whereas measurements are made in this imperfect world. In these labs (and often in real life) these complications are set aside; instead of waiting for perfect theory, experiment can test if we have "the right end of the stick". Thus a Spherical Cow is born. The theory should of course be re-done using the actual geometry, but often such calculations are extremely complex. Engineering can often proceed perfectly adequately with such a first approximation (with due allowance for a safety factor) and, practically speaking, we simply may not need accuracy beyond a certain number of sigfigs. Indeed it takes a special breed of physicist to push for the next sigfig; such folks are often found in national standards labs like `nist.gov`.

Special Problem: Transducer Calibration

High precision measurement is most commonly electrical in nature: voltmeters, frequency counters, etc.... Transducers are devices that convert a quantity of interest (e.g., pressure, temperature, light intensity) into an electrical quantity for precise measurement. The formula that relates the measured electrical quantity back to the quantity-of-interest is called a calibration. Commonly the uncertainty in calibration vastly exceeds the uncertainty in the electrical measuring device. Thus the error in, say, pressure δP has nothing to do with the error in, say, the voltmeter δV. This can come about for a variety of reasons. For example, the calibration may assume that there is a simple single-variable relationship between P and V ($P = f(V)$), whereas in fact there may be an uncontrolled second variable involved (as in $P = f(V, X)$). Uncontrolled variations in X produce mis-calculated P. Commonly a complex relationship between the quantities has been over-simplified by choice of a simple

formula relating the quantities. In this case the error will be systematic as actual P will be systematically above or below the value calculated from V using the formula.

The Fit Elephant

I remember a public lecture in the late 1970s by the theoretical astrophysicist Don Cox, in which he said

> Give me one free parameter and I'll give you an elephant. Give me two and I'll make it wag its tail

Cox was certainly not the originator[6] of this sort of statement, for example, Freeman Dyson writes[7] that in 1953 Enrico Fermi quoted Johnny von Neumann as saying:

> with four parameters I can fit an elephant and with five I can make him wiggle his trunk

The fit elephant is the opposite of the spherical cow: totally unconstrained parameters are added willy-nilly to the model in order to chase the data. The parameter k_2 in the hot resistor equation (Eq. 2) is potentially such a dangerously free parameter: I will accept any value the computer suggests if only it improves the fit. While I have provided a story which suggests why such a term might be present, I have not actually checked that there is any truth in the story (for example, by measuring the actual temperature of the resistor at high voltage and by measuring the resistance of the resistor when placed in an oven). Skepticism about such inventions is expressed as Occam's razor[8] and the law of parsimony.

Purpose:

In all your physics labs we have stressed the importance of 'error analysis'. However, in this course you will have little use for that form of error analysis (because it was based on computer reports of random errors). Instead, my aim in this course is to introduce you to the problems of non-random error. In the bubble chamber lab you will see how increasingly sophisticated analysis can reveal systematic error not important or evident in more elementary analysis. In the other labs you will see how systematic error can be revealed by measuring the same quantity using different methods. In all of these labs you will use too simple theory to extract characterizing parameters, which are not exactly the same quantity as might occur in a perfect version of the problem.

---

[6]Brown & Sethna, Phys.Rev.E, **68** 021904 (2003), reports attributions to C.F. Gauss, Niels Bohr, Lord Kelvin, Enrico Fermi, and Richard Feynman; I would add Eugene Wigner. The first Google Books hit is in 1950.

[7]*Nature* **427**, 297 (2004), Mayer, el al., Am. J. Phys. **78**, 648 (2010), found the required 4+1 (complex) parameters

[8]"entia non sunt multiplicanda praeter necessitatem", roughly (Wiki) translated as "entities must not be multiplied beyond necessity".

Comment:

The lesson: "measure twice using different methods" is often impractical in real life. The real message is to be constantly aware that the numbers displayed on the meter may not be the truth. Be vigilant; check calibrations and assumptions whenever you can.

> "Many people do not appreciate that measurements which are both reproducible and of low uncertainty can still be inaccurate."

> Sam Benz (NIST, 2013) on the development of a 1 part per billion voltage source.

1: Electrical Measurements Review

Pre-Lab: Read and understand this chapter before coming to lab!

Lab: See page 38. Work as individuals!
Lab in room PE116. Begin: ASAP; Due: before class 3
Complete the following problems: 1a, 4–10, 12, 16–22
Record the results in your notebook.

Reference: Horowitz & Hill, *The Art of Electronics*
Chapter 1 & Appendix A
Diefenderfer & Holton, *Principles of Electronic Instrumentation*
Chapter 6
Paul Scherz, *Practical Electronics for Inventors*, 2nd edition

Purpose

This review aims to re-familiarize you with the electronic instrumentation you used in Physics 200 lab: components like resistors and capacitors, measuring devices like digital multimeters and oscilloscopes, and electrical sources like function generators and d.c. power supplies. In addition the usual schematic symbols for these devices are presented.

Components

In Physics 200 you learned about three passive, linear components: resistors (R), capacitors (a.k.a., condensers, C) and inductors (a.k.a., chokes or coils, L). These devices are called passive because they require no outside power source to operate (and thus circuits involving just these components cannot amplify: at best power in = power out). These devices are called linear because the current *through* these devices is linearly proportional to the voltage *across* them[1]:

$$I = \frac{1}{Z}V \tag{1.1}$$

---

[1]Unless otherwise stated, you should always assume that "voltage" and "current" refer to the root-mean-square (rms) value of that quantity. That is what meters always report. Of course, this equation would also apply to peak or peak-to-peak values as long as they are consistently used.

$$X_C = \frac{1}{2\pi f C} \qquad X_L = 2\pi f L$$
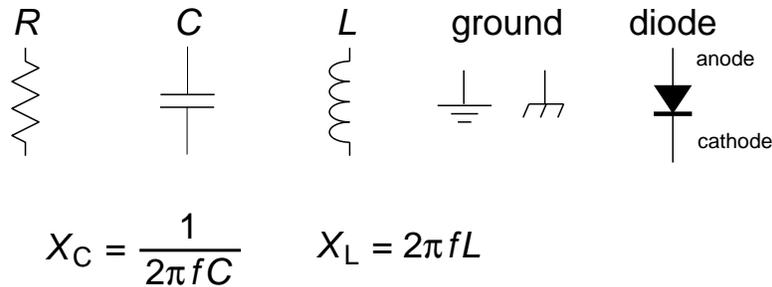
Figure 1.1: The schematic symbols for common components including resistors (R), capacitors (C), and inductors (L). For these three "linear" devices there is a linear relationship between the current *through* the device (I) and the voltage *across* the device (V). For resistors, the resistance, $R = V/I$ is constant. For capacitors and inductors the reactance $X = V/I$ depends on frequency (f) as shown above. The two symbols for ground (zero volts) are, respectfully, earth and chassis ground.

The impedance Z (unit: Ω) determines the proportionality constant. Large impedances (think MΩ) mean small currents (think μA) flow from moderate driving voltages. Small impedances (think 1 Ω) mean large currents (1 A) flow from moderate driving voltages. Impedance[2] is an inclusive term: for resistors the impedance is called resistance; for inductors and capacitors the impedance is called reactance. Inductors and capacitors are useful only in circuits with changing voltages and currents. (Note: changing voltage and/or current = alternating current = a.c.; unchanging current/voltage = direct current = d.c..) The reactance (or impedance) of inductors and capacitors depends on the frequency f of the current. A capacitor's impedance is inversely proportional to frequency, so it impedes low frequency signals and passes high frequency signals. An inductor's impedance is proportional to frequency, so it impedes high frequency currents but passes low frequency currents. Recall that current and voltage do not rise and fall simultaneously in capacitors and inductors as they do in resistors. In the inductors the voltage peaks before the current peaks (voltage leads current, ELI). In capacitors the current peaks before the voltage peaks (current leads voltage, ICE).

Diodes are non-linear passive devices. Positive voltages on one terminal (the anode) results in large current flows; positive voltages on the other terminal (the cathode) results in essentially no current flow. Thus the defining characteristic of diodes is easy current flow in only one direction. The arrow on the schematic symbol for a diode shows the easy direction for current flow. On a diode component a white line often marks which terminal allows easy outward flow.

Light Emitting Diodes (LED) are specialized diodes in which part of the electrical power ($I\Delta V$) is dissipated as light rather than heat. The color of the emitted light (from IR to UV) depends on the material and ΔV.

---

[2]Impedance is often distinguished as being a complex quantity (as in $Z = a + bi$, where $i = \sqrt{-1}$ and $a, b \in \mathbb{R}$). Resistors then have real impedances whereas Z is purely imaginary for inductors and capacitors. This advanced approach is followed in Physics 338. In Physics 200 this reality was hidden behind 'phasers'.
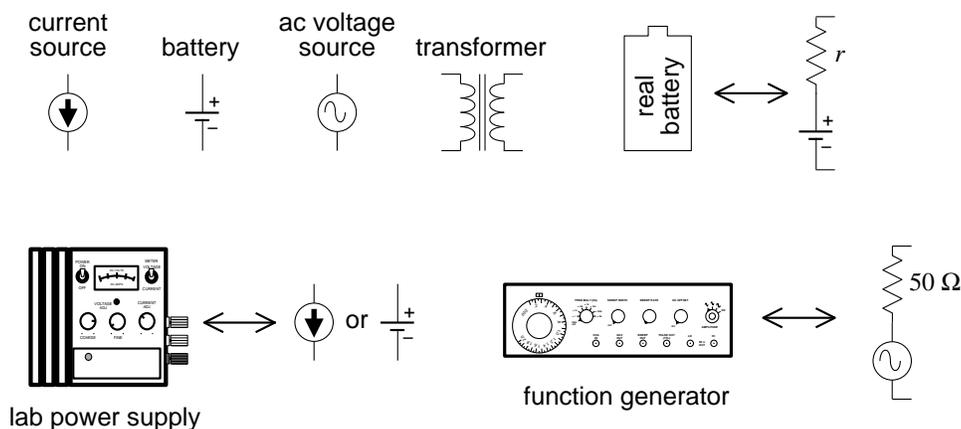
Figure 1.2: The schematic symbols for common electric sources. Real sources can be modeled as ideal sources with hidden resistors. Lab power supplies are fairly close to ideal sources, if operated within specified limits. For example, the Lambda LL-901 specifications report an internal resistance less than 4 mΩ.

Sources

D.C. Current and Voltage Sources

An ideal voltage source produces a constant voltage, independent of the current drawn from the source. In a simple circuit consisting of a voltage source and a resistor, the power dissipated in the resistor (which is the same as the power produced by the voltage source) is V^2/R. Thus as $R \to 0$ infinite power is required. I hope it comes as no surprise that infinite power is not possible, so ideal voltage sources do not exist. Every real voltage source has some sort of current limit built in. (Unfortunately it is not uncommon that the current limiting feature is the destruction the device — beware!!!) Batteries can be thought of as an ideal voltage source in series with a small resistor[3], r, (the internal resistance). The maximum battery current flow (achieved if the external circuit is a "short" i.e., $R \to 0$) is V/r. Laboratory power supplies ("battery eliminators") usually have an adjustable maximum current limit that can be achieved without damaging the device. When this current limit is reached the supplied voltage will be automatically reduced so no additional current will flow. When operating in this mode (current pegged at the upper limit, with actual output voltage varying so that current is not exceeded) the power source is acting as a nearly ideal current source. An ideal current source would produce a constant current, arbitrarily increasing the voltage if that currents meets a big resistance. In a simple circuit consisting of a current source and a resistor, the power dissipated in the resistor (which is the same as the power produced by the current source) is I^2R. Thus as $R \to \infty$ infinite power is required. No surprise: infinite power is not available, so ideal current sources do not exist. Every real current source has some sort of voltage limit built in. Real current sources can be modeled as ideal current sources in parallel with a (large) internal resistance[4].

---

[3]Thévenin's Theorem claims most any two terminal device can be thought of this way!

[4]Norton's Theorem claims most any two terminal device can be thought of this way!

A.C. Voltage Sources

A function generator is a common source of a.c. signals. A function generator can produce a variety of wave shapes (sinusoidal, square, triangle, ...) at a range of frequencies, and can even 'sweep' the frequency (i.e., vary the frequency through a specified range during a specified period). Usually the signals are balanced (i.e., produces as much positive voltage as negative), but a d.c. offset can be added to the signal, for example, producing a voltage of the form

$$A\cos(2\pi ft) + B \tag{1.2}$$

(In this case the d.c. offset would be B, the amplitude would be A, and the peak-to-peak voltage would be $2A$.) Most function generators are designed to have an internal resistance of 50 Ω and maximum voltage amplitude of around 10 V. Generally they have a power output of at most a few watts.

Certainly the most common a.c. source is the wall receptacle: 120 V at a frequency of 60 Hz. Transformers can be used to reduce this voltage to less dangerous levels for lab use. A 'variac' (a variable transformer) allows you to continuously vary the voltage: 0–120 V. Relatively large power (> 100 W) and current (> 1 A) can be obtained in this way. Of course the frequency is not changed by a transformer; it would remain 60 Hz.

Electrical Measurement

Digital Multimeter (DMM)

The most common measurement device is the digital multimeter (DMM). Feel free to call these devices 'voltmeters', but in fact they can measure much more than just volts. For example, the Keithley 169 is fairly generic, measuring volts (a.c. and d.c.), amps (a.c. and d.c.), and ohms. The hand-held Metex M-3800 measures the above and adds transistor h_{FE} and diode test. The bench-top DM-441B measures all of the above and frequency too. The ease of switching measurement options should not blind you to the fact that these measurement options put the DMM into radically different modes. If, for example, the DMM is properly hooked up to measure voltage and — without changing anything else — you switch the DMM to measure amps, most likely something will be destroyed: either the DMM or the device it is connected to. Please be careful!

Recall that voltage (or more properly potential difference) is a measurement of the electrical 'push' applied to an electron as it moves through a section of the circuit. It is analogous to water pressure in that the difference in the quantity determines the driving force. (No pressure difference; no net force.) Note that the presence of big 'push', in no way guarantees that there will be a large resulting flow (current). A large resistance (or for a.c. circuits, impedance) can restrict the flow even in the presence of a big push. In fact, large current flows are often driven by very small voltage differences as a very fat (small resistance) wire is provided for the flow. Wires work by having very small resistance; an ideal wire has zero resistance, and hence nearly zero voltage difference between its two ends.

an ammeter acts
like a short circuit

a voltmeter acts
like a open circuit
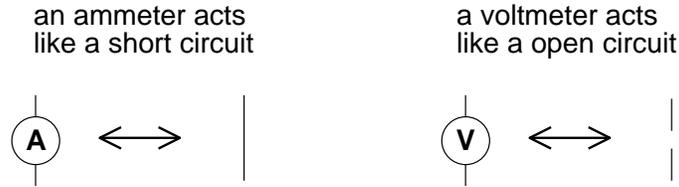
Figure 1.3: The schematic symbols for basic meters. An ammeter must substitute for an existing wire to work properly, whereas a voltmeter can be attached most anywhere.

A voltmeter measures the potential difference across or between the two selected points. A good voltmeter is designed to draw only a small current so it must be equivalent to a large resistance. Modern DDMs typically have input impedances greater than 1 MΩ. Voltmeters with even larger resistance (TΩ) are called electrometers. An ideal voltmeter would draw no current; it would be equivalent to an 'open circuit'. (An open circuit $[R \to \infty]$ is the opposite of 'short circuit' $[R \to 0]$ which is obtained if the two points are connected by an ideal wire.) Since voltmeters draw only a small current, they should not affect the circuit being measured. This makes the voltmeter an excellent diagnostic tool.

Voltmeters always measure the voltage difference between the two probes (e.g., to measure the 'voltage across' a device place probes at opposite ends of the device). More commonly you are asked to measure <u>the</u> voltage at a point. Those words imply that the black (a.k.a. common) probe is to be placed at ground. (In fact you'll soon learn that oscilloscopes are designed as one-probe-at-ground voltmeters.)

Measurement of the current flow through a wire, necessarily requires modification of the circuit. The flow normally going through the wire must be redirected so it goes through the ammeter. This requires breaking the path that the current normally uses, i.e., cutting the wire and letting the ammeter bridge the two now disconnected ends. (With luck, the wire may not need to be literately cut, perhaps just disconnected at one end.) Because current measurements require this modification of the circuit under study, one generally tries to avoid current measurements, and instead substitute a voltage measurement across a device through which the current is flowing. Knowledge of the impedance of the device will allow you to calculate the current from the voltage. Because an ammeter substitutes for a wire, it should have a very small resistance. An ideal ammeter would have zero resistance, i.e., be a short circuit between its two leads. (Note that this is the opposite of a voltmeter, which ideally has an infinite resistance between its two leads.) Real ammeters require a small voltage drop (typically a fraction of a volt for a full scale reading) to operate. This small ΔV is called the *voltage burden*.

I say again: converting a DMM from voltmeter to ammeter makes a drastic change from open circuit to short circuit. Making such a switch in a DMM connected to a circuit usually results in damaging something. Poking around in a circuit with a voltmeter is unlikely to cause damage, because the voltmeter acts like a huge resistor (not that different from the air itself). Poking around in a circuit with an ammeter is quite likely to cause damage, as it is linking two points with a wire, i.e., adding short circuits between points in the circuit.

A DMM measures resistance by forcing a current through the leads and, at the same time, measuring the potential difference between the leads. The resistance can then be calculated by the DMM from $R = V/I$. Note that since a DMM in resistance mode is sending a

current through its leads ('sourcing current') and assuming that this current is the only current flowing through the device, you cannot measure the resistance of a powered device. Furthermore, you can almost never use the DMM to measure the resistance of a device attached to an existing circuit because the current injected by the DMM may end looping back through the circuit rather than through the device. (In addition injecting current into a circuit at random places may damage some components in the circuit.) Thus to measure the resistance of something, you almost always have to disconnect at least one end of it from its circuit.

Lab Reminder: For accurate measurement you must use the appropriate SCALE: the smallest possible without producing an overscale. DMM's may report an overscale condition by a flashing display or a nonsense display like: $\boxed{I.\ \ \ }$. Similarly, for significant DMM measurements you should *record in your notebook every digit displayed by the DMM.*
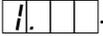
Lab Reminder 2: Manufacturers typically report DMM errors as a percentage of the reading plus a certain number of "digits". In this context, one digit means a 1 in the rightmost displayed digit and zeros everywhere else; two digits means a 2 in the rightmost displayed digit and zeros everywhere else; etc. (One digit usually means the same thing as the 'resolution' of the meter.) Consider a DMM display: 1.707 with an error reported to be "2% + 3 digits". The error is: $1.707 \times .02 + .003 = .037$. Recall the Lab Lint rules for reporting errors: at most 2 sigfigs in the error and the rightmost displayed decimal place in the error matches (in place value) the rightmost decimal place in the number.

A.C. DMM Measurements

Some special considerations are needed when using a DMM to measure a.c. currents or voltages. First, DMMs give accurate readings only for frequencies in a limited range. DMMs fail at low frequencies because DMMs report several readings per second and, in order to be properly measured, the signal needs to complete at least one cycle per reading frame. Thus $f > 20$ Hz or so for accurate readings. At the high frequencies, the input capacitance (~ 100 pF) of the DMM tends to short out the measurement (recall the impedance of a capacitor at high frequency is small). No SJU DMM operates accurately above 0.3 MHz; some DMMs have trouble above 1 kHz. The DMM's manual, of course, reports these specifications.

Recall that a.c. signals are time-variable signals ... there is no steady voltage to report as "the" voltage. The solution is to report root-mean-square ('rms') quantities. (The square root of the average of the square of the voltage.) Since this is a complex thing to calculate, most cheap DMMs assume that the signal is sinusoidal so that there is a relationship between the rms value and the peak value:

$$V_{\text{rms}} = V_{\text{peak}}/\sqrt{2} \tag{1.3}$$

These cheap DMMs find the peak voltage, divide it by $\sqrt{2}$ and report the result as if it were an rms voltage. This of course means the meter reports faulty values if non-sinusoidal signals are applied. "True rms" meters properly calculate the rms quantities.
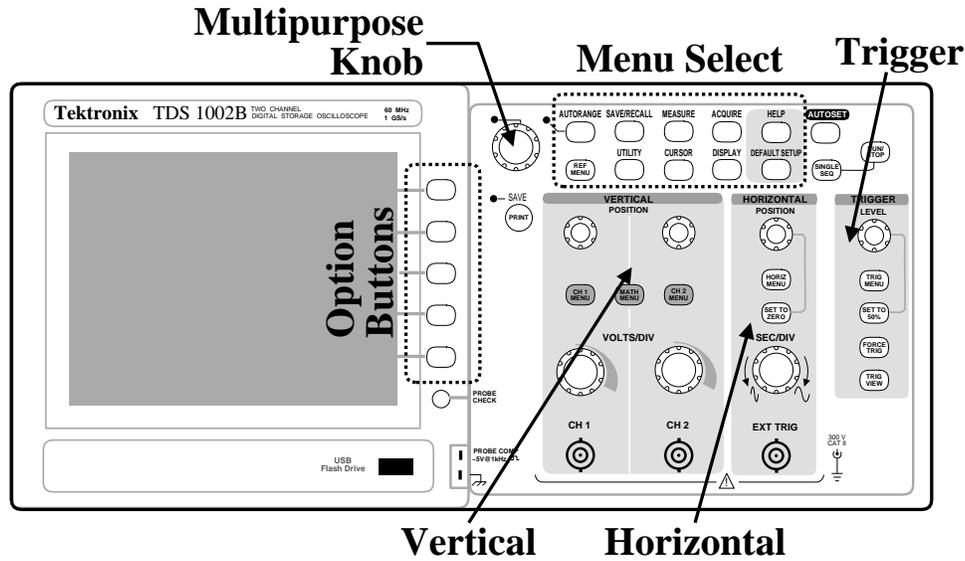
Figure 1.4: The Tektronix TDS 1002B is a two channel digital storage oscilloscope (DSO). Pushing a button in the Menu Select region displays a corresponding menu of the items to be re-configured adjacent to the option buttons. (The multifunction knob allows a continuous variable to be modified.) The VERTICAL region has knobs that control the size (VOLTS/DIV) and POSITION of vertical (y) scales for channel 1 (CH 1) and channel 2 (CH 2). Push buttons in this region control the display of menus for those channels and combinations of those channels (MATH). The HORIZONTAL region has knobs that control the size (SEC/DIV) and POSITION of horizontal (x) scales. In addition to the Main time-base, this dual time-base scope can blow up a selected portion ("Window") of the display. The controls to do this are in the HORIZ MENU. The TRIGGER region has a knob that controls the voltage LEVEL for the triggering and the TRIG MENU button allows the display of configurable options for the trigger. Note particularly the AUTOSET, DEFAULT SETUP, and HELP buttons to the right of the Menu Select region.
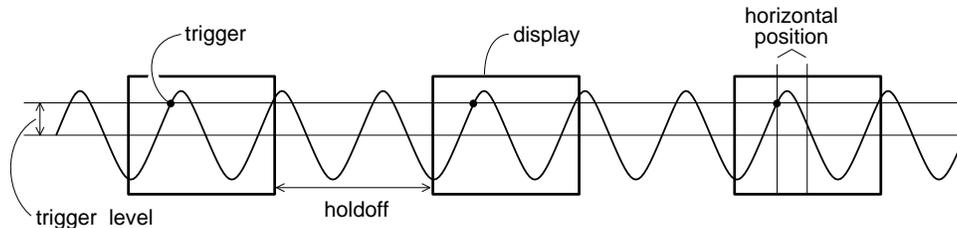
Figure 1.5: An oscilloscope displays one wave-section after another making an apparently steady display. Determining when to start a new wave-section is called triggering. The level and slope of the signal determine a trigger point. The trigger point is placed in the center of the display, but it can be moved using the horizontal POSITION knob. The holdoff is an adjustable dead time following a triggered wave-section.

Oscilloscope

Generally speaking DMMs work in the 'audio' frequency range: 20 Hz – 20 kHz. 'Radio frequency' (rf, say frequencies above 1 MHz) require an alternative measuring device: the oscilloscope (a.k.a., o'scope or scope). Unlike the DMM, the oscilloscope usually measures voltage (not current). Also unlike the DMM, the scope effectively has only one lead: the 'black' lead of the scope is internally connected to ground; the voltage on the 'red' lead is displayed on the screen. (Note: with a DMM you can directly measure the 1 V potential difference between two terminals at 100 V and 101 V. You cannot do this with a scope—its 'black' lead is internally connected to ground so if you connect it's 'black' lead to the 100 V terminal you will cause a short circuit [the 100 V terminal connected to ground through the scope, which will probably damage either the device or the scope].) While a DMM takes a complex waveform and reduces it to a single number: V_{rms}, a scope displays the graph of voltage vs. time on its screen.

Oscilloscopes are generally used to display periodic signals. Every fraction of a second, a new section of the wave is displayed. If these successively displayed wave-sections match, the display will show an apparently unchanging trace. Thus the TRIGGERing of successive wave-sections is critical for a stable display. And an unsteady display (or a display with 'ghosts') is a sign of a triggering problem. In addition, the scales used for the HORIZONTAL and VERTICAL axes should be in accord with the signal. (That is you want the signal to neither be off-scale large or indistinguishable from zero. A too large time (horizontal) scale will result in displaying hundreds of cycles as a big blur; a too small time scale will result in just a fraction of a cycle being displayed.)

Oscilloscope Controls and How To Use Them

Pre-lab Exercise   The knob-filled face of an oscilloscope may appear intimidating, but the controls are organized in a logical and convenient way to help you recall their functions. The class web site contains a line drawing of an oscilloscope (`TDS1002Bscope.pdf`). Print out this diagram and have it in hand as you read this section. As each control is discussed below a circled number (e.g., ①) appears. Find this control on the line drawing and label it with that number. Attach your diagram in your notebook. The name of each control or feature will be printed in SMALLCAPS TEXT.

Display Section The left hand side (lhs) of the scope is dominated by the display or screen ①. Note that there is a USB port ② below the display—this allows you to save scope data and display images on a thumb drive! There is an additional USB port in the back. The power switch is on top of the scope, lhs.

Vertical Sections Right of the option buttons ㉕ – ㉙ are the knobs and buttons that control the vertical portions of the graph. Typically the display shows a graph of voltage (y or vertical) vs. time (x or horizontal). This scope has two BNC[5] inputs so the vertical section is divided into two sections with identical controls for each input. The inputs are called channel 1 (CH 1) ⑤ and channel 2 (CH 2) ⑩. The scale factor for each input is determined by the corresponding VOLTS/DIV knob ⑥ & ⑪; the vertical location of zero volts on the screen is determined by the corresponding POSITION knobs ⑧ & ⑬. Note that the scale factors and the zero location for the channel traces are independently set. Therefore the graph axes can not show values in physical units (volts), rather the graph is displayed on a 8×10 grid with units called divisions. (Note that a division is about a cm.)

The VOLTS/DIV (or sensitivity) knobs are similar to the range switch on a multimeter. To display a 2 volt signal, set the VOLTS/DIV knob to 0.5 volts/div. A trace 4 div high will then be obtained, since 4 div $\times$ 0.5 V/div $= 2$ V. The current settings of these sensitivity knobs is displayed in the lower lhs of the display. *You should always try to adjust the sensitivity so that the signal displayed is at least 3 div peak-to-peak.*

The traces from the two channels look identical on the screen; the symbols 1-▶ and 2-▶ on the lhs of the display show the position of zero volts for each trace. If you are unsure which trace is which, moving a POSITION knob will immediately identify the corresponding trace.

The CH 1 and CH 2 MENU buttons ⑦ & ⑫ produce menus controlling how the corresponding input is modified and displayed. In addition pushing these menu buttons toggles the display/non-display of the corresponding signal.

The top menu item in CH 1 and CH 2 MENUs, is Coupling, with options: DC, AC, Ground. These options control how the signal supplied to the BNC is connected (coupled) to the scope's voltage measuring circuits. The Ground option means the scope ignores the input and instead the display will graph a horizontal line at zero volts (ground) — this allows you to precisely locate and POSITION the zero volt reference line for that channel on the grid. When AC is selected a capacitor is connected between the inputted signal and scope's electronics. This capacitor prevents any d.c. offset voltage from entering the scope's circuits. Thus any d.c. offset will be subtracted from the signal and the display will show the remaining signal oscillating around a mean of zero volts. The usual selection for Coupling is DC, which means the signal is displayed unmodified.

Proper Practice: Use Coupling▶DC almost always. Exceptional circumstances like a d.c. offset larger than an interesting a.c. signal (i.e., $B \gg A$ in Eq. 1.2) or a requirement to MEASURE an rms voltage in the usual way, i.e., with any d.c. offset removed, may force occasional, brief uses of Coupling▶AC, but don't forget to switch back to Coupling▶DC ASAP.

---

[5]According to Wiki, this denotes "bayonet Neill-Concelman" connector. This coaxial cable connector is very commonly used when signals below 1 GHz are being transmitted. You can also use Wiki to learn about banana connectors, RCA plugs, and alligator clips (a.k.a. crocodile clips).
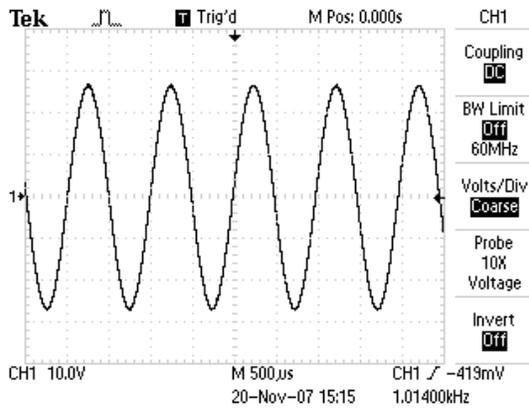
(a) Channel 1 Menu
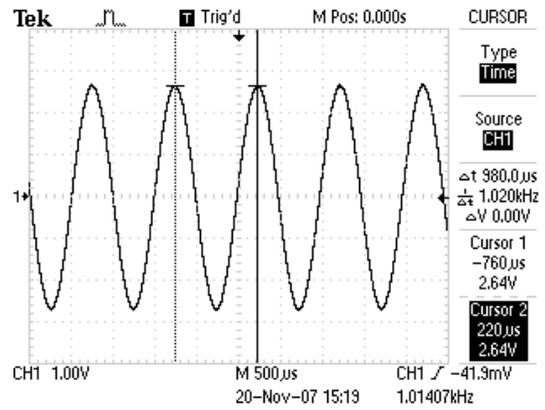
(b) Cursor Menu
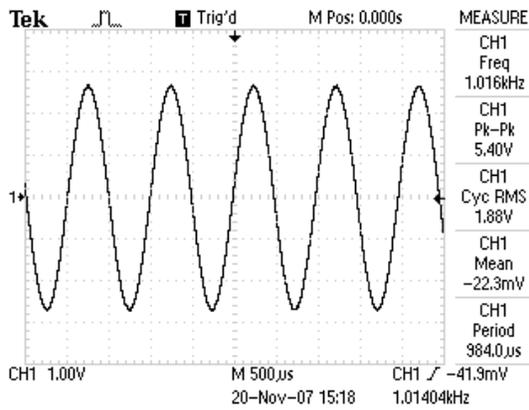
(c) Measure Menu
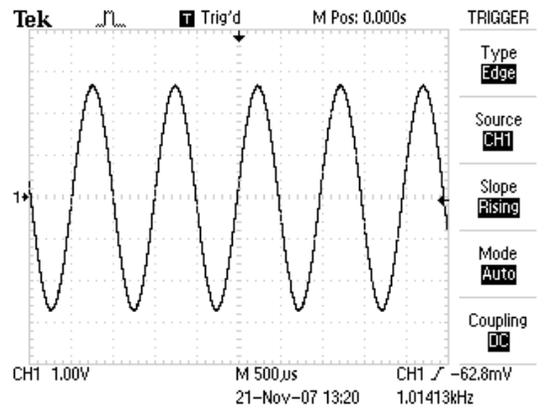
(d) Trigger Menu

Figure 1.6: Buttons (often in the Menu Select region) control which menu appears on the rhs of the display. Here are four examples.

In between the CH 1 and CH 2 MENU buttons, find the MATH MENU ⑨ button. which is used to display combinations of CH 1 and CH 2 ($+, -, \times$) and Fourier transforms (FFT) of either signal.

Probe The second item from the bottom in the CH 1 and CH 2 MENUs is Probe. "Probe" is the name for the device/wire that connects the scope to the signal source. In this course most often your probe will be nothing more complex than a wire, so the choice should be 1X Voltage. Note that this is <u>not</u> the factory default choice (which is 10X Voltage). So one of the first things you should do on turning on a scope, is check that the the probe actually attached to the scope matches what the scope thinks is attached to the scope. (If there is a mis-match, all scope voltage measurements will be wrong.) There is a PROBE CHECK button ③ on the scope to help you establish the attenuation of an unlabeled probe, but usually probes are labeled and it is faster just to immediately set the probe type yourself in the corresponding channel menu. Note: most probes used in this class have a switch to select either $10\times$ or $1\times$ attenuation. All SJU probes are voltage probes.
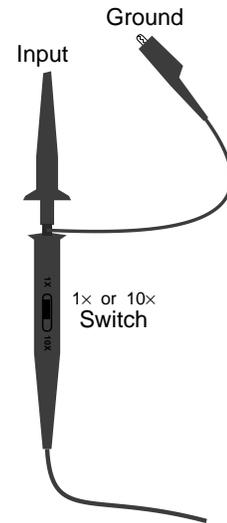
Note: A probe is a scope-only accessory: it should not be attached to any other device. (While a simple wire can act as a scope probe, a scope probe is in general more than a simple wire and therefore should not substitute for a simple wire.)

FYI: The name "$10\times$" on a probe is quite confusing: "$\div 10$" would be a better name as the voltage that reaches the scope has been reduced by a factor of 10. Why reduce a signal before measuring it? Because it reduces the probe's impact on the circuit it is connected to. A $10\times$ probe has a larger impedance (smaller capacitance and larger resistance) than a $1\times$ probe and hence affects the circuit less. This is particularly important for high frequency measurements (which are not the focus of this class).

Horizontal Section In the center-right of the scope face, find the HORIZONTAL section. Just as in the vertical sections, there are knobs that control the horizontal scale (SEC/DIV ⑮) and horizontal POSITION ⑱. In a single time-base scope, all the input channels must be displayed with the same horizontal scale (unlike the vertical scale). In this dual time base scope, a portion of the display can be expanded in a Window. The window controls are found in the HORIZ MENU ⑰. When using the window feature the <u>M</u>ain sec/div setting is labeled M and the <u>W</u>indow sec/div setting is labeled W.

Trigger Section As you might guess, the process of determining when to trigger and display the next wave-section is the most complex part of a scope. Luckily most often the default settings will work OK. Generally you will want to trigger when the wave has reached a particular LEVEL ㉓. But which wave? The TRIG MENU ㉒ allows you to set the Source: CH1, CH2, Ext (the signal connected to the EXT TRIG BNC ⑭ in the horizontal section), Ext/5 (the same signal, but first attenuated by a factor of 5—useful for larger triggering signals), or AC Line (which uses the 60 Hz, 120 V receptacle power line as the triggering signal—useful for circuits that work synchronously with the line voltage). Just as in the vertical section, the Coupling of this source to the triggering electronics can

occur in a variety of ways: subtract the dc offset (AC), filter out (attenuate or remove) high frequency (HF Reject, "high" means > 80 kHz), filter out low frequency (LF Reject, "low" means < 300 kHz), use hysteresis to reduce the effects of noise (Noise Reject), or directly connected (DC). Note: triggering with Coupling▶AC is a common choice as then a LEVEL of zero is sure to match the wave at some point. Similarly Noise Reject is not an uncommon choice. The above options go with Type▶Edge. There are additional sophisticated and useful triggering modes for Type▶Pulse and Type▶Video.

Measure Menu The MEASURE menu ③⑥ allows up to five measurements to be continuously updated and displayed. Push on one of the option buttons and a new menu is displayed allowing you to set the Source: CH1, CH2, MATH, and the Type: Freq, Period, Mean (voltage), Pk-Pk (peak-to-peak, i.e., the full range of voltage from the lowest valley to the highest peak), Cyc RMS (the root-mean-square voltage of the first complete cycle), Min (minimum voltage), Max (maximum voltage), Rise Time, Fall Time (10% to 90% transitions), Pos(itive) Width, Neg(itive) Width (using the 50% level).

Warning: Unlike a DMM on AC, the option Cyc RMS does not subtract the d.c. offset before calculating rms voltage. You can assure yourself of a DMM-like rms result only if the channel is switched to Coupling▶AC.

If a measurement is displayed with a question mark, try switching scales. (Generally the scope wants signals that are several divisions high and complete at least one—but not too many—cycles in the display.)

Cursor Menu The CURSOR menu ③⑦ enables a pair of Type▶Amplitude or Type▶Time measuring lines. With Amplitude cursors, a pair of horizontal lines ("cursors") appears. Hitting the appropriate option button allows the multifunction knob to move each cursor up or down to the required place. The voltage for each cursor is displayed along with the difference (ΔV). With Time cursors, a pair of vertical lines appears. Hitting the appropriate option button allows the multifunction knob to move each line (cursor) right or left to the required place. The voltage and time for each cursor is displayed along with the differences (Δt, ΔV), and frequency $1/\Delta t$.

Display Values The bottom of the display is used to report key numerical values like scale settings. A typical example:

CH1 500mV   CH2 2.00V   M 1.00ms   CH1 ⌐ 0.00V
                                  23-Nov-07 13:03   1.01407kHz

The first two numbers of the first line are the VOLTS/DIV for channels CH1 and CH2, M refers to the main time-base of 1 ms/div, and the final sequence reports that positive edge triggering at a level of 0.00 V is being used with channel 1 as the source. The second line shows the date/time and the frequency of the triggering signal.

Run/Stop In normal operation, the scope is constantly updating the display. It is possible to freeze the display (i.e., take a snapshot of the voltage vs. time graph) using the

RUN/STOP ㊹ or SINGLE SEQ ㊸ buttons.

Hit Me First   The scope remembers option settings between uses. Thus unless you are the sole user of the scope it is wise to set it to a well defined initial state before proceeding. The DEFAULT SETUP ㊶ achieves this goal (but it sets the Probe▶10X Voltage, in the channel menus—which is not usually desired in this class). Similarly the AUTOSET ㊷ button will attempt to make rational choices for scale factors, etc. given the signals connected to the scope. If you want you can save commonly used setups using the SAVE/RECALL ㉞ menu button.

Problems

1. (a) A 1.5 V battery can be modeled as an ideal 1.5 V voltage source in series with a 1 Ω resistor. If this battery is connected to a 10 Ω resistor (see below left), what voltage is actually across the 10 Ω resistor?

   (b) A 1 mA current source can be modeled as an ideal 1 mA current source in parallel with a 1 MΩ resistor. If this source is connected to a 10 kΩ resistor (see below right), how much current actually flows through the 10 kΩ resistor.?
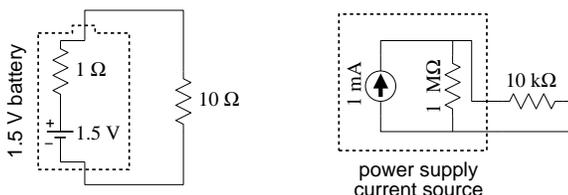

2. (a) A voltage source produces a voltage V when its terminals are disconnected (open circuit). When a device that draws a current I is connected across its terminals, the voltage decreases to $V - \Delta V$. What is the internal resistance?

   (b) A current source produces a current of I when a wire connects the terminals. When a device is instead connected to the terminals the current drops to $I - \Delta I$ and the voltage across the terminals is V. What is the internal resistance?

3. An a.c. voltage source follows the equation: $v(t) = A\sin(\omega t + \phi) + B$
   where $A = 5$ V, $B = 1$ V, $\omega = 1000$ rad/s, and $\phi = .45$ rad. Report values for the following:

   (a) average voltage (d.c. offset)   (e) rms voltage

   (b) voltage amplitude   (f) voltage at $t = 0$

   (c) peak voltage   (g) frequency

   (d) peak-to-peak voltage   (h) period

4. The manual for a stereo amplifier warns that it can be damaged if its "outputs are too heavily loaded". What sort of resistor would constitute a "heavy load": (A) $R = 1$ MΩ or (B) $R = 1$ Ω? Explain!

5. (a) A current of 3 mA flows into a circuit consisting of 3 resistors, and 10 mA flows out (see below left). Report the readings on the three voltmeters. Draw a schematic diagram showing which lead on each voltmeter is the 'red' lead.

   (b) An unknown device is connected in a circuit with a 9 V battery and a 15 kΩ resistor (see below right). The ammeter reads 0.5 mA. What does the voltmeter read?
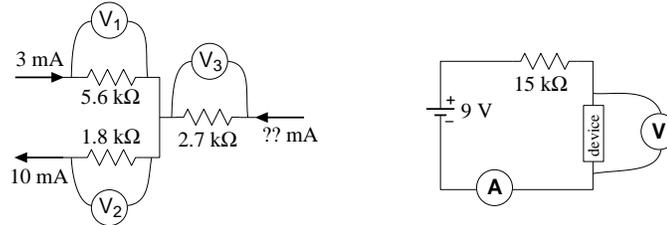

6. What is the resistance (when operating) of a 100 W light bulb operating from a 120 V source?

7. An ideal ammeter should act like a wire and hence have zero volts between its terminals. However real ammeters are less than perfect. The specifications for Keithley 196 in d.c. amps mode reports it has a *voltage burden* of about .15 V when measuring 100 mA on the proper scale. If you use a 196 to measure the current through a 15 Ω resistor powered by a 1.5 V battery, what current does it read?

8. Which light bulb in the below left circuit shines the brightest? Why? Which light bulb shines the dimmest? Why?
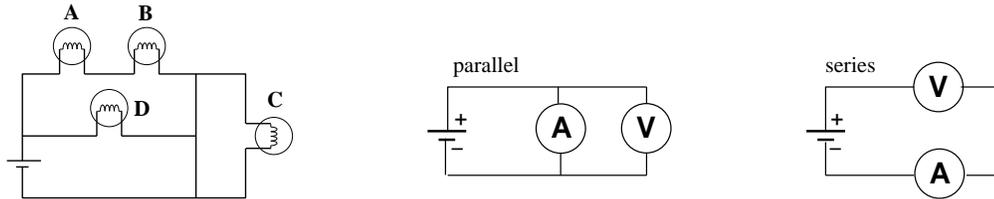
9. (a) An ammeter and a voltmeter are connected in parallel. Are either likely to be damaged? Why? Will either read the current or voltage of the battery? Why?

   (b) An ammeter and a voltmeter are connected is series. Are either likely to be damaged? Why? Will either read the current or voltage of the battery? Why?

10. An ideal voltmeter V_1 is used to measure the voltage across the series combination of a battery and 1 kΩ resistor; an ideal voltmeter V_2 is used to measure the voltage across the battery alone. Which of the below is correct?

   (a) $V_1 > V_2$

    (b) $V_1 = V_2$

    (c) $V_1 < V_2$
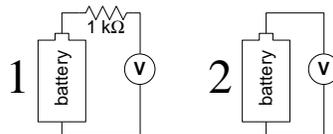
11. The picture (right) shows a circuit in which a battery powers a light bulb.

    (a) Make a careful drawing showing how the voltage produced by the battery could be measured. Include details like exactly where the red and black leads on the voltmeter would be attached.

    (b) Make a careful drawing showing how the current produced by the battery could be measured. Include details like exactly where the red and black leads on the ammeter would be attached.
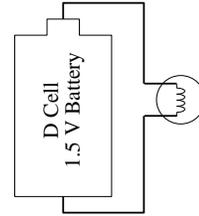
12. The specifications for a Keithley 169 DMM say that, when operating in a.c. volts mode, the inputs look like 1 MΩ in parallel with 100 pF (see right). At what frequency is the current equally shared by the capacitor and the resistor?
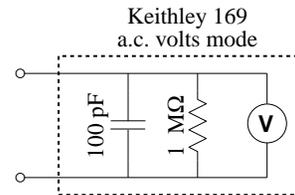
13. A typical lab power supply has knobs labeled VOLTAGE ADJUST and CURRENT ADJUST. If you turn the voltage knob the output voltage changes, but if you turn the current knob nothing seems to change and the current meter continues to read zero. Explain!

14. A function generator has an output impedance of 50Ω and, when unloaded and adjusted to produce its maximum output, produces a voltage amplitude of 10 V. What is the maximum power that can be transferred to an external device attached to the function generator?

15. In the circuit shown right, find the rms voltage drop across each component and phase shift between that voltage and the current.
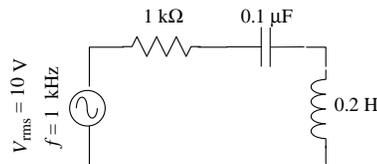
16. Oscilloscope — True or False:

    (a) When the vertical input coupling is set to DC mode, the voltage of an a.c. waveform cannot be measured.

    (b) When the vertical input coupling is set to AC mode, the voltage of a battery cannot be measured.

17. A circuit consists of an inductor (inductance L) connected directly to a 120 V, 60 Hz wall receptacle. What is the smallest L you could use and avoid blowing the 20 A fuse? A similar circuit consists of a capacitor connected directly to a wall receptacle. What is the largest C you could use and avoid blowing the fuse?

18. Manufacturers typically report DMM errors as a percentage of the reading plus a certain number of "digits". In this context, one digit means a 1 in the rightmost displayed digit and zeros everywhere else; two digits means a 2 in the rightmost displayed digit and zeros everywhere else; etc. Consider a DMM display: 0.707. Find the absolute error in this reading if the device is:

    (a) METEX-3800 DC current, 2 mA range.

    (b) METEX-3800 AC current at 500 Hz, 2 A range.

    (c) Sinometer DM-97 resistance, 4 kΩ range.

    (d) Sinometer DM-97 AC volts at 200 Hz, 4 V range.

    The specification sheets can be found posted in the lab (PE116) or in the manuals in the physics library. Please note that errors should ALWAYS be rounded to one or two significant figures.

19. Work the previous problem assuming the display reads: 0.007

20. The section describing oscilloscope controls identified controls with a circled number like: ①. On the class web site, find and print the file `TDS1002Bscope.pdf` which is a line drawing of an oscilloscope. On this hardcopy, locate every control and label each with the appropriate number.

21. The below left diagram shows a single sinusoidal scope trace. Determine: the peak-to-peak voltage, the voltage amplitude, the rms voltage, the wave period and frequency. Assume that the bottom of the scope display reads:

    CH1 500mV   M 1.00ms   Ext ╱ 0.00V
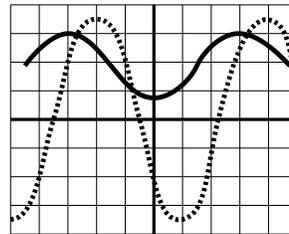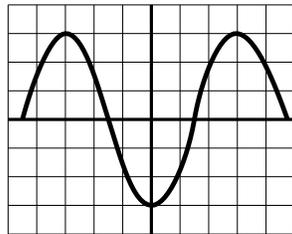
22. The above right diagram shows a pair sinusoidal scope traces. Assume that the scope controls are set as in the previous problem with CH 2 (dotted) and CH 1 (solid) identical. Which trace is lagging: dotted or solid? What is the phase shift in degrees?

2: Electrical Measurements Lab

Electrical Measurement Lab
DC & AC Measurements
Oscilloscope & Filters
Work individually please!

0. **Basic DC measurements.**

1. Select a resistor from the "370 Resistors" cup. Sketch your resistor carefully recording the color of the bands on it. Decode the color bands to find the manufacturer's reported resistance. Hang on to this resistor! You must turn it in with your lab report.

2. Using a Sinometer DM-97 DMM measure the resistance of your resistor. Record the result with an error. Using a Metex 3800 DMM, measure the resistance of your resistor. Record the result with an error. Are your results consistent?

3. Find one of the dual battery packs with black and red leads attached. Using a Sinometer DM-97 DMM measure the voltage of the pack. Record the result with an error. Using a Metex 3800 DMM, measure the voltage of the pack. Record the result with an error. Are your results consistent?

4. Calculation: $R = V/I$. Build the circuit shown (extra wires required).

   Measure the voltage across your resistor with a Sinometer DM-97 DMM (record the result with an error). Simultaneously measure the current through the resistor with a Metex 3800 DMM (record the result with an error). Calculate the resistance of the resistor along with its error. Is your calculated resistance consistent with those measured in #2? (If it isn't consult your lab instructor.) I bet the voltage measured here is less than that in #3 above. This is a consequence of the internal resistance of the battery pack that I call *voltage droop* and also related to the *voltage burden* of the ammeter discussed in problem 6 and measured in #8 below. You are now done with the battery pack.
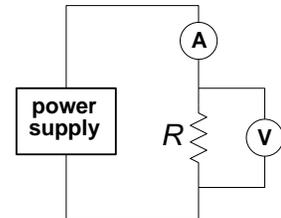
5. Find a LAMBDA power supply. Record its model number. Turn all three of its knobs to mid-range values; leave the terminal plugs unconnected. Notice that the terminal plugs are labeled: + (red), ⏚ (white), − (black). Which terminals should you use to have the power supply function like the battery pack? What is the purpose of the third terminal?

6. Plug-in/turn on the power supply. Switch the METER switch to VOLTAGE and adjust the VOLTAGE ADJ knobs (both COURSE and FINE) until the meter reads approximately the same as #3 above. Now attach a DMM to the power supply. Using the FINE VOLTAGE ADJ knob try again to match the output of the battery pack. (It need not be perfect.)

7. Reform the circuit of #4 now using the power supply adjusted to match the battery pack. Measure of the voltage across and current through your powered resistor and compare to those obtained in #4 above. Why the difference?

8. Voltage Burden of Ammeter. Move the voltmeter in the above circuit so as to measure the voltage drop across the ammeter (rather than your resistor). Report the result.

9. Disconnect the DMMs, and make a circuit where the power supply is directly attached to the resistor. Holding the resistor between your fingers carefully and slowly raise the voltage produced by the power supply. At some point (typically > 20 V) the resistor should start to warm up. Using the power supply meter, record the voltage that produced a noticeable heating effect. Calculate the power (watts) required for this heating effect. (No need to calculate error.)

10. Return the power supply to the approximate voltage level produced by the battery pack and disconnect your resistor. Switch the METER switch to CURRENT. Use a banana plug wire to short circuit the output of the power supply. (Note that short circuiting most devices will result in damage. These lab power supplies are [I hope] an exception to this rule.) Notice that the CURRENT ADJ knob now allows you to set the current through the wire (and of course the voltage across the wire is nearly zero). As long as the output voltage is less than the set voltage limit, the CURRENT ADJ knob is in control. If (as before) the current is below the set current limit, the VOLTAGE ADJ knob is in control. Thus if the voltage limit is set high, you have an adjustable current source; if the current limit is set high, you have an adjustable voltage source.

11. Voltage divider. "Pot" = potentiometer.

    Set up the potentiometer circuit shown and power it with the LAMBDA power supply. Measure V_{out} with a DMM. This is a classic voltage divider. Notice that by adjusting the pot, any fraction of the supplied voltage can be produced.
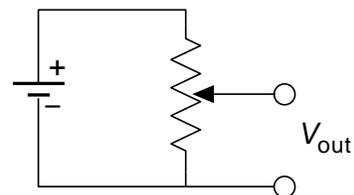
0. **Basic AC measurements.** For this and following sections you need not report errors in measurements, but always record every digit displayed on the DMM or scope and use devices on proper scales/ranges.

1. WAVETEK 180 as ac power source

   Set up the circuit shown using your resistor and your WAVETEK 180 (as the ac voltage source). Start with all the WAVETEK knobs fully counter-clockwise. (This is basically everything off and sine wave function selected.) Using the big knob and the FREQ MULT knob, set the frequency to 1 kHz. This will also turn on the WAVETEK.
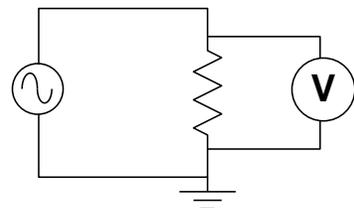
2. Adjusting the AMPLITUDE knob will change the output voltage. What range of voltages can be obtained from the HI BNC output? How about the LO BNC output?

3. Set an output voltage of approximately 1 V. Increase the frequency to find the highest frequency for which the reported voltage remains in the range 0.95–1.05 V (i.e., within 5% of the set value). Report the voltmeter you're using, its reported voltage, and the frequency. Further increase the frequency until the voltmeter reports a voltage under 0.5 V. (Record that your voltage and frequency.) The apparent change in voltage you are seeing is really mostly due to the failure of the voltmeter to work at high

frequency. As you'll see below the WAVETEK's output really is fairly constant as the frequency is varied.

4. With the output voltage still set at approximately 1 V, set the frequency to 100 Hz, and measure the ac current through and the ac voltage across your resistor. (The required circuit is analogous to #4 in DC measurements. You might want to have your instructor check your circuit before powering up.) Calculate the resistance (no error calculation required). Your result should be consistent with previous resistance measurements. Measure the current again at a frequency of 1 kHz. Does the current (and hence resistance) vary much with frequency?

5. Select a capacitor from the "370 Capacitor" cup. (All of these capacitors are identical.) Replace your resistor with that capacitor. With the 100 Hz, 1 V output, measure the current thorugh and voltage across the capacitor using the same circuit as you used for the resistor. Calculate the "resistance" (actually reactance, i.e., V/I). Using your calculated reactance, calculate the capacitance (see Fig. 1.1 on page 22 if you're unsure of the definition of C). Measure again at a frequency of 1 kHz. Does the current (and hence reactance) vary much with frequency? Again calculate the capacitance—it should be nearly the same even though the current should be nearly ten times bigger.

0. **Basic scope measurements. Note:** A "scope trace sketch" should include: horizontal and vertical scale settings (and the size of a DIV on your sketch—this is one place where quad ruled notebook paper helps!), the location of ground (zero volts), and the signal frequency. All of these numbers are on the scope display. In addition report the type of Coupling being used on the displayed channel. Feel free to report "same settings as previous" if that is the case.

1. Set up the scope and a DMM to monitor the function generator as shown. Set the function generator to produce a ~1 V sine wave at a frequency of ~1 kHz. Turn on the scope and hit DEFAULT SETUP. You must now change both CH1 and CH2 to 1× probe from the default value of 10× probe. Fiddle (or not) with the controls until you have a nice stable display of the signal on the scope. Sketch the scope trace.
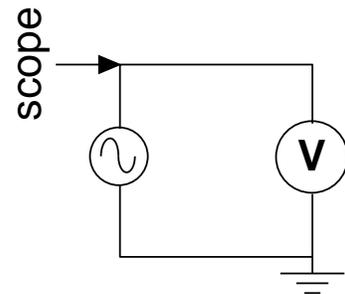
2. Record the size (in units of divisions on the display) for: the peak-to-peak voltage (V_{pp}), the amplitude or peak voltage (V_p), and the period (T) of the wave. Convert these to physical units (volts, seconds) using the scale factors. Calculate the frequency from the period and compare to the set value. Calculate the rms voltage from the peak voltage and compare to the DMM value. Hit the MEASURE menu and arrange the simultaneous display of Freq, Period, Pk-Pk voltage, Cyc RMS voltage, and Max voltage. Compare these values to those you calculated from divisions.

3. Run the function generator frequency up to and beyond the limiting value determined in previous section, #3. Notice that the scope (correctly) shows a constant amplitude even as the voltmeter (incorrectly) shows a changed voltage at high frequency.

4. Return the frequency to 1 kHz. Vary the function produced by the function generator: try triangle and square waves. For each wave, record the scope-reported peak-to-peak and rms voltages. Also record the rms voltage reported by the DMM. For a square

wave: $V_{rms} = V_{pp}/2$; for a triangle wave: $V_{rms} = V_{pp}/2\sqrt{3}$. Compare these calculated rms voltages to those directly reported by the scope and DMM. Which device appears to be a "true rms" meter? Explain.

5. In the MEASURE menu change the measurement of Period to Mean (voltage). Find the DC OFFSET knob on the function generator. Monitor the scope display and produce a ~1 V amplitude sine wave with a ~1 V dc offset. Sketch the resulting scope trace. Record the DMM ac voltmeter reading; record the scope Cyc RMS voltage. Switch the DMM to dc volts and record the dc voltage the reading; record the scope Mean voltage.

Switch the scope's input to Coupling▶AC, and repeat the above sketches and measurements.

Questions:

(a) Does a dc offset affect the reading of the DMM on dc volts?

(b) What is feature of the signal is recorded by the DMM on dc volts?

(c) Does a dc offset affect the reading of the scope Mean voltage:
When Coupling▶AC? When Coupling▶DC?

(d) Under what coupling will a DMM's dc volts and a scope's Mean voltage agree (approximately)?

(e) Does a dc offset affect the reading of the DMM on ac volts?

(f) Does a dc offset affect the reading of the scope Cyc RMS voltage:
When Coupling▶AC? When Coupling▶DC?

(g) Under what coupling will a DMM's ac volts and a scope's Cyc RMS voltage agree (approximately)?

0. **Filters.**

In many circumstances in electronics we are concerned with the ratio of voltages (or powers, currents, etc.), e.g., $\mathcal{A} \equiv V_{\text{out}}/V_{\text{in}}$. For example the aim of an amplifier is to produce a $V_{\text{out}} \gg V_{\text{in}}$ and one is usually primarily concerned with the amplification (or *gain*), which is that ratio. ($\mathcal{A} > 1$ for amplification.) On the other hand, sunglasses aim to reduce or attenuate light before in enters the eye. ($\mathcal{A} < 1$ for attenuation.) Such ratios are most often reported in *decibels* (dB):

$$20 \ \log_{10} \left(V_{\text{out}}/V_{\text{in}} \right) \tag{2.1}$$

Thus if an amplifier increases the voltage by a factor of 10 (i.e., $\mathcal{A} = 10$), we would say the amplifier has a gain of 20 dB. Some common examples:

| amplification		attenuation	
$\mathcal{A}$	dB	$\mathcal{A}$	dB
100	40	$\frac{1}{100}$	-40
10	20	$\frac{1}{10}$	-20
2	6	$\frac{1}{2}$	-6
$\sqrt{2}$	3	$\frac{1}{\sqrt{2}}$	-3
1	0	1	0

An electronic filter aims to pass certain frequencies and attenuate others. For example, a radio antenna naturally picks up every frequency; an electronic filter must attenuate all but the desired frequency. We start here with a *low pass filter*: a filter that lets low frequencies pass nearly untouched, but attenuates high frequencies. It is impossible to have step changes in attenuation, so as the frequency is increased, $\mathcal{A}$ goes smoothly to zero. Traditionally the curve of $\mathcal{A}$ vs. f is presented as a log-log plot and is called a Bode[1] plot. Filters are usually characterized by their "-3 dB" frequency ($f_{-3\text{dB}}$), i.e., the frequency[2] at which $\mathcal{A} = 1/\sqrt{2} \approx .707$.

For the following measurements you will use a sinusoidal signal with no dc offset and a very large range of frequencies and voltages: Remember to adjust your scope scales appropriately. Your voltage MEASUREments are most easily done as Cyc RMS (as a ratio is involved you could use Pk-Pk—theoretically as long as the voltages are in proportion the ratio will be invariant) with Coupling▶AC.

1. Solder together a lead from your resistor and a lead from your capacitor.

2. *RC* low pass filter.

   Use your *RC* combination to construct the low pass filter shown. You will monitor the input to the filter (V_{in}) on CH 2 of the scope and the output (V_{out}) on CH 1. Describe how this filter's 'gain' (but here $\mathcal{A} < 1$, so 'attenuation' might be a better word) changes as the frequency is varied from 50 Hz to 1 MHz. Find the filter's $f_{-3\text{dB}}$ (at which $V_{\text{out}} = V_{\text{in}}/\sqrt{2}$). Make a table of attenuation ($V_{\text{out}}/V_{\text{in}}$), dB, and f measured at frequencies of about 0.1, 0.2, 1, 2, 10, 20, and 100 times $f_{-3\text{dB}}$. (Note that if, by adjusting the AMPLI-TUDE on the Wavetek you make $V_{\text{in}} = 1$, you won't need a calculator to find $V_{\text{out}}/V_{\text{in}}$.) By hand, make a log-log graph of attenuation vs. f using the supplied paper. Check that the filter's Bode-plot slope for $f > f_{-3\text{dB}}$ is 6 dB per octave (or, equivalently, 20 dB per decade). Sketch (together) the scope traces at $f_{-3\text{dB}}$. Does V_{out} lead or lag V_{in}? (Note: a *RC* low pass filter is behind your scope's HF Reject option; a *RC* high pass filter is behind your scope's Coupling▶AC option.)
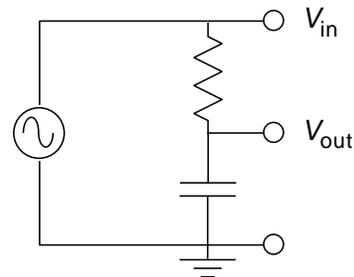
3. *LC* Filter

   I've constructed the beastie shown. (For the curious, it's a 5-pole, low-pass But-terworth filter in the π-configuration: see Appendix H of Horowitz and Hill if you want more details.) At low frequencies the gain never exceeds about $\frac{1}{2}$; therefore $f_{-3\text{dB}}$ is defined as the frequency at which $\mathcal{A}$ is .707× its low frequency value, i.e.,

---

[1] "Bo–Dee" plots were popularized by Hendrik Wade Bode's book *Network Analysis and Feedback Am-plifier Design* 1945. Bode worked at Bell Labs and became head of the lab's Mathematics Department in 1944.

[2] **Note:** The general definition of $f_{-3\text{dB}}$ is <u>not</u> that the gain is .707, rather that the gain has <u>changed</u> from some standard value by a factor of .707, i.e., the gain is 3 dB less than 'usual'. In #2 the 'usual' $\mathcal{A}$ is 1, so $f_{-3\text{dB}}$ is where $\mathcal{A} = .707$ In #3 the 'usual' $\mathcal{A}$ is .5, so $f_{-3\text{dB}}$ is where $\mathcal{A} = .354$

$\mathcal{A} = .707 \times .5 = .354$. Using a sine wave as input, take data of attenuation vs. frequency and plot them on log-log paper. Use frequencies of about 0.01, 0.1, 0.2, 0.5, 1, 2, 4, and 8 times $f_{-3\text{dB}}$. (At high frequencies the attenuation is so large that V_{out} may be lost in the noise unless you increase V_{in} to well above 1 V.) How does the steepness of this filter's cut-off compare with that of the simple RC filter? (I.e., compare the dB-per-octave falloff of this filter to the simple RC filter.)
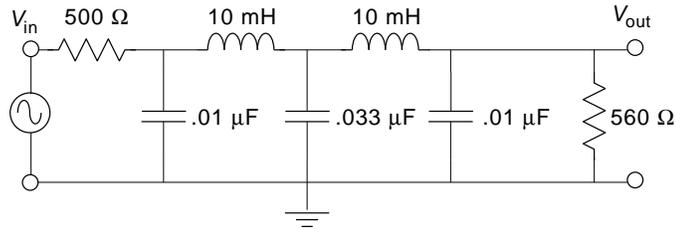
4. Tape your resistor+capacitor combination in your notebook.

3: Bubble Chamber

Well, by 4 A.M. Saturday we had made a fundamental discovery that would change the way cyclotrons were used... because of the properties of the magnetic field near and beyond the rim of the cyclotron magnet, [the pion beam] curved around machine in a tight beam... The only catch was the [ten-foot-thick] wall of concrete between the machine and the experimental area... In a few days my cloud chamber registered more pions than all the other labs in the world put together. Each photograph (we took one each minute) had six to ten beautiful tracks of pions. Every three or four photographs would show a kink in a pion track as it disintegrated into a muon and "something else."

Leon Lederman (1988 Nobel Laureate) *The God Particle* (1993) p. 223

Purpose

The purpose of this experiment is to determine the rest mass of the pion (m_π) and the rest mass of the muon (m_μ).

Introduction

Particle physics (a.k.a. high energy physics) is the division of physics which investigates the behavior of particles involved in "high" energy collisions. ("High" here means energies greater than those found in nuclear reactions, i.e., more than 100 MeV=0.1 GeV. The highest energy particle accelerators available today produce collisions with energies of a few million MeV = TeV.)

The first "new" particles discovered (circa 1940) by particle physicists were the pion (π) and the muon (μ). In spite of roughly similar masses (near 100 MeV, compare: electron mass = .511 MeV and proton mass = 938 MeV), these two particles have quite different properties.

The muon is a relative of the electron (and hence is called a lepton). It comes in particle (μ^-) and anti-particle (μ^+) versions and has spin $\frac{1}{2}$. Unlike the electron, the muon is unstable. It decays into two neutrinos (ν) and an electron (or positron) after a mean life of

2×10^{-6} s:

$$\mu^+ \longrightarrow \bar{\nu} + \nu + e^+ \tag{3.1}$$

$$\mu^- \longrightarrow \nu + \bar{\nu} + e^- \tag{3.2}$$

The pion belongs to the class of particles called mesons. Unlike leptons, mesons interact with protons and neutrons through an additional force called the strong nuclear force (a.k.a., color force). (Particles that can feel this force are called hadrons.) Unlike leptons, mesons are known to be composite particles: each is made of a quark and an antiquark. The pion comes in three versions: π^+, π^0, and π^- and has spin 0. All the pions are unstable; the π^+ decays after a mean life of 3×10^{-8} s:

$$\pi^+ \longrightarrow \mu^+ + \nu. \tag{3.3}$$

(The π^0 has a slightly smaller mass and decays much faster than the $\pi^\pm$. It is not seen in this experiment.)

Particle Detection

Since the particles studied by particle physics are sub microscopic and decay "quickly", particle detection is a problem. Most existing particle detectors rely on the fact that as a charged particle moves by an electron (e.g., an electron in an atom of the material through which the charged particle is moving), the electron feels a net impulse. If the charged particle comes close enough to the electron and/or is moving slowly enough (so the interaction is long enough), the impulse on the electron will be sufficient to eject the electron from its atom, producing a free electron and an ion. Thus a charged particle moving through material leaves a trail of ions. This trail can be detected in many ways (e.g., by direct electronic means as in a modern wire chamber or chemically as when the material is a photographic plate or emulsion). In this experiment the ion trail is made visible by vapor bubbles which are seeded by individual ions in boiling material (here liquid hydrogen). The bubbles are large enough to be photographed whereas the ion trail itself is much too narrow.

Relativistic Kinematics

Recall the following from Modern Physics:

$$E = \gamma m c^2 \tag{3.4}$$

$$T \equiv E - m c^2 \tag{3.5}$$

$$pc = \gamma m v c = \gamma m c^2 \beta = E\beta \tag{3.6}$$

$$E^2 - (pc)^2 = \left[\gamma m c^2\right]^2 \left(1 - \beta^2\right) = \left[m c^2\right]^2 \tag{3.7}$$

where:

$$\beta = v/c \tag{3.8}$$

$$\gamma = \frac{1}{\sqrt{1 - \beta^2}} \tag{3.9}$$

and v is the velocity of the particle with rest mass m, momentum p, total energy E and kinetic energy T. Note that E, T, pc, and mc^2 all have the dimensions of energy; it is customary to express each in MeV and even say "the momentum of the particle is 5 MeV" or "the mass of the particle is 938 MeV." (Of course, technically the momentum of the particle would be 5 MeV/c and the mass 938 MeV/c^2. Basically what we are doing is redefining "momentum" to be pc and "mass" to be mc^2. Since the "c" has disappeared, this re-naming is sometimes called "setting $c = 1$".)

For future reference, note from Equation 3.6 that if $\beta \to 1$, $E \approx pc$ and from Equation 3.7 that if $m = 0$, $E = pc$. Of course, massless particles (like light) must travel at the speed of light (i.e., $\beta = 1$).

Momentum Measurements

Classically a charged particle (with mass m and charge q) moving through a magnetic field $\vec{\mathbf{B}}$ has an acceleration, $\vec{\mathbf{a}}$, given by

$$m\vec{\mathbf{a}} = q\vec{\mathbf{v}} \times \vec{\mathbf{B}} \tag{3.10}$$

Because of the cross product, the acceleration is perpendicular to both $\vec{\mathbf{v}}$ and $\vec{\mathbf{B}}$. Thus there is zero acceleration in the direction of $\vec{\mathbf{B}}$, so $v_\parallel$, the component of velocity parallel to $\vec{\mathbf{B}}$, is constant. On the other hand in the plane perpendicular to $\vec{\mathbf{B}}$, the acceleration and the velocity are perpendicular resulting in centripetal (circular) motion. Thus the particle moves in a circle of radius R even as it travels at constant speed in the direction of $\vec{\mathbf{B}}$. The resulting motion is a helix (corkscrew). Using $\perp$ to denote components in the plane perpendicular to $\vec{\mathbf{B}}$, we have:

$$ma_\perp = \frac{mv_\perp^2}{R} \quad = \quad qv_\perp B \tag{3.11}$$

$$p_\perp = mv_\perp \quad = \quad qBR \tag{3.12}$$

$$p_\perp c \quad = \quad qcBR \tag{3.13}$$

This last relationship continues to hold for relativistic particles.

SHOW: For a positron, Equation 3.13 means the momentum $p_\perp c$ (in MeV) can be calculated as a simple product $3BR$:

$$p_\perp c \quad \text{(in MeV)} = 3BR \tag{3.14}$$

where B is in Tesla and R is in cm.[1]

In this experiment, positrons (and electrons) from muon decay circle in an applied magnetic field. You will measure the radii of the positron orbits to determine positron $p_\perp$. Since the rest mass of the muon has been converted to kinetic energy of its decay products, positron $p_\perp$ depends on muon mass and measurement of $p_\perp$ allows calculation of m_μ.

---

[1]This is an example of a "calculator equation" where we seemingly ignore units. That is if $B = 2$ T and $R = 5$ cm, this equation says $p_\perp c = 3 \times 2 \times 5 = 30$ MeV, units seemingly just tacked onto the answer. To 'derive' such an equation, you must demonstrate (once!) how the units work out. In particular, $p_\perp c$ — which in MKS units in going to naturally come out in Joules — must be converted to the energy unit MeV. You can start your derivation by assuming $B = 1$ T, $R = 1$ cm and calculate the resulting $p_\perp c$ in Joules and then convert that to MeV. The conversion factor is 1 MeV=1.6022×10^{-13} J. Of course, you already know 100 cm=1 m.

Kinetic Energy Measurement

As stated above, a charged particle moving through a material leaves a trail of ions. The energy needed to form these ions must come from the kinetic energy of the charged particle. Thus, every cm of travel results in a kinetic energy loss. It can be shown (Bethe-Block) that the decrease in kinetic energy depends on the inverse of the velocity squared:

$$\frac{dT}{dx} = -\frac{2.1\rho}{\beta^2} \qquad \text{MeV/cm} \tag{3.15}$$

where ρ is the density of the material ($\rho = .07$ g/cm^3 for liquid H_2). This is the second example of a "calculator equation"[2].

A particle with some initial kinetic energy T_0 will travel some definite distance, L, before all of its kinetic energy is lost and it comes to rest. The relationship between T_0 and L can be determined from the energy loss per cm:

$$L = \int_{T_0}^{0} \frac{dx}{dT}\, dT = \int_{0}^{T_0} \frac{\beta^2}{2.1\rho}\, dT \tag{3.16}$$

For particles moving much slower than the speed of light, Newton's mechanics is a good approximation: $T = \frac{1}{2}\, mv^2 = \frac{1}{2}\, mc^2\beta^2$

$$L = \frac{2}{2.1mc^2\rho} \int_{0}^{T_0} T\, dT = \frac{T_0^2}{2.1mc^2\rho} \tag{3.17}$$

so $T_0 \propto L^{1/2}$.

In this experiment muons produced by pion decay travel a distance L before coming to rest. You will measure the muon path length to determine muon T_0. Since the kinetic energy of the muon comes from the rest mass of the decaying pion, the mass of the pion can be calculated from muon T_0.

Perspective Effects

In real particle physics experiments, decay events are reconstructed in three dimensions. However in this experiment you will measure *apparent* muon path lengths from photographs. Because of perspective effects, typically the true path length (L) is longer that the apparent (photographic) path length ($L_\perp$), as the particle will generally be moving towards or away from the camera in addition to sideways. In this experiment we need to "undo" the perspective effect and determine L from the measurements of $L_\perp$.

There are several ways this could be done. Perhaps the easiest would be to pick out the longest $L_\perp$, and argue that it is longest only because it is the most perpendicular, i.e.,

$$\max\left(\{L_\perp\}\right) \approx L \tag{3.18}$$

Essentially this is a bad idea because it makes use of only one collected data point (the maximum $L_\perp$). For example, it is likely you will make at least one misidentification or

---

[2]Thus if $\rho = .1$ g/cm^3 and $\beta = .5$ we would conclude that dT/dx was $-.84$ MeV/cm.
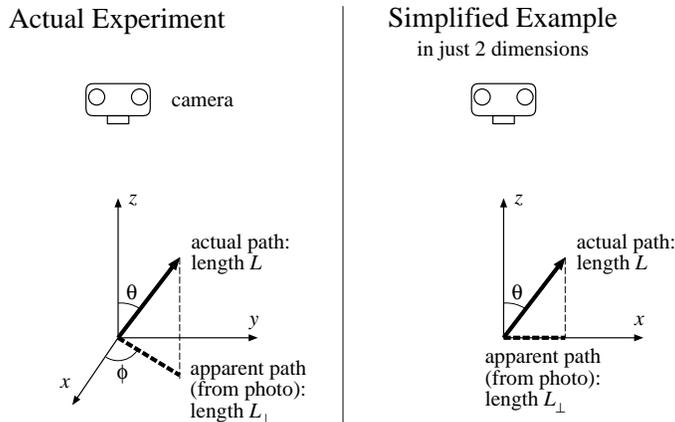
Figure 3.1: This experiment uses photographs of particle paths in a bubble chamber. Two angles ($\theta \in [0, 180°], \phi \in [0, 360°]$) are required to describe the orientation of the path in three dimensional space. The photographic (apparent) path length, $L_\perp$, is shorter than the actual path length, L, (of course, if $\theta = 90°$, $L_\perp = L$). In general: $L_\perp = L \sin\theta$. The angle ϕ just describes the orientation of the apparent path in the photograph. We can make an easier-to-understand model of perspective effects by just dropping ϕ and considering a two dimensional experiment. In this case to generate all possible orientations $\theta \in [0, 360°]$.
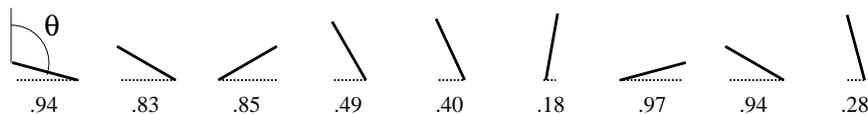
Figure 3.2: Nine randomly-oriented, fixed-length segments are placed on a plane and the corresponding horizontal lengths $L_\perp$ (dotted lines) are measured (results displayed below the segment). The resulting data set $\{.94, .83, .85, .49, .40, .18, .97, .94, .28\}$ of $L_\perp$ can be analyzed to yield the full segment length L. (The angle $\theta \in [0, 360°]$ describes the orientation, but it is not measured in this "experiment": only $L_\perp$ is measured.)

mismeasurement in your 60+ measurements. If the longest $L_\perp$ happens to be a bad point, the whole experiment is wrong. Additionally since L is the net result of interactions with randomly placed electrons, L is not actually exactly constant. (That is, Equation 3.15 is true only "on average".) Paths that happen to avoid electrons are a bit longer. The L-T_0 relationship is based on average slowdown; it should not be applied to one special path length.

One way of using all the data is to note that randomly oriented, fixed-length paths will produce a definite average $L_\perp$ related to L. So by measuring the average $L_\perp$ (which we will denote with angle brackets: $\langle L_\perp \rangle$), you can calculate the actual L.

It will be easier to explain this method if we drop a dimension and start by considering randomly oriented, fixed-length segments in two dimensions. Figure 3.2 shows[3] nine randomly oriented segments in a plane with the corresponding measured $L_\perp$. The different measured

---

[3]Note that if we applied Equation 3.18, we would conclude $L = .97$ with no estimate for the uncertainty in this result (i.e., δL).

$L_\perp$ are a result of differing orientations of a fixed-length segment:

$$L_\perp = L|\sin\theta| \tag{3.19}$$

From a sample of N measurements of the horizontal distance $L_\perp$ (i.e., a data set of measured $L_\perp$: $\{x_i\}$ for $i = 1, 2, \ldots, N$, with corresponding orientations $\{\theta_i\}$ with $\theta_i \in [0, 2\pi]$), the average $L_\perp$ could be calculated

$$\langle L_\perp \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{L}{N}\sum_{i=1}^{N}|\sin\theta_i| \tag{3.20}$$

The θ_i should be approximately evenly distributed with an average separation of $\Delta\theta = 2\pi/N$ (because there are N angles distributed throughout $[0, 2\pi]$). Thus, using a Riemann sum approximation for an integral:

$$\langle L_\perp \rangle = \frac{L}{N}\sum_{i=1}^{N}|\sin\theta_i| = L\sum_{i=1}^{N}|\sin\theta_i|\left(\frac{\Delta\theta}{2\pi}\right) \tag{3.21}$$

$$\approx \frac{L}{2\pi}\int_0^{2\pi}|\sin\theta|\,d\theta = L\frac{\int_0^{2\pi}|\sin\theta|\,d\theta}{\int_0^{2\pi}d\theta} \tag{3.22}$$

The above integral is easily evaluated:

$$\int_0^{2\pi}|\sin\theta|\,d\theta = 2\int_0^{\pi}\sin\theta\,d\theta = 2\Big[-\cos\theta\Big]_0^{\pi} = 4 \tag{3.23}$$

Thus we have the desired relationship between $\langle L_\perp \rangle$ and L:

$$\langle L_\perp \rangle = L\frac{2}{\pi} \tag{3.24}$$

With the example data set we have: $\langle L_\perp \rangle = 0.653$ with standard deviation $\sigma_{L_\perp} = 0.314$. Using the standard deviation of the mean we have:

$$0.65 \pm \frac{.314}{\sqrt{10}} = 0.65 \pm .10 \;=\; L\frac{2}{\pi} \tag{3.25}$$

$$1.03 \pm .16 \;=\; L \tag{3.26}$$

Note that our argument for finding averages is quite general, so if random values of x are uniformly selected from the interval $[a, b]$, the average value of any function of x, $f(x)$, can be calculated from:

$$\langle f(x)\rangle = \frac{\int_a^b f(x)\,dx}{\int_a^b dx} \tag{3.27}$$

For the actual experiment, the path orientations have a uniform distribution in space. That is, if all the paths originated from the same point, the path ends would uniformly populate the surface of a sphere of radius L. The element of surface area of a sphere of radius L is:

$$L^2\,d\Omega = L^2\,\sin\theta\,d\theta\,d\phi \tag{3.28}$$

where Ω is called the solid angle and plays an analogous role to radian measure in a plane:

$$\text{plane angle in radians} \;=\; \frac{\text{arc length}}{R} \tag{3.29}$$

$$\text{solid angle in steradians} \;=\; \frac{\text{sphere surface area}}{R^2} \tag{3.30}$$

Thus the relationship between $\langle L_\perp \rangle$ and L in three dimensions is:

$$\langle L_\perp \rangle = L \, \langle \sin \theta \rangle = L \, \frac{\int \sin \theta \, d\Omega}{\int d\Omega} = L \, \frac{\pi}{4} \tag{3.31}$$

SHOW this result! Note: $d\Omega = \sin \theta \, d\theta \, d\phi$ and the range of the double integral is $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$

Comment: The above discussion has been phrased in terms of position vectors, but it applies as well to any vector. In particular, you will be measuring the perpendicular component of momentum, $p_\perp$, and need to deduce the actual momentum, p. Exactly as above, if the particles have the same speed with direction uniformly distributed in space:

$$\langle p_\perp \rangle = p \, \frac{\pi}{4} \tag{3.32}$$

If the particles actually have differing speeds we can still conclude:

$$\langle p_\perp \rangle = \langle p \rangle \, \frac{\pi}{4} \tag{3.33}$$

Displaying Distributions

As discussed above, when finding L it is best to use the entire data set. Although the $\langle L_\perp \rangle$ method uses all the data, it quickly reduces the whole data set to one number. Is there some way of graphically displaying and using all the data? In particular, is there some way of checking to see if the data have the expected distribution (i.e., the right proportion of long and short $L_\perp$s)?

Perhaps the easiest way to understand the idea of a distribution is to consider the idea of the cumulative fraction function for some data set: $\{x_i\}$, for $i = 1, 2, \ldots, N$. The cumulative fraction function[4], $c(x)$, reports the fraction of the data set $\{x_i\}$ that is less than or equal to x. Obviously if $a < \min(\{x_i\})$, $c(a) = 0$; if $b \geq \max(\{x_i\})$, $c(b) = 1$; and if $c(x) = .5$ then x is the median (i.e., middle) data point. In the 2d example data set, $c(.60) = 4/9$, because four of the nine data points are smaller than .6. Similarly $c(.84) = 5/9$. Every time x increases past one of the x_i, $c(x)$ has a jump. See Figure 3.3 for a plot of this function.

The function $c(x)$ depends on the data set, so if the experiment is repeated generating a new data set $\{x_i\}$, a new function $c(x)$ is also generated. The new $c(x)$ should be slightly different, but generally similar to the old $c(x)$. If the data set is sufficiently large, the new and old $c(x)$ will be quite similar and both $c(x)$ would approximate the function $\hat{c}(x)$, the cumulative fraction function that would be generated from an infinite-sized data set[5]. How can $\hat{c}(x)$ be best approximated from one finite-sized data set $\{x_i\}$? To answer this question it will be convenient to consider the data set $\{x_i\}$ already sorted so x_1 is the minimum and x_N is the maximum. Thus our example data set:

$$\{.94, .83, .85, .49, .40, .18, .97, .94, .28\} \tag{3.34}$$

---

[4]The cumulative fraction function is also known as the empirical distribution function, and is closely related to percentiles and order statistics.

[5]The usual name for $\hat{c}(x)$ is the distribution function.
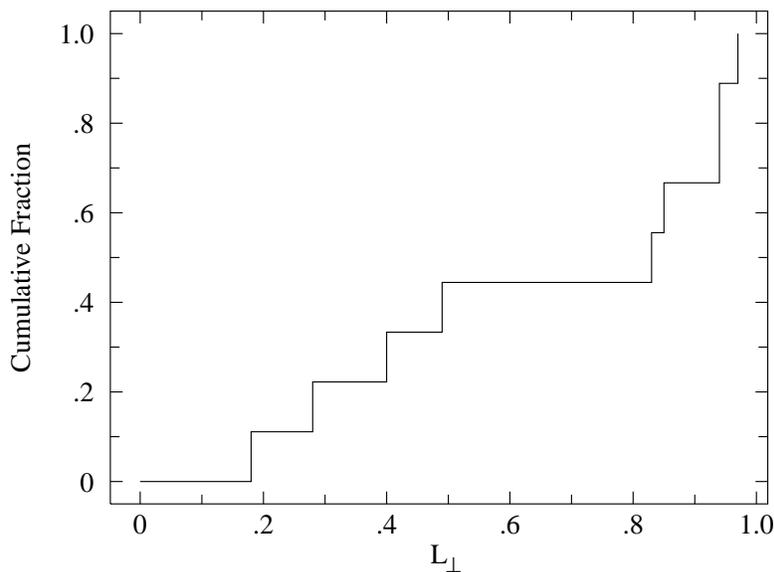
Figure 3.3: The distribution of data set: {.94, .83, .85, .49, .40, .18, .97, .94, .28} of Figure 3.2 (nine randomly oriented segments) displayed as a cumulative fraction.

becomes:

$$\{.18, .28, .40, .49, .83, .85, .94, .94, .97\} \tag{3.35}$$

As defined above, $c(x)$ is given by:

$$c(x) = \frac{i}{N} \qquad \text{where } i \text{ is such that: } x_i \le x < x_{i+1} \tag{3.36}$$

That is to determine $c(x)$ for some x, we see how far down the sorted list we must travel to find the spot where x fits between two adjacent data points: $x_i \le x < x_{i+1}$. Clearly there are a total of i data points less than or equal to x (out of a total of N), so $c(x) = i/N$. If x happens to equal one of the data points, things are a bit undefined because $c(x)$ has a jump discontinuity at each x_i. It turns out that the best estimate for $\hat{c}$ at these discontinuities is:

$$\hat{c}(x_i) = \frac{i}{N+1} \equiv \alpha_i \tag{3.37}$$

Of course this estimate can be wrong; it has an uncertainty of

$$\sigma = \sqrt{\frac{\alpha_i(1 - \alpha_i)}{N+2}} \tag{3.38}$$

See Figure 3.4 for a comparison of the estimated $\hat{c}(x_i)$ (called the percentile) and the cumulative fraction function. (The mathematics of these results is covered under the topic "order statistics" or "nonparametric methods" in advanced statistics books.)

Your experimental estimate of $\hat{c}$ should be compared to the theoretically expected $\hat{c}$. The example data set was generated from randomly oriented line segments in a plane. As shown
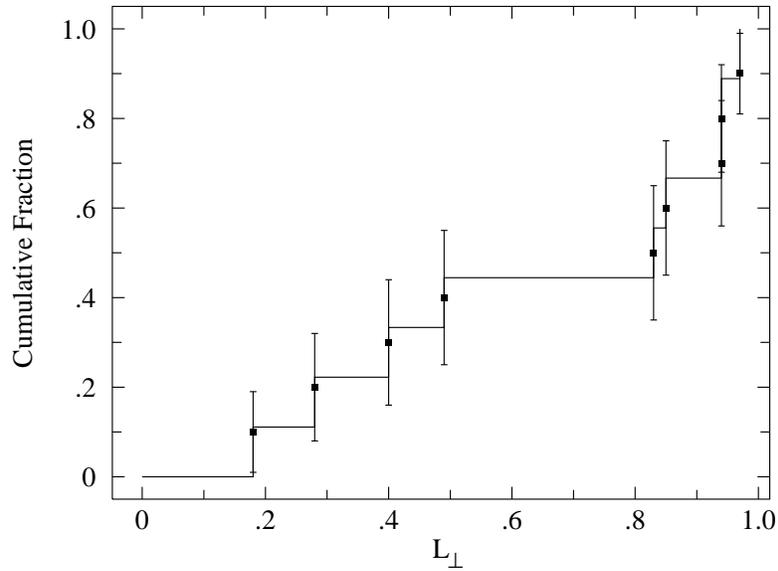
Figure 3.4: The cumulative fraction function for the example data set is plotted along with the data points for the percentile estimate of the distribution function $\hat{c}(x)$.
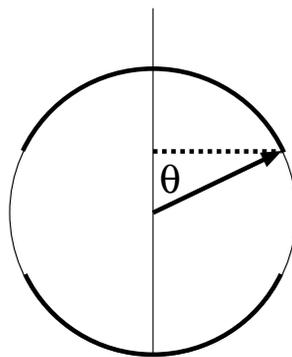
Figure 3.5: A particular line segment is displayed along with the measured $L_\perp$ (dotted line). What fraction of randomly oriented segments would have a $L_\perp$ smaller than this particular segment? The darkly shaded part of the circle shows possible locations for these small $L_\perp$ segments. The fraction of such small $L_\perp$ segments should be the same as the dark fraction of the circle: $4\theta/2\pi$.

in Figure 3.5, it is expected that the fraction of a data set less than some particular value
of $L_\perp$ is:

$$\hat{c}(L_\perp) \quad = \quad \frac{4\theta}{2\pi} \qquad \text{where: } \theta = \arcsin(L_\perp/L) \tag{3.39}$$

$$= \quad \frac{2}{\pi} \arcsin(L_\perp/L) \tag{3.40}$$

Warning: the above function is for the 2-d example, not your real 3-d data!

Our formula for $\hat{c}$ involves the unknown parameter L; we adjust this parameter to achieve
the best possible fit to our experimental estimate for $\hat{c}$. Using the program `fit`:

```
tkirkman@bardeen 7% fit
* set f(x)=2*asin(x/k1)/pi
* read file cf.L.dat
* set k1=1
* fit
Enter list of Ks to vary, e.g. K1-K3,K5 k1
 FIT finished with change in chi-square=  5.4810762E-02
  3 iterations used
 REDUCED chi-squared=  0.2289333     chi-squared=   1.831467
 K1=  0.9922597
 Display covariance/curvature matrices? No, Screen, File [N,S,F] s
 ...
 COVARIANCE MATRIX:
  0.65E-03

 CURVATURE MATRIX:
  0.15E+04
```

Using the covariance matrix to determine errors[6], we conclude `k1` $= 0.992 \pm .025$. This
reported random error is about $\frac{1}{6}$ that obtained above using $\langle L_\perp \rangle$.

SHOW: Derive yourself the theoretical function $\hat{c}(L_\perp)$ for line segments in space. Hint:
Begin by noting that if the segments shared a common origin, the segment ends would
uniformly populate the surface of a sphere of radius L. Segments with measured $L_\perp$ less
than some particular value would lie on a spherical cap, the three dimensional version of
arc caps displayed in Figure 3.5. The ratio of the area of these caps to the total surface
area of the sphere gives the expected value for $\hat{c}$. You will need to calculate the area of a
spherical cap by integration.

The above discussion has focused on path lengths as that is the quantity measured in pion
decay. In muon decay, the radius of positron orbits in the applied magnetic field is measured.
Weinberg-Salam theory provides a complete description of the decay process, including the
distribution of positron momentum (which in turn determines the radius of positron orbits

---

[6]Reference 2, Press et al., says usually error estimates should be based on the square root of the diagonal
elements of the covariance matrix
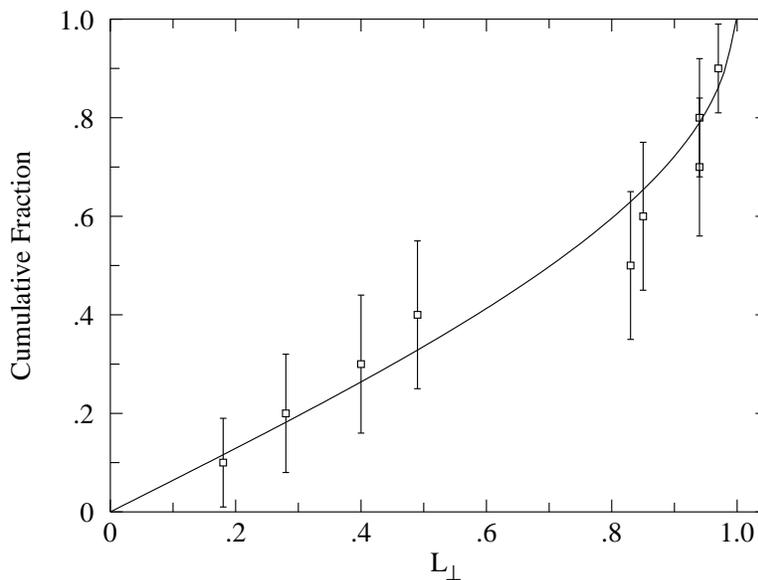
Figure 3.6: The theoretical distribution function (Equation 3.40) fit to the "experimental" data points derived (Equations 3.37 & 3.38) from the example data set. As a result of the `fit` we estimate: $L = 0.992 \pm .025$.

R). Kirkman has shown that the the Rs should be distributed according to

$$\hat{c}(R) = \left(\frac{3}{2}\, u^2 - 1\right) \sqrt{1 - u^2} + 1 - \frac{1}{2}\, u^4 \log\left(\left(1 + \sqrt{1 - u^2}\right)/u\right) \qquad (3.41)$$

where $u = R/R_m$, and R_m, the maximum value of R, is the value of R that corresponds to $p_\perp c = m_\mu c^2/2$. The adjustable parameter R_m (of course called `k1` in `fit`), can be adjusted to give the best possible fit to the experimental distribution. From R_m, m_μ can be determined.

Biases and Robust Estimation

The bane of every particle physics experiment is bias. Biases are data collection techniques that produce nonrepresentative data. For example, short $L_\perp$ are harder to notice than long $L_\perp$, and thus long $L_\perp$ tend to be over represented in the data sample, producing a high $\langle L_\perp \rangle$. Use of the cumulative fraction function allows this biases to be detected. In addition to biases, the cumulative fraction function allows you to detect likely mistakes: for example, particle path lengths that are extraordinary given the entire data set.

The detection of a likely mistake suggests corrective actions like removing the "bad" point. You should almost never do this! (You will find a chapter in Taylor on this "awkward" and "controversial" problem.) A better option is to use analysis methods that are "robust", i.e., that are insensitive to individual "bad" points.[7] Imagine we modify our example data set

---

[7] Removing a data point is a lie. A more subtle sort of lie comes from the existence of <u>choice</u> of methods.

by adding a "bad" point: $L_\perp = 2$:

$$\{.18, .28, .40, .49, .83, .85, .94, .94, .97, 2.00\} \tag{3.42}$$

Adding this outlying[8] data point totally messes up the $\max(\{L_\perp\})$ method (the least robust method). Since it increases both the mean and the standard deviation, the estimated L based on the $\langle L_\perp \rangle$ method shifts from $1.03 \pm .16$ to $1.24 \pm .26$.

Changing the number of data points requires recalculating the estimated distribution function for every point (because the value of $\hat{c}$ depends on the set size N). If we carry through the total analysis with our enlarged data set we find the `fit` L shifts from $0.992 \pm .025$ to $1.05 \pm .05$. We can conclude that the cumulative fraction method is less sensitive to bad data than the average method.[9]

Unconscious (uncontrolled) biases produce tainted data which can be rescued in part by robust estimation. Once bias is recognized the experiment can be rearranged to adjust for its effects. This requires that the bias be exactly reproducible. For example, short apparent muon paths (paths mostly towards or away from the camera) are inconspicuous and hence more likely to be missed on some occasions. One solution is formalize this bias and intentionally ignore all photographic paths shorter than say .3 cm long (about 5% of the data). This <u>cut</u> (formalized noncollection of data), can be included in the theoretical distribution function so it will not affect parameter estimation.

Experimental Arrangement

Our bubble chamber photographs were taken using the 385 MeV proton accelerator at Nevis Lab which is a part of Columbia University. A pion beam was produced by colliding accelerated protons with a copper target. The pion beam was directed through an absorber to slow the pions so that a sizable fraction of the pions would come to rest in the adjacent bubble chamber. See Figure 3.7. The path of charged particles from the pion decay ($\pi \to \mu \to e$) as recorded by a nearby camera, encodes the information needed to calculate m_π and m_μ. Apparent muon path length, $L_\perp$, will allow you to determine the actual muon path length L, from which in initial muon kinetic energy T_0 can be determined. A magnetic field ($B = .88$ T, directed toward the camera) bends the path of all charged particles into helices; but the effect is most visible with the low mass electrons. The radius of the electron helix determines (Equation 3.14) the electron's $p_\perp c$. From the distribution of electron $p_\perp c$, you can determine both $\langle pc \rangle$ and the maximum pc, from which m_μ can be determined.

Figure 3.8 shows an idealized decay sequence as might be recorded in a bubble chamber

---

Clearly, you can analyze the data several different ways, and then present only the method that produces the answer you want. Darrell Huff's book *How to Lie with Statistics* (Norton, 1954) can help you if that is your goal. I probably don't need to remind you that schools with "Saint" in their name do not recommend this course of action. Choice and ethics are interlocking concepts.

[8] While not exactly relevant, this data point is $2.34 \times \sigma$ above the mean and hence an outlier by Chauvenet's criterion (see Reference 4, Taylor). It is also an outlier by Tukey's criterion (see Reference 3, Hogg & Tanis).

[9] Do note that both results remain consistent with the intended value of 1.00. Also note that the median could have provided a more robust alternative to the average. However, that would have required a discussion of the uncertainty in the median, which is beyond the intended aims of this lab. In this lab—and in most any experiment—there are many possible ways to analyze the data. Choice of method often involves art and ethics.
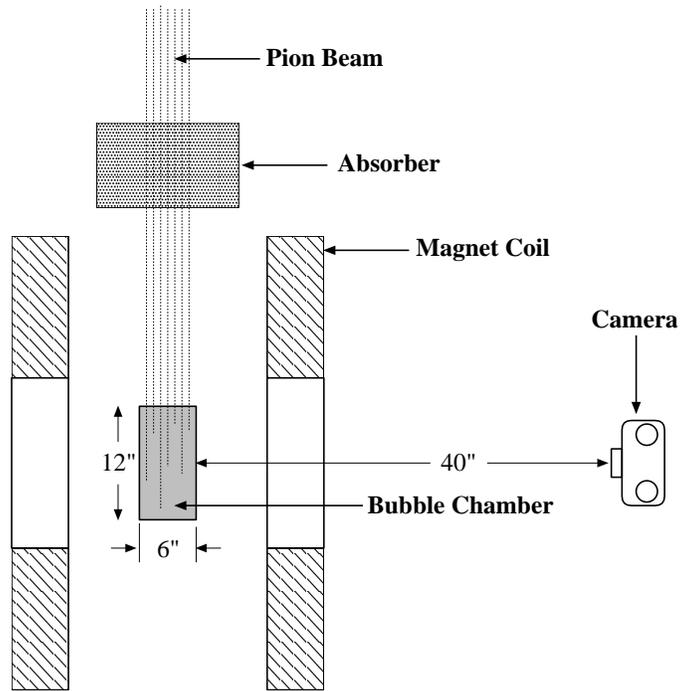
Figure 3.7: A pion bean is slowed by an absorber so the pions are likely come to rest inside a liquid hydrogen bubble chamber. The path of charged particles from the pion decay $(\pi \to \mu \to e)$ is recorded by a nearby camera. (Of course, the uncharged neutrinos from the decay leave no ion trail, and hence no bubbles grow to mark their path.) A magnetic field ($B = .88$ T, directed toward the camera) produced by the current in the coil, bends the path of all charged particles into helices, but the effect is most visible with the low mass electrons. The radius of the helix, as recorded by the camera, can be related to the particle's $p_\perp$. Since the muon decays into three particles, allowing varying distribution of energy, the electron's momentum can vary from 0 up to a maximum of $m_\mu c/2$. Since the pion decays into just two particles, there is only one way to distribute the released energy so the muon's initial kinetic energy is determined uniquely which produces a fixed stopping distance L. (Of course, the apparent muon path length, $L_\perp$, recorded by the camera will vary.)
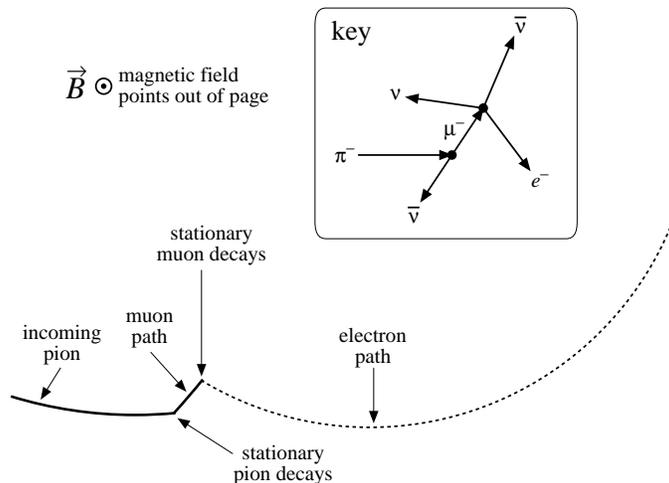
Figure 3.8: A typical decay process as recorded in a bubble chamber photograph. A pion (π^-) slows and comes to rest inside the bubble chamber. A short time later it decays into a muon and an antineutrino ($\pi^- \to \mu^- + \bar{\nu}$). A kink in the path marks the decay location. The muon is in turn slowed and comes to rest after traveling a short distance. A short time later the muon decays into an electron and two neutrinos ($\mu^- \to e^- + \bar{\nu} + \nu$). The high speed electron leaves a sparse track in accord with the Bethe-Block Equation (3.15) (reduced energy loss due to large β means fewer ions produced and hence fewer bubbles). You will be measuring apparent muon path lengths, $L_\perp$, and electron helix radii R.

photo. The paths of interest start on the left (pions from the accelerator), have a short ($\sim$1 cm) kink (muon), connecting to a sparse loop (electron).

Procedure

Measurement of Rs and $L_\perp$s is computerized. (Indeed almost all of your data collection this semester will be computerized.) As I'm sure you know, while computers can be useful devices, they seemingly have a knack for unintended/unexpected disasters, which are called 'user error'. Thus the most important lesson of computer use is: GIGO ('Garbage In; Garbage Out'[10]) — a computer's output must be considered unreliable until, at the very least, you know the limitations/uncertainties in the input data producing that output. In this lab you will be using a CalComp 2500 digitizing tablet to measure distances. The process seems simple (aligning a point between crosshairs and clicking to take the data-point) but involves problem of definition errors (including systematic biases, see page 13) in addition to more familiar device limitations (random and calibration errors). To have some justified confidence in this process, you must measure a known and see what the computer reports ('trust, but verify'). I have provided you with a simulated bubble chamber photo in which all the path lengths are 1 cm and all the curvatures are 20 cm. (If you don't trust this fiducial—and you might not since it depends on the dimensional stability of printers and paper—you can measure the 'tracks' with an instrument you do trust.) Begin by logging into your linux account using the Visual 603 terminal with attached CalComp 2500

---

[10]Sometimes this acronym is reported as 'Garbage In, Gospel Out' stressing many people's (mistaken) faith in computer output.

digitizer, and running the program `bubbleCAL`:

```
tkirkman@linphys8 1% bubbleCAL
```

The following directions are displayed:

```
The general procedure will be to place the cursor crosshair at
the needed place and press a cursor button to digitize.  Press
cursor button "0" when obtaining muon path lengths (digitize
beginning and end of track); press cursor button "1" when obtaining
electron radii of curvature (digitize three points on curve, from
which the computer can figure R); press cursor button "2" to cancel
an in-progress data point or clear error; press cursor button "3" to
remove the last data point of the presently selected type.  Additional
data points may be removed by number when done.  Digitizer will beep
between data points.  Hit ^D (control D) on keyboard when done.
Files containing your data (unsorted) will be created: Lcal.DAT & Rcal.DAT.
THIS PROGRAM IS FOR CALIBRATION/TESTING NOT DATA COLLECTION
```

With the simulated bubble chamber photo taped in place on the digitizer, check some long distances (> 10 cm) on the scales. Then measure 16 path lengths and 16 curvatures. *Record the data reported by the program (means, standard deviations, 95% confidence intervals, etc.).* Does the probable range for each mean include the known value? (If not discuss the problem with Dr. Kirkman.) Note that if you took more data points, the probable range for the mean would become increasingly small and eventually you would detect a systematic error limiting the ultimate accuracy of the device.

The program `bubble` works very much like `bubbleCAL`, except in the end it will produce files containing the cumulative fraction of $L_\perp$ (`cf.L.dat`) and R (`cf.R.dat`). Please do not deface the bubble chamber photos! Lightly tape each bubble chamber photograph to the digitizer to keep the photo from moving as you collect data. Collect ~ 64 $L_\perp$s and ~ 64 Rs. (This will require scanning about 20 bubble chamber photos.)

Common Mistakes: Collecting the radius of pion paths rather than electron paths. (If you've forgotten how to tell them apart, you should probably ask!) Collecting a biased sample of $L_\perp$: just the big obvious ones and skipping the ones ($\sim 5\%$!) shorter than .3 cm.

The program will make four files: `cf.R.dat` contains the usual three columns: R, the percentile estimate $\hat{c}(R)$ calculated from your data, and the error in the estimate, `cf.L.dat` similarly contains the sorted $L_\perp$, estimated $c(L_\perp)$ and error, `L.DAT` and `R.DAT` contain the (unsorted) raw data.

Use the web[11] or `gnumeric` (Linux spreadsheet) to calculate $\langle L_\perp \rangle$ and $\langle R \rangle$ and their standard deviations. Note that the raw data files (`L.DAT` and `R.DAT`) can be used to transfer the data to these applications.

Use `fit` (for a tutorial and introductory video see page 189; see page 177 for a command summary) to fit each dataset (`cf.R.dat` & `cf.L.dat`) to the appropriate theoretical distribution function. Produce a hardcopy of your `fit` results to include in your notebook. Use

---

[11] http://www.physics.csbsju.edu/stats/cstats_paste_form.html

`plot` (see page 183 for a command summary) to produce a hardcopy plot of your data with fitted curve to include in your notebook. (Copy and paste is the easiest way to accurately transfer `K1` between `fit` and `plot`.)

The formula for $\hat{c}(R)$ (Eq. 3.41) is a bit tricky to type in, so I've provided you with a shortcut: $\hat{c}(R)$ will be automatically entered as `f(x)` into `fit` or `plot` if you type:

```
   * @cr.fun
```

In this `f(x)`, `K1` holds the parameter R_m, `x` is a measured R, and `f(x)` is the expected value $\hat{c}(R)$.

Calculation of m_μ

Since energy is conserved, the rest energy of the muon must end up in its decay products:

$$E_\mu = m_\mu c^2 = E_e + E_\nu + E_{\bar{\nu}} \tag{3.43}$$

where E_μ, E_e, E_ν, $E_{\bar{\nu}}$ are respectively the total energy of the muon, electron, neutrino, and antineutrino.

You should expect that the rest energy of the muon would, on average, be evenly divided between the three particles. Thus

$$\langle E_e \rangle \approx \langle E_\nu \rangle \approx \langle E_{\bar{\nu}} \rangle \approx \frac{1}{3}\, m_\mu c^2 \tag{3.44}$$

In fact, the Weinberg-Salam theory of weak decays predicts

$$\langle E_e \rangle = .35 m_\mu c^2 \tag{3.45}$$

Since electrons with that much energy have $\beta = .9996$, to a good approximation $E_e = p_e c$. Thus

$$.35 m_\mu c^2 = \langle E_e \rangle = \langle p_e c \rangle = \frac{4}{\pi}\, \langle p_{e\perp} c \rangle \tag{3.46}$$

so you will find $m_\mu c^2$ from $\langle p_{e\perp} c \rangle$ (which, in turn, may be found from $\langle R \rangle$).

Additionally, you will calculate $m_\mu c^2$ from your `fit` value of R_m (using your data and the theoretical expression for $\hat{c}(R)$, Equation 3.41). From the above discussion you already know

$$3BR_m = p_\perp c = m_\mu c^2/2 \tag{3.47}$$

allowing you to determine $m_\mu c^2$ from B and R_m.

Recall: Standard deviation of the mean is used for errors in averages (see Taylor). The square root of the diagonal elements of the covariance matrix determine the errors in `fit` parameters (See Press, *et al.*).

Calculation of m_π

The pion mass can be determined from the muon path length L: From L you can find the initial kinetic energy of the muon (T_0); adding the rest energy of the muon (use a high accuracy book-value for $m_\mu c^2$) gives you the total muon energy, E_μ.

$$E_\mu = T_0 + m_\mu c^2 \tag{3.48}$$

From momentum conservation and the fact the neutrinos have nearly zero rest mass and hence travel at the speed of light, you can show

$$E_\nu = p_\nu c = p_\mu c = \left[E_\mu^2 - (m_\mu c^2)^2\right]^{1/2} \tag{3.49}$$

Finally, energy conservation of the decaying pion requires

$$m_\pi c^2 = E_\mu + E_\nu = E_\mu + \left[E_\mu^2 - (m_\mu c^2)^2\right]^{1/2} \tag{3.50}$$

from which $m_\pi c^2$ can be calculated.

We have discussed two ways of determining L (using Equation 3.31 with a measured value for $\langle L_\perp \rangle$ and by `fit`ting the theoretical $\hat{c}(L_\perp)$ curve to your experimental data), so you will produce two values for m_π.

Note: The error in the book value of m_μ used to calculate m_π should be small enough to ignore, so the error in $m_\pi c^2$ ($\delta m_\pi c^2$) is due to the uncertainty in L (given by `fit`) or the uncertainty in $\langle L_\perp \rangle$ (given by the standard deviation of the mean). To calculate $\delta m_\pi c^2$, find $\dfrac{\partial m_\pi c^2}{\partial E_\mu}$ and $\delta E_\mu = \delta T_0$ and apply Eq. E.9 (or E.12) in the form:

$$\delta m_\pi c^2 = \frac{\partial m_\pi c^2}{\partial E_\mu}\, \delta E_\mu \tag{3.51}$$

Some Words About Errors

We noted above that different analysis methods yield different statistical error estimates. And robust methods that produce smaller uncertainties are preferred. But no amount of statistical gamesmanship can erase a systematic error in the original measurements so there is little point in reducing statistical errors once systematic errors dominate. Said differently: the aim is to understand the systematic errors and then reduce the statistical errors until they become irrelevant.

In this experiment one finds potential systematic errors in constants given without error ($B = .88$ T, $\rho = .07$ g/cm^3, the "2.1" in the Bethe-Block Equation 3.15) and theoretical simplifications (the bubble chamber's 6" thickness means paths closer to the camera are enlarged in photographs compared to those further from the camera, 'straggling' where some muons would have stopping distances a bit more or less than that calculated from the Bethe-Block Equation, varying magnetic field within the bubble chamber, expansion/contraction of the photographs due varying to humidity).

$m_\mu c^2$ (MeV)		$m_\pi c^2$ (MeV)	
113	±1	140.36	±.02
109		140.10	
106		139.61	
102		139.98	
103		140.17	
110		140.15	
112		140.26	
114		139.55	
114		139.92	
105		140.09	
97		140.12	
109		140.24	
109		140.49	
106		140.24	
112		140.21	
100		140.48	
99		140.24	
99		140.44	
106		139.81	
105		140.23	

	$m_\mu c^2$	$m_\pi c^2$
mean:	106.5	140.13
standard deviation:	5	0.26
discrepancy (mean−known):	1	0.56

Table 3.1: Three years of student results (same photos, same apparatus) for the muon mass ($m_\mu c^2$ in MeV) and the pion mass ($m_\pi c^2$ in MeV) as determined using a fit to the appropriate cumulative fraction function. Values using averages were similar (if slightly more discrepant) but with much larger errors. The error listed here is just typical: clearly each experiment will report both differing errors and values. **Note particularly** the failure of the reported error to accurately predict the actual variation (i.e., reported errors smaller than standard deviation of masses, e.g., ±.02 vs. 0.26). Almost all of these results would declare the actual value wrong. Evidently there are unrecognized sources of random variation which result in a miss estimation of random error. In the case of $m_\pi c^2$ the discrepancy is several times the SDOM; systematic error is the likely cause. (Note that the `fit`-reported error in L is usually unbelievably small.)

The error in L determined from the fit to the theoretical cumulative fraction function is often obviously too-small[12]. The covariance matrix typically suggests an L error of about .001 cm. Whereas `bubbleCAL` typically suggests systematic errors in L measurements greater than .01 cm. Further the likely error in the ρ and 2.1 that occur together with L in Eq. 3.17, would suggest L accuracy below 5% is irrelevant.

It is helpful to review multiple results collected over several years given in Table 3.1. The mean discrepancy is not too worrisome: a few percent change in the given constants would erase that difference. Of greater concern is large mismatch between the calculated statistical error and the standard deviation upon repeated measurement with identical equipment (but different data collectors). I'm sure part of this variation is due to 'personal placement decision' detected by `bubbleCAL`, but additionally varying ability to notice short $L_\perp$ will create varying results. (Most student cumulative fraction plots show a deficit of short $L_\perp$.) In any case we see here an example of calculated statistical errors reduced way below other uncontrolled uncertainties.

Given this table of experiment repeats, would it be fair to report the standard deviation of the mean as the uncertainty in, say $m_\pi c^2$? Numerically you can see that the result would be disastrous (as the SDOM is about $\frac{1}{10}$ the discrepancy). Since we are attributing the deviation to different observers (since the equipment and photos are the same) and we have no right to suggest that the average observer is the perfect observer (indeed one might say the perfect observer is exceptional), the answer must be "No". In short you can only use the SDOM as the error if you are convinced the deviation is caused by a balanced (exactly as much high as low) process. (And that implies you have a good idea of what is causing the deviations.)

In our experiment, the actual experimental error is available to us (since m_μ and m_π have since been measured to high accuracy), but how could the original investigators estimate these errors? A short answer to this question is calibration: "placing" known pathlengths and radii into the bubble chamber, and comparing the known values to the measured results. Indeed, experiments exactly like this lab are used to calibrate the detection systems in modern experiments.

Histograms

The cumulative fraction function described above is generally used when fewer then a thousand data points are available. Histograms are a more familiar way of displaying distributions. Histograms are made by dividing the range of the data set $\{x_i\}$ into several (usually equal-sized) sub intervals call bins. The data are sorted into these bins, and the number of data points in each bin is recorded as the y value for the average x value of the bin. (According to Poisson statistics, the uncertainty in y would be the square root of y.) The resulting plot is closely related to dc/dx (also known as the probability density function). Histograms are valued because they immediately show which bins are highly populated, i.e., what values of x occur frequently.

---

[12]The reason too-small errors are found when fitting to cumulative fraction functions may be found in Reference 2, Press, *et al.*

Report Checklist

1. Write an introductory paragraph describing the basic physics behind this experiment. Why did different decays of pions result in different $L_\perp$? Why did different decays of muons result in different R? What is the relationship between $L_\perp$ and more usual variables of motion? What is the relationship between R and more usual variables of motion? (This manual has many pages on these topics; your job is condense this into a few sentences and no equations.)

2. Book values for muon ($\mu^\pm$) and charged pion ($\pi^\pm$) 'mass' in MeV. Find (and cite) a source that reports at least five significant digits.

3. Derivations:

   (a) Equation 3.14
   (b) Equation 3.31
   (c) Area of a spherical cap
   (d) Use above to derive $\hat{c}(L_\perp)$

4. Read: `http://www.physics.csbsju.edu/stats/display.distribution.html`
   Let's call the last two digits of <u>your</u> CSB/SJU ID number: *XY*.

   (a) Go to the web site: `http://www.physics.csbsju.edu/370/data/` and select the file: `4_X.dat`. This file contains 4 random data points. *By hand* draw in your notebook the cumulative fraction step-curve for this dataset along with the percentile values with errors. The result should be similar to Figure 3.4. Show your work!

   (b) From the same web site, select the file: `1000_Y.dat` (or `1000_Yw.dat` with the same data in multiple columns). Using $\sim$10 bins *by hand* draw the histogram for this dataset including error bars. Show your work! Copy & paste the 1000 data points into the web[13] site and produce a hardcopy percentile plot. Comment on the relationship between features in your histogram and features in the percentile plot. E.g., a peak in the histogram corresponds to what in the percentile plot?

   (Note: I'm requesting individual work—rather than partnered work—to check that everyone understands these plots. Do feel free to help your partner succeed, but don't just do the work for him/her.)

5. Results (copied from the screen) and conclusions from use of the program `bubbleCAL`: Estimate the random error in <u>a</u> L and R measurement (think σ). Was systematic error detected? (Is the actual value outside the reported 95% confidence limits?) If so, how much? If not, how large could the systematic have been and nevertheless go undetected? (This is called an upper limit on the error.) Compare your systematic error (detected or upper limit) in L and R to the `fit` supplied estimates of random error in L and R_m for the actual experiment. Which is (or might be) more significant: random or systematic error?

   This is a good time to make a distinction between accuracy and precision. *Accuracy* is the deviation between the actual value and our experimentally determined

---

[13]`http://www.physics.csbsju.edu/stats/cstats_paste_form.html`

value, whereas *precision* is our estimate of the error in our experimentally determined value (typically based on a standard deviation or a computer reported parameter error). (This distinction between accuracy and precision is very commonly ignored since "accuracy" requires knowing the actual value (which is usually impossible), so we commonly are forced to use our estimate of precision as our estimate of accuracy. However, in this lab we do have "book" values which we assume are accurate.) Inaccurate results (actual value outside your reported precision) are usually the result of systematic errors.

6. Report value (and error) for $\langle R \rangle$ and $\langle L_\perp \rangle$. (An easy way to do this is to open the files `cf.R.dat` and `cf.L.dat` with a spreadsheet like `gnumeric` and use the spreadsheet statistical functions.) Report how you calculated the errors.

7. Two values (with errors) for $m_\mu c^2$ (in MeV): one based on $\langle R \rangle$ (see Eq. 3.46) and the other on a `fit` to Equation 3.41. Hardcopies of the `fit` results and a `plot` of the best-fit curve with your data points. (The plot is analogous to Figure 3.6.) Note: `fit` is described in Appendix A; `plot` is described in Appendix B. But perhaps the easiest starting point for these programs is the video that goes along with Appendix C.

8. Two values (with errors) for $m_\pi c^2$ (in MeV): one based on L found from $\langle L_\perp \rangle$ using Eq. 3.31 and the other based on L found from the `fit` to the equation for $\hat{c}(L_\perp)$ you derived in 3(d) above. Hardcopies of the `fit` results and a `plot` of the best-fit curve with your data points.

9. A comparison of the mass values you reported above (6 & 7) and the "known" values cited above in 2. (Make a nice table.)

10. Typically in this lab, the error estimates based on `fit`s that are much smaller than the errors based on averages. Are these error estimates accurate? (If your masses (with errors) are inconsistent with book values, it is quite likely that something is wrong with your error estimates.) Explain how your error estimates could be improved. (For example, sometimes error estimates can be improved (and reduced) by simply taking more data. In other situations, improved errors come from proper consideration of systematic errors and result in larger estimates of error.) A few words about statistical and systematic errors are probably required; you might consult Chapter 0 if nothing occurs to you.

References

1. Lederman, Leon & Teresi, Dick *The God Particle* (1993)— In this general-audience book Nobel Laureate Lederman describes his work as an experimental particle physicist. Of particular interest: his work with pions in a cloud chamber using the Nevis accelerator (pp. 219–224); building cloud and bubble chambers (pp. 244–247); parity violation detected in the $\pi \to \mu \to e$ decay sequence (pp. 256–273).

2. Press, Teukolsky, Vetterling & Flannery *Numerical Recipes* (1992)— Title says it all. Cumulative fraction is discussed in the section on the Kolmogorov-Smirnov Test; covariance matrix in chapter on modeling data.

3. Hogg & Tanis *Probability and Statistical Inference* (1993)— Order statistics are discussed in Chapter 10.

4. Taylor, J.R. *An Introduction to Error Analysis* (1997)— derivative rule error propagation, standard deviation of the mean, histograms, random and systematic errors

5. I know of no book that derives Equation 3.41. A starting point is the Kurie Plot for weak decays, for example, from Halzan & Martin *Quarks and Leptons* (1984) p. 263:

$$\frac{d\Gamma}{dE} = \frac{G^2}{12\pi^3} \, m^2 \, E^2 \, \left(3 - \frac{4E}{m} \right) \tag{3.52}$$

6. Enge, H.A. *Introduction to Nuclear Physics* (1966)— Chapter 7 has a nice discussion on stopping power, range, the Bethe-Block equation and particle detectors including bubble chambers.

7. Coughlan, G. D. & Dodd, J. E. *The Ideas of Particle Physics: An Introduction for Scientists* (2003)— An excellent introduction to particle physics at undergraduate physics major level.

8. This lab is based on "The Pi-Mu-e Experiment" 33-3908 (1966) by Ealing Corporation.

4: Photometry

The mathematical thermology created by Fourier may tempt us to hope that... we may in time ascertain the mean temperature of the heavenly bodies: but I regard this order of facts as forever excluded from our recognition.

Auguste Comte *Cours de la Philosophie Positive*[1] (1835)

...within a comparatively few years, a new branch of astronomy has arisen which studies the sun, moon, and stars for what they are in themselves and in relation to ourselves.

Samuel Pierpont Langley *The New Astronomy* (1888)

Purpose

In 1835 the French philosopher Auguste Comte noted that since we know stars only by their light (and cannot take bits of stars into the laboratory), our knowledge of stars would be forever limited essentially to location. Fifty years later astrophysics—the study of the physics of stars—was beginning, and the first measurements of the Sun's temperature and composition were being made. Evidently, careful measurement of starlight (photometry) allows the intrinsic properties of stars to be determined. In this lab you will use broadband photometry to measure the temperature of stars.

Introduction

We begin our study of stars with blackbody radiation, which you studied in Modern Physics and you will yourself measure in the Thermionic Emission lab (p. 121). The light produced by a hot object is akin to audio noise: in both cases a random process produces a simultaneous superposition of a wide distribution of wavelengths. In his *Opticks*, Newton noted certain systematic dependencies in the light emitted by incandescent objects and suggested something fundamental was behind the process[2]. Apparently the light produced by an incandescent object is fundamentally related to its temperature not its material composition. In 1879 Joseph Stefan proposed that every object emits light energy at the rate (power in

---

[1] `http://socserv2.mcmaster.ca/~econ/ugcm/3ll3/comte/` provides this as an etext translated and edited by Harriet Martineau.

[2] *Query 8*: Do not all fix'd Bodies, when heated beyond a certain degree, emit Light and shine...
Query 11:...And are not the Sun and fix'd Stars great Earths vehemently hot...

watts):

$$P = \epsilon_T \sigma T^4 \mathcal{A} \tag{4.1}$$

where σ is the Stefan-Boltzmann[3] constant, T is the temperature of the body, $\mathcal{A}$ is the surface area of the body, and ϵ_T is the total emissivity.

Experimentally it was found that the wavelength distribution of the light from hot objects was broadly similar to the Maxwell-Boltzmann speed distribution: a bell-shaped curve in which the location of the peak depended on temperature. In 1893 Wilhelm Wien[4] concluded that the wavelength of this peak must be inversely proportional to temperature. Experiment confirmed that the wavelength of this peak was given by:

$$\lambda_{max} = \frac{2898 \ \mu\text{m} \cdot \text{K}}{T} \tag{4.2}$$

Together the Stefan-Boltzmann and Wien displacement laws—both the results of classical physics—explain much of what is commonly observed in hot objects. At room temperature objects do not appear to be a source of light: they appear dark (unless externally illuminated). This is a result both of the small rate of emission given by Stefan-Boltzmann and (from Wien) for $T = 300$ K, we find $\lambda_{max} \sim 10 \ \mu$m, that is the wavelengths typically emitted are well outside the range of light that can be detected by the eye[5]. On the other hand for an object like the Sun: $T \approx 6000$ K, so $\lambda_{max} \sim 0.5 \ \mu$m — right in the middle of of the visible range.

Our 'spherical cow' model[6] of a star is an incandescent ball (radius R) of gas shining as a blackbody. Thus the total power output of a star (called the *luminosity*[7]) is

$$L = \sigma T^4 \ 4\pi R^2 \tag{4.3}$$

The equation for the exact distribution of photon wavelengths produced by a blackbody was actively researched with erroneous equations produced by Wien and Rayleigh & Jeans. In 1900 Max Planck[8] published his derivation of the wavelength distribution of light from a blackbody. However before we can discuss his result we must explain what precisely is meant by 'wavelength distribution'.

In Physics 211 you learned about the speed distribution[9] of the molecules in a gas: a bell-shaped curve that shows that slow molecules are rare and supersonic molecules are

---

[3]In 1879 Jožef Stefan (1835–93) proposed the law based on experimental measurements; five years later Ludwig Boltzmann (1844–1906) derived the result from fundamental thermodynamics. Today the result is known as the Stefan-Boltzmann Law.

[4]Wilhelm Wien (1864–1928) German physicist, 1911 Nobel Prize in Physics

[5]The human eye can detect light with a wavelength in the range 0.4 μm (violet) to 0.7 μm (red).

[6]This model of a star works best with stars similar in temperature to the Sun, but every star's light is considerably modified by absorption as it travels through the star's atmosphere. In hot stars, UV wavelengths that can photoionize hydrogen ($\text{H}(n=2)+\gamma \rightarrow \text{H}^+ + e^-$) are highly attenuated producing the Balmer jump. The spectra of cool stars shows broad absorption bands due to molecules.
`http://www.jb.man.ac.uk/distance/life/sample/java/spectype/specplot.htm` is an applet allowing a comparison of blackbody light to actual star light.

[7]The MKS unit for luminosity is watt, but the numbers are so large that it is usually expressed as a multiple (or fraction) of the Sun's luminosity (a.k.a., solar luminosity) $L_\odot = 3.846 \times 10^{26}$ W.

[8]Max Planck (1858–1947) German physicist, 1918 Nobel Prize in Physics

[9]Derived by James Clerk Maxwell (1831–1879); Ludwig Boltzmann (1844–1906) provided a firm foundation for the result using statistical mechanics. Today the result is generally known as the Maxwell-Boltzmann distribution.
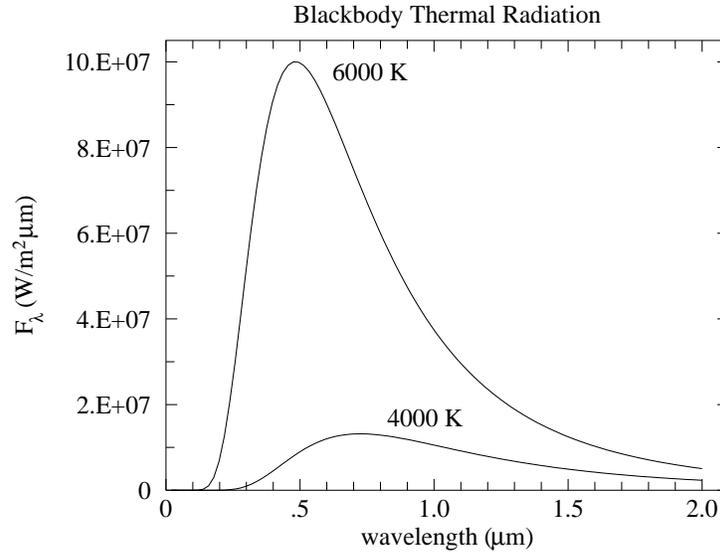
Figure 4.1: The distribution of the light emitted by a blackbody plotted as a function of wavelength. Hotter objects emit much more light than cool ones, particularly at the shorter wavelengths.

rare, whereas a molecular kinetic energy near $\frac{3}{2}kT$ is common. 'Speed probability' is a problematic concept. For example the probability that a molecule has a speed of exactly π m/s (or any other real number) must be essentially zero[10] (as an infinite number of digits must match, but there are only a finite [but huge] number of molecules). Probability density provides a meaningful context to talk about speed distribution: determine the fraction of molecules ('hits') that have speed in an interval (bin) $(v - \Delta v/2, v + \Delta v/2)$ for various speeds v. (Here the bin is centered on the speed v and has bin-size Δv.) Clearly the number of hits in a bin depends on the bin-size (or range), but you should expect that the hit *density* (number of hits divided by the bin-size) should be approximately independent of the bin-size. A plot of fraction-per-bin-size vs. bin center provides the usual bell-shaped curve. Bins centered on both slow and supersonic speeds include few molecules whereas bins that correspond to molecular kinetic energy near $\frac{3}{2}kT$ include a large fraction of molecules.

In a similar way we can sort the light by wavelength into a sequence of wavelength intervals (bins), and calculate the total light *energy* per bin-size. For example, if there were[11] a total of 2 J of light energy (about 5×10^{18} photons) with wavelengths in the interval $(.50, .51)$ μm we would say the intensity of light at $\lambda = .505$ μm was about 200 J/μm. In the case of the thermal radiation continuously emitted from the surface of a body, we are interested in the *rate* of energy emission per area of the body, with units $(\mathrm{W/m^2})/\mu$m. This quantity is called the monochromatic flux density and denoted F_λ[12]. Planck's result was:

$$F_\lambda = \frac{2\pi h c^2}{\lambda^5} \frac{1}{\exp(hc/\lambda kT) - 1} \tag{4.4}$$

---

[10]A mathematician would say the probability is zero for *almost every* speed.

[11]These numbers correspond approximately to 0.1 sec of full sunlight on 1 m^2 of the Earth.

[12]One could just as well measure bin-size in Hz. The result is denoted F_ν with units $(\mathrm{W/m^2})/\mathrm{Hz}$. A helpfully sized unit for F_ν is the jansky: 1 Jy $= 10^{-26} \mathrm{W \cdot m^{-2} \cdot Hz^{-1}}$. F_ν is commonly used in radio astronomy whereas F_λ is commonly used in optical astronomy.

where h is Planck's constant, whose presence is a sign of the importance of quantum mechanical effects. The primary assumption in Planck's derivation of this distribution was the *photon*: an indivisible packet of light carrying total energy $E = hc/\lambda$. With this identification notice the Boltzmann factor in Planck's equation: $\exp(E/kT)$. As before, this factor means that high energy (i.e., short wavelength) photons are rarely in the mix of emitted photons. Thus only 'high' temperature stars will emit an abundance of UV radiation.

Of course, the light emitted from the surface of a star gets spread over an ever larger area as it moves away from the star. Thus the flux coming into a telescope a distance r from the center of the star would be:

$$F_\lambda = \frac{2\pi hc^2}{\lambda^5} \frac{R^2}{r^2} \frac{1}{\exp(hc/\lambda kT) - 1} \tag{4.5}$$

Theoretically its now easy to measure the temperature of stars: simply measure the wavelength distribution ('spectra') of the starlight and fit the measurements to the above theory with two adjustable parameters: T and R/r. (Do notice that increasing R or decreasing r produces exactly the same effect—an overall increase in the flux density—so we cannot separately determine R and r from the spectra. A large, distant $T = 6000$ K star could have exactly the same spectra as a small, close $T = 6000$ K star.) Let me enumerate a few (of the many) reasons this does not work perfectly.

1. We assumed above that space was transparent: that the only effect of distance was diluting the light over a greater area. However, our Galaxy is filled with patchy clouds of dust. Unfortunately the dust in these clouds has the property of being more likely to absorb blue light than red. Thus a dust-obscured star will lose more blue than red light and the observed spectra will be distorted from that emitted. A dust-obscured star would measure cooler and appear more distant (i.e., dimmer) than it actually was. These two effects of dust are called 'reddening' and 'extinction'. Of course substantial dust absorption is more likely for more distant stars. It is usually 'small' for stars within 300 Ly of Earth. However, 300 Ly covers a very tiny fraction of our Galaxy. In this lab we will side step this problem by selecting stars with little dust absorption. However, with a bit of additional work, dust absorption could be measured and corrected for using photometry.

2. Starlight measured from the surface of the Earth must traverse the Earth's atmosphere. The non-transparency of the Earth's atmosphere will also modify the observed starlight. In addition, the telescope and detector typically introduce additional 'non-transparency', that is blue and red photons at the telescope aperture are not equally likely to be counted. The 'efficiency' of the system depends on wavelength. Unless extraordinary care is taken, this efficiency can change on a daily basis. (And, of course, the atmosphere's transparency can change in a matter of minutes.) As a result frequent calibration of the system is required. Calibration at the multitude of wavelengths required to make a full spectra is obviously more difficult than calibration at just a few wavelengths.

3. There is a competition between the accurate measurement of the value of F_λ and the bin-size used. As you may have noticed in the Bubble Chamber lab, if you select a small bin-size, each bin captures relatively few hits which results in relatively large $\sqrt{N}$ errors. So if you have only a few thousand photons, you may do better to use a big bin-size (so you capture enough counts to get an accurate measurement of F_λ),

but then only have a few bins (each spanning a large interval of in wavelength) so, unfortunately, the spectrum has been reduced to just a few points. Of course, you could always collect starlight for a longer period of time (longer 'integration time') or use a telescope with a larger aperture. However, these solutions miss the point: The boundary between the known and unknown in astronomy is almost always at the edge of what you can just barely detect. Thus you must always come up with the maximally efficient use of the available photons. Telescope time is allocated for a detailed spectra only when it is expected that the results cannot be obtained in a 'cheaper' way.

4. Stars are not exact blackbodies, and of course they are not at <u>a</u> temperature. Certainly if we move from the surface to core of a star we would experience quite different temperatures. But even on the 'surface' of our Sun we see regions ('sunspots') with unusually low temperature. In the Langmuir Probe lab, you will find that the electrons in a tube of gas may have a different temperature from the co-mingled atoms. Similarly, in the atmospheres of stars it is not unusual for the various components of the gas to be at different temperatures (that is to say the gas is not in local thermodynamic equilibrium [LTE]). Hence it may make no sense to try to obtain high-precision measurements of 'the' temperature.

   Oddly enough the fact that stars are not blackbodies is actually helpful as it allows a variety of information to be decoded from the starlight. (A blackbody spectra includes just two bits of information the temperature T and R/r.) Varying absorption in the star's atmosphere means that some light (i.e., λ at which the star's atmosphere is largely transparent) comes from deeper in the star and hence represents a higher temperature. Ultimately the absorption lines in a star's spectra provide the bulk of the information we have about stars. Chemical composition of the star's atmosphere and temperature—accurate[13] to a few percent—are best obtained this way. However, consideration of these absorption lines is beyond the aims of this lab.

Temperature I

Your aim in this lab is to measure the temperature of stars without resorting to a detailed measurement of the star's spectra. We begin by considering measurement of F_λ at just two wavelengths: F_1 at λ_1 and F_2 at λ_2 where $\lambda_1 < \lambda_2$. (Think of λ_1 as blue light and λ_2 as red light.) There is a huge cancellation of factors if we look at the ratio: F_2/F_1:

$$\frac{F_2}{F_1} = \frac{\lambda_1^5}{\lambda_2^5} \frac{\exp(hc/\lambda_1 kT) - 1}{\exp(hc/\lambda_2 kT) - 1} \tag{4.6}$$

In general, ratios are great things to measure because (A) since they are dimensionless they are more likely to be connected to intrinsic properties and (B) it is often the case that systematic measurement errors will (at least in part) cancel out in a ratio. For starlight, the flux will depend on distance ($\propto 1/r^2$), but in a flux ratio the effects of distance will (in theory) cancel out.

---

[13]Statements like this—that imply we know the true accuracy of our measurements—should be read with the knowledge that history shows unexpected jumps in results as systematic errors are discovered. The history of the measured chemical composition of stars would show graphs much like those in Figure 1 on page 10.

While the ratio has considerably reduced the complexity of the formula, it would help if we had an even simpler formula. Towards that goal we make the approximation that the exponential terms are much larger than 1:

$$\frac{F_2}{F_1} = \frac{\lambda_1^5}{\lambda_2^5} \frac{\exp(hc/\lambda_1 kT) - 1}{\exp(hc/\lambda_2 kT) - 1} \tag{4.7}$$

$$\approx \frac{\lambda_1^5}{\lambda_2^5} \frac{\exp(hc/\lambda_1 kT)}{\exp(hc/\lambda_2 kT)} = \frac{\lambda_1^5}{\lambda_2^5} \exp\left[\frac{hc}{kT}\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right)\right] \tag{4.8}$$

$$\log(F_2/F_1) \approx \frac{hc}{kT}\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right)\log e + 5\log(\lambda_1/\lambda_2) \tag{4.9}$$

Example

Consider the case where $\lambda_1 = .436$ μm (blue) and $\lambda_2 = .545$ μm (green or 'visible'). For historical reasons, astronomers prefer to consider $2.5\log_{10}(F_2/F_1)$, and if we evaluate all of the constants we find:

$$2.5\log_{10}(F_2/F_1) \approx \frac{7166 \text{ K}}{T} - 1.21 \tag{4.10}$$

Do notice that for cool stars this quantity is large (more 'visible' than blue light) and that extremely hot stars ($T \to \infty$) will all have much the same value for this quantity: -1.21. This last result is a consequence of the 'Rayleigh-Jeans' approximation for the Planck distribution, valid for $\lambda \gg \lambda_{max}$. In this large wavelength (or high temperature) limit, the Boltzmann factors are nearly one and:

$$\exp(hc/\lambda kT) - 1 \approx hc/\lambda kT \tag{4.11}$$

so[14]

$$F_\lambda \approx \frac{2\pi hc^2}{\lambda^5} \frac{\lambda kT}{hc} = kT \frac{2\pi c}{\lambda^4} \tag{4.12}$$

Since temperature is now an over-all factor, it will cancel out in any flux ratio.

Color Index

At increasing distance from the star, both F_2 and F_1 will be diminished, but by exactly the same fraction (assuming space is transparent, or at least not colored). Thus this ratio is an intrinsic property of the star, and you saw above that this ratio is related to the temperature of the star. The ratio is, of course, a measure of the relative amounts of two colors, and as such is called a *color index*. (Our experience of color is independent of both the brightness of the object and the distance to the object. It instead involves the *relative* amounts of the primary colors; hence the name for this ratio.) Any two wavelengths λ_1, λ_2 can be used to form such a color index, but clearly one should select wavelengths at which the stars are bright and $\lambda \lesssim \lambda_{max}$ (to avoid the Rayleigh-Jeans region where the ratio is independent of temperature).

---

[14]Do notice the absence of h in this classical result.
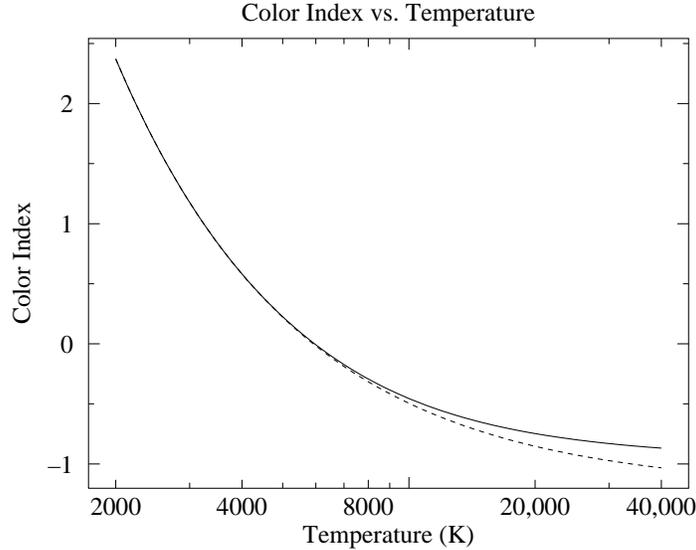
Figure 4.2: In the case where $\lambda_1 = .436$ μm (blue) and $\lambda_2 = .545$ μm (green or 'visible'), the 'color index' $2.5 \log(F_2/F_1)$ is plotted as a function of temperature. The dotted line shows the approximate result Eq. 4.10; the solid line shows the actual ratio of Planck functions. Do remember that high temperature stars have a smaller, even negative, color index.

Magnitude

For historical reasons, astronomers typically measure light in *magnitudes*. Magnitudes are related to fluxes by:

$$m = -2.5 \log_{10}(F/F_0) \tag{4.13}$$

where F_0 is the standardizing[15] flux that corresponds to $m = 0$. In astronomy log commonly[16] refers to $\log_{10}$ (and so in the rest of this chapter I now drop the $_{10}$), although we could equivalently write this equation in terms of other bases:

$$m = -2.5 \log_{10}(F/F_0) = -1.085736 \ \ln(F/F_0) = -\log_{2.511886}(F/F_0) \tag{4.14}$$

The minus sign in this definition means that magnitudes work backwards: the brighter the star, the *smaller* the magnitude. Very bright objects would have negative magnitudes. Since the measured flux of a star depends on distance, magnitude is not an intrinsic property of a star. Assuming transparent space, magnitude can be related to luminosity and distance:

$$m = -2.5 \log \left(\frac{L}{4\pi r^2 F_0} \right) \tag{4.15}$$

Thus if a star's luminosity is known and its magnitude measured, its distance can be calculated. Astronomers define the *absolute magnitude* of a star to be the magnitude the star

---

[15]Historically F_0 was defined as the flux of the star Vega, so Vega by definition had magnitude zero and color index zero.

[16]Pun intended. Use of $\log_{10}$ is a bit old fashion. If you see 'log' (without an explicit base or further comment) in a modern physics document, you should probably assume $\ln = \log_e$ is intended, unless something historical is being discussed (like stellar magnitudes, dB or pH).

would have if observed from a distance of 10 pc[17]. Thus:

$$m = -2.5 \log \left[\frac{L}{4\pi r^2 F_0} \right] \tag{4.16}$$

$$= -2.5 \log \left[\left(\frac{L}{4\pi (10 \text{ pc})^2 F_0} \right) \cdot \left(\frac{10 \text{ pc}}{r} \right)^2 \right] \tag{4.17}$$

$$= -2.5 \log \left[\frac{L}{4\pi (10 \text{ pc})^2 F_0} \right] + 5 \log \left[\frac{r}{10 \text{ pc}} \right] \tag{4.18}$$

$$= M + 5 \log \left[\frac{r}{10 \text{ pc}} \right] \tag{4.19}$$

where M is the absolute magnitude and the remaining term: $5 \log(r/10 \text{ pc})$ is known as the distance modulus. The main point here is that distance affects magnitude as a constant offset. This result is a lemma of the general result:

Theorem: If the measured flux F_m is a fraction (ϵ) of the actual flux F (as would occur due to a less-than-perfectly efficient detector, a dusty mirror, atmospheric absorption, absorption due to interstellar dust,...) the resulting magnitude as measured (m_m) is just a constant ($2.5 \log \epsilon$) off from the actual magnitude (m).

Proof:

$$m_m = -2.5 \log(F_m/F_0) \tag{4.20}$$

$$= -2.5 \log(\epsilon F/F_0) \tag{4.21}$$

$$= -2.5 \log(F/F_0) - 2.5 \log(\epsilon) \tag{4.22}$$

$$= m - 2.5 \log(\epsilon) \tag{4.23}$$

In order to calculate flux (W/m^2) from photon count rates, we must divide by the photon collection efficiency, multiply by the energy per photon and divide by the collecting area of the telescope. But why bother: the above theorem says these multiplicative factors just result in a constant offset between the correctly calculated magnitude and one that simply ignores these factors. Furthermore, many of these complicating factors (e.g., the fraction of photons absorbed in the atmosphere) cannot be known in advance. Since some of these factors cannot be known in advance, the plan is to ignore all of them (so the computer program[18] calculating magnitude will not ask you for the energy/photon, quantum efficiency, telescope aperture, etc.), and instead experimentally determine the offset between the computer's magnitudes (which we will denote in lowercase, e.g., b) and the standard magnitude (uppercase: B). So part of this lab will involve finding the constant offset that relates our measured (instrumental) magnitudes to the actual (standardized) magnitude.

Filters

In the above equation for magnitude, the flux might be the total (all wavelengths or 'bolometric') light flux, or a monochromatic flux density F_λ, or the flux in a particular range of wavelengths. Much of astronomy is concerned with the flux through the five standard[19]

---

[17]pc = parsec = 3.0857×10^{16} m = 3.2616 Ly

[18]`gaia` magnitudes are calculated as: $50 - 2.5 \log(\text{gain} \times \text{net\_counts/exposure\_time})$

[19]Johnson, H. L. and Morgan, W. W. (1951) ApJ **114** 522 & (1953) ApJ **117** 313
Cousins, A.W.J. (1976) memRAS **81** 25

Figure 4.3: The characteristics of the standard filters: U (ultraviolet), B (blue), V (visible), R (red), I (infrared) from *Allen's Astrophysical Quantities* p. 387 and The General Catalogue of Photometric Data: `http://obswww.unige.ch/gcpd/system.html`

U, B, V, R, and I filters. The characteristics of these filters are detailed in Figure 4.3, but essentially each filter allows transmission of a range of wavelengths $\Delta\lambda$, centered on a particular wavelength. The name of the filter reports the type of light allowed to pass: U (ultraviolet), B (blue), V ('visible', actually green), R (red), and I (infrared). The filters are broad enough that the resulting bin really doesn't well represent a value for F_λ. (An appendix to this chapter describes the $u'g'r'i'z'$ filters[20] used in the Sloan Digital Sky Survey.)

The magnitude of a star as measured through a B filter is called the B magnitude and is simply denoted: B. Notice that we can now form color indices just by subtracting two magnitudes. For example, the most common color index is $B - V$:

$$B - V \;\; = \;\; 2.5\left(\log(F_V/F_{V0}) - \log(F_B/F_{B0})\right) \tag{4.24}$$

$$= \;\; 2.5\left(\log(F_V/F_B) + \log(F_{B0}/F_{V0})\right) \tag{4.25}$$

$$= \;\; 2.5\log(F_V/F_B) + \text{constant} \tag{4.26}$$

So $B - V$ is related to the flux ratio F_V/F_B and so is an intrinsic property of the star related to temperature. Furthermore, while Eq. 4.10 was derived assuming monochromatic

---

[20]Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996 AJ, 111, 1748
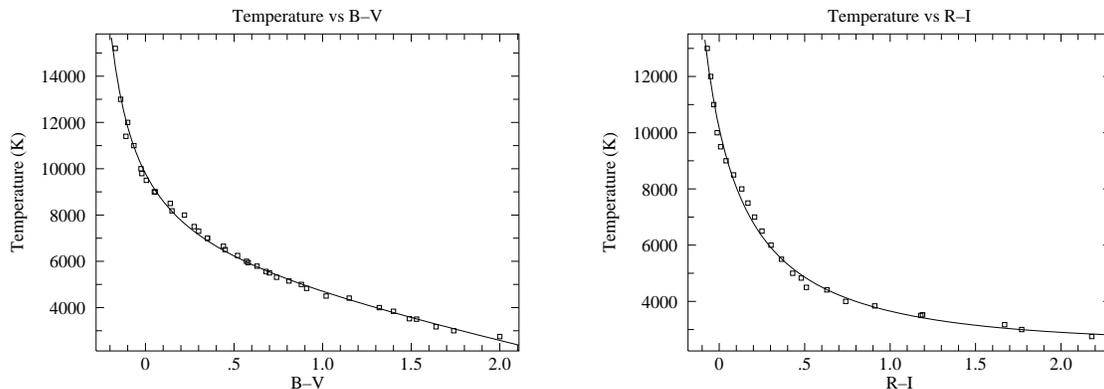
Figure 4.4: Using data from *Allen's Astrophysics Quantities* calibration curves relating stellar temperature to color index can be obtained. The fit equations for these curves is given in the text.

flux densities (in contrast to the broad band fluxes that make up B and V), we can still hope equations similar to Eq. 4.10 can be derived for $B - V$. *Allen's Astrophysical Quantities* provides calibration data to which Kirkman has fit a curve:

$$B - V = \frac{35800}{T + 3960} - 3.67 + 1.08 \times 10^{-4} \, T \tag{4.27}$$

and the inverse relationship:

$$T = \frac{1700}{(B - V) + .36} + 5060 - 1600(B - V) \tag{4.28}$$

The data with fitted curve is plotted in Figure 4.4. Notice that for $B - V < 0$ small uncertainties in $B - V$ result in large uncertainties in T: it would probably we wise to switch to a different color index like $U - B$ to measure the temperature of such a hot star.

Similar work can be done for $R - I$, with results:

$$T = \frac{2750}{(R - I) + .314} + 1400 + 160(R - I) \tag{4.29}$$

For the Sun, *Allen's Astrophysics Quantities* reports: T=5777 K, $B - V$=0.65, $R - I$=0.34 whereas Eq. 4.28 gives 5703 K and Eq. 4.29 gives 5659 K. In general errors of a few percent should be expected.

Temperature II

The starting point for physics is usually hard intrinsic quantities like temperature, density and pressure. However, in astronomy the transducers[21] used to measure these quantities are

---

[21] In normal usage, a transducer is a device which converts a physical quantity (e.g., pressure, temperature, force,...) to an electrical quantity (volts, amps, Hz ...) for convenient measurement: for example, a microphone. Here I've broadened the usual meaning to include natural 'devices' whose emitted or absorbed *light* allows a physical quantity to be determined. Obviously we are connected to the stars by light, not copper wires.

Telescope	Meade LX200	CCD	Kodak KAF-1001E
aperture	12"	pixel	24 μm $\times$ 24 μm
focal ratio	$f/10$	full-well	$\sim 150,000\ e^-$/pixel
focal length	120"	read noise	15 e^- rms
		gain	2.03 e^-/ADU
Filter Wheel	SBIG FW8-STL		
filters	blank, $BVRI$, $g'r'i'$	λ	Quantum Efficiency
Camera	SBIG ST-1001E NABG	.95 μm	20%
pixels	1024×1024	.80 μm	50%
image	$28' \times 28'$	.65 μm	65%
scale	1.65"/pixel	.50 μm	50%
weight	7.5 lbs	.35 μm	20%

Table 4.1: CCD photometry equipment used at the SJU observatory.

parts of the star itself (say the absorption lines of a particular element in the atmosphere of the star). We will always be a bit uncertain about the exact situation of these transducers, and hence the calibration of these transducers is correspondingly uncertain. For example stars with unusual chemical composition (like population II stars) or unusual surface gravity g (like giant stars) really deserve separate calibration curves. Since the usual physical quantities are in astronomy provisional and subject to recalibration, astronomers attach primary importance to the quantities that are not subject to the whims of revised theory: the actual measurements of light intensity. Thus astronomers use as much as possible hard quantities[22] like $B - V$ directly without converting them to temperature using some provisional formula. For example just using $B - V$, stars can be arranged in temperature-increasing order, since there is a monotone relationship between $B - V$ and T. Of course, in this lab the aim is to measure star temperatures in normal units.

Summary

In this lab you will measure the temperature of stars by measuring their B, V, R, I magnitudes, calculating the color indices $B-V$ and $R-I$, and then, using the supplied calibration curves, find T. The calibration curves are based on detailed spectra of bright, normal stars; In using these calibration curves we are automatically assuming our (distant and hence dimmer) target stars are also 'normal'.

Detector System

In this lab photons are counted using a *charge coupled device* or CCD. Our Kodak KAF-1001E CCD consists of a typical 26-pin DIP integrated circuit with a window that exposes to

---

[22]In this lab you will see that even as simple a quantity as B is actually only distantly related to direct meter readings. The direct meter readings must be 'reduced' to eliminate confounding factors like detector efficiency or atmospheric absorption. Nevertheless, since these problems are 'at hand' astronomers believe that they can be properly treated. Astronomers willing archive B data, whereas T calculations are always considered provisional.

light a 1"$\times$1" field of 1024$\times$1024 light sensitive regions or pixels. Photons incident on a pixel free electrons via the photoelectric effect[23]. Each pixel stores its charge during an exposure, and then each bucket of charge is transferred (in a manner similar to a bucket brigade) to a capacitor where it produces a voltage proportional to the number of electrons. An analog-to-digital converter then converts that voltage to a 16-bit binary number (decimal: 0–65535). A CCD image is thus a 1024$\times$1024 matrix of integers. The units of these integers is simply ADU for analog to digital units, but ideally the result should be proportional to the number of incident photons. The combined effects of photon$\rightarrow e^-$ efficiency ('quantum efficiency') and $e^- \rightarrow$ADU ('gain') means that there are ~ 4 times as many incident photons as ADU counts. Electron storage in a pixel is limited to a full-well capacity of $\sim 150,000\ e^-$; additional photoelectrons will simply leak out of the well.

The Kodak CCD is contained within (and controlled by) a STL-1001E camera made by the Santa Barbara Instrument Group[24]. In front of the camera, an SBIG FW8-STL filter wheel allows computer controlled placement of one of the B, V, R, I, g', r', i' filters or a blank.

The SBIG camera is mounted on a Meade 12" LX200 Schmidt-Cassegrain Catadioptric telescope. The orientation of the telescope is controlled by a Paramount ME mount using TheSky6[25] software. Once the telescope has been initialized the telescope control software allows the telescope to be directed to a long list of cataloged objects or to any location specified by RA/dec[26].

Linearity

In an ideal CCD, the count in a particular pixel should be proportional to the number of photons absorbed, which in turn should be proportional to the exposure time. This proportionality fails for large exposures and the CCD saturates as the count approaches 60,000 ADU. (The count actually jumps *up* to the 16-bit maximum 65535 on saturation.) Of course, if a pixel is saturated, additional photons are missed, and the (now pegged) ADU count falls below actual photon flux. CCD linearity was tested by varying the exposure time. The results are displayed in Figure 4.5.

The calculated magnitude of a star—a measurement of the photon flux (counts/time)—should not depend on the exposure time. Of course, for very short exposures the handful of counts above background is subject to a relatively large random fluctuation, and so the magnitude will have a relatively large uncertainty. As we've just learned, sufficiently long exposures will result in pixel saturation and hence a systematically low count and a too-large (dim) magnitude. As shown in Figure 4.6, there is a large range of exposures (including those with slight saturation) which can produce an accurate magnitude.

---

[23]Albert Einstein (1879–1955) received the 1921 Nobel prize for his 1905 theory explaining this effect. Robert Millikan's [(1868–1953), B.A. Oberlin College, Ph.D. Columbia] experimental studies (1912–1915) of the photoelectric effect were cited, in part, for his Nobel in 1923.

[24]`www.sbig.com`

[25]both from `bisque.com`

[26]See page 85 for a list of online tutorials covering basic astronomy vocabulary.
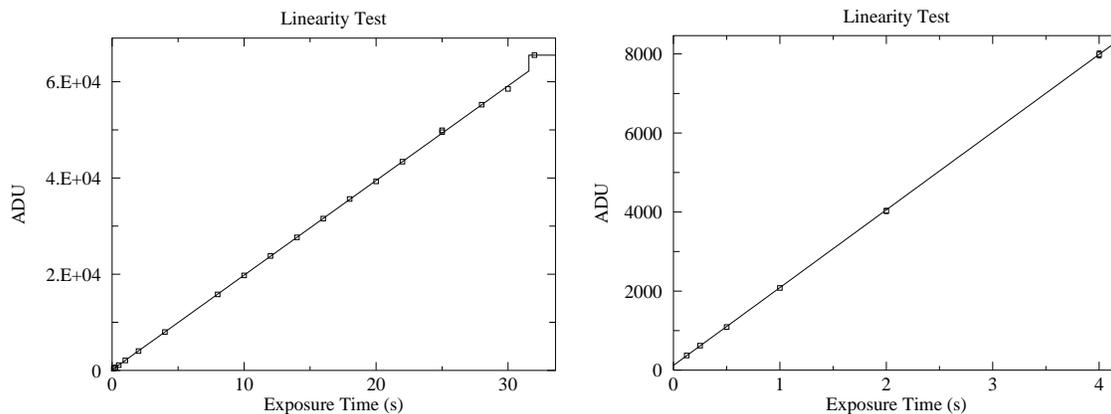
Figure 4.5: Below $\sim 60,000$ ADU (30 second exposure) the response of our CCD seems to be linear (reduced $\chi^2 \sim 1$ for errors approximated as $\sqrt{\mathrm{ADU}}$—generally an overestimate for error), with saturation evident in the 32 second exposure. (At $\sim 62,000$ ADU the count jumps up to the 16-bit maximum 65535.) The linear fit even looks good at minimum exposure times where systematic error in the shutter speed control is expected. Note that the y intercept is not exactly zero: with this camera zero-time exposures are designed to produce an output of about 100 ADU.
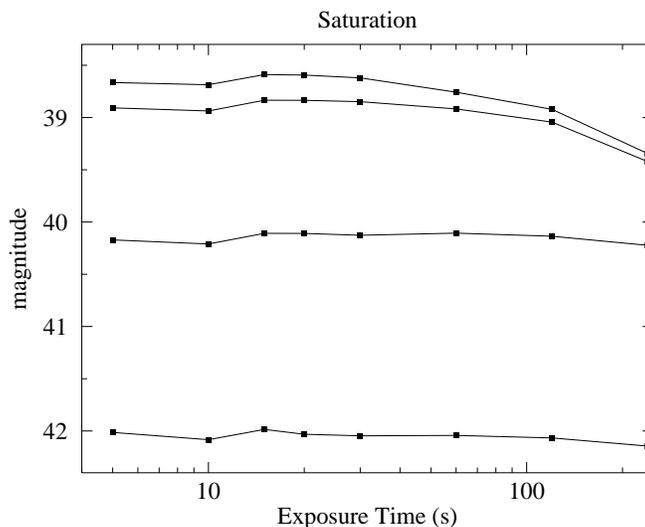
Figure 4.6: The magnitude of four stars was calculated from frames with various exposures. The brightest two stars have saturated pixels in 20 sec exposures and the dimming magnitude is evident in exposures longer than 30 sec. The third brightest star begins to be saturated at 240 sec. The fourth star is not saturated. Notice that a handful of saturated pixels will generally not result in a large magnitude error.

Flat Frame

Individual pixels have different sensitivities to light. While the effect is not large ($\sim 5\%$), it would be a leading source of uncertainty if not corrected. The solution is (in theory) simple: take a picture of a uniformly illuminated surface, and record the count in each pixel. Evidently this 'flat frame' count records how each pixel responds to one particular flux level. If in another frame a pixel records some fraction of its flat-frame count, linearity guarantees that the measured flux level is that same fraction of the flat-frame flux level. Thus to correct for varying sensitivities we just divide (pixel by pixel) the raw CCD frame by the flat frame.

In practice it is difficult to obtain a uniformly illuminated surface. In a *sky flat*, one assumes that the small section of a blue sky covered in a CCD frame is uniform. (Typically this is done after sunset so the sky is dim enough to avoid saturation in the short ($\lesssim 1$ sec) exposure, but not so dim that stars show through.) In a *dome flat* we attempt to uniformly illuminate a white surface with electric lights. Whatever the source of the flat, it is best to average several of them to average out non-uniformities in the source.

It should be clear that a flat frame records the *overall* efficiency of the system: both the varying quantum efficiency of individual pixels and the ability of the telescope to gather light and concentrate it on a particular pixel. Generally the optical efficiency of a telescope decreases for sources far from the optical axis. The off-axis limits (where the optical efficiency approaches zero) define a telescope's maximum field-of-view. The limited optical efficiency near the edge of the field of view leads to 'vignetting': the gradual fade of optical efficiency to zero at the extreme edge. In addition optical flaws, like dust on optical surfaces, are recorded in a flat frame. As a result a flat frame depends on the entire detection system (telescope, filter, chip) and will gradually change, as for example, dust accumulates on the mirror.

Bias Frame

A careful examination of the linearity plots in Figure 4.5, shows that they do not go through the origin. That is, a zero-time[27] exposure produces some counts, simply by the process of reading the CCD. The uncertainty in this bias frame (the 'read noise') represents the minimum background noise in any pixel. Measurements on our CCD shows that the difference between two bias frames shows an approximately normal distribution (aside from the flaws) with a standard deviation ~ 11 ADU. Also note that the system artificially introduces an offset of 100 into each pixel, so $\frac{2}{3}$ of the pixels in a single bias frame are expected in the range 100 ± 8 ADU. ($8 \approx 11/\sqrt{2}$)

Dark Frame

Even in the total absence of light, pixels will slowly accumulate charge. The electric current leaking into the pixels is strongly affected by temperature, hence the name 'thermal current' and it produces accumulated charge in the absence of starlight, hence the name 'dark current'. The professional solution is to cool the CCD using liquid nitrogen, which

---

[27]With the STL-1001E the minimum exposure is actually 0.12 s
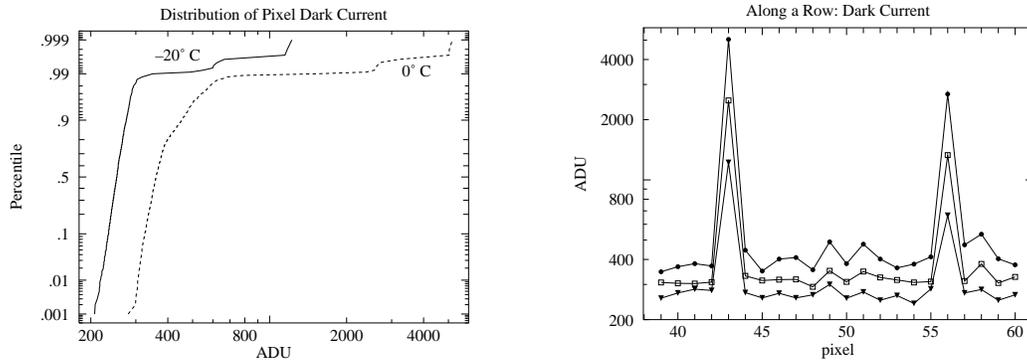
(a) The distribution of pixel dark current at $0°C$ and $-20°C$. All but a few of the pixels have a (log)normal distribution of dark current. About 1% percent of the pixels show much larger dark current: these are the 'hot' pixels. Note that both classes of pixels respond to temperature: lowering the temperature reduces every pixel's dark current.

(b) Part of a row of pixels is displayed at three temperatures: $\bullet = 0°C$, $\square = -10°C$, $\blacktriangledown = -20°C$. Isolated hot pixels are randomly sprinkled throughout the image (here at columns 43 and 56). The data here and in (a) are from 15 minute dark frames.
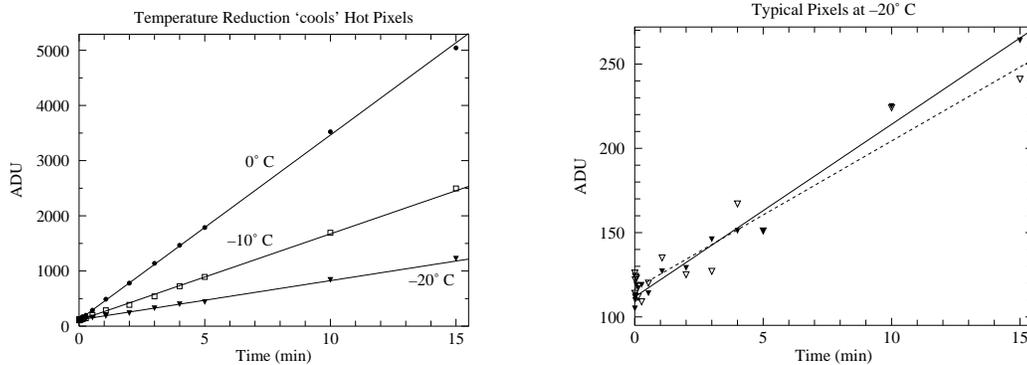
(c) Pixel counts grow with 'exposure' time as the dark current produces ever larger stored charge in a pixel; of course reducing the temperature reduces the dark current.

(d) While it is more evident in this low count data, there is always random deviation in counts. While we can subtract the average dark charge, deviations from the average produce 'noise' in our images. Reducing the temperature reduces this noise.

Figure 4.7: Dark frames are 'exposures' with the shutter closed. The source of these counts is not starlight, rather a thermally induced current allows charge to accumulate in each pixel. This 'dark current' can be exceptionally large in 'hot pixels'.

reduces the dark currents to near zero. The STL-1001E uses a less cumbersome (but less effective) solution: a thermoelectric cooler which can cool the CCD $\sim 40°$C below ambient temperature, reducing the dark currents by more than a factor of 10. The resulting dark currents are *not* negligible for exposures longer than a few minutes. A 'dark frame' is simply an 'exposure' with the shutter closed; it is a bias frame plus accumulated charge due to dark currents. A dark frame can be subtracted from an (equally timed) exposed frame to compensate for this non-photon-induced current, however variation in the dark current ('noise') results in a corresponding uncertainty in the adjusted pixel counts.

There is tremendous pixel-to-pixel variation in the dark current. About 1% of the pixels have a dark currents ~ 10 times the typical value. The 'hottest' pixels[28] will saturate in a 5 minute exposure at $0°$C. Thus a dark frame looks black except for a handful of bright ('hot') pixels; it could be mistaken for the night sky, except the 'stars' are not bright disks rather isolated, single hot pixels. Since taking a dark frame takes as much time as a real exposure, use dark frames only when required: when you are taking real data. Simply learn to disregard the sprinkling of isolated hot pixels in non data-taking situations.

Since the dark current depends on temperature, the CCD needs to achieve a stable temperature before the dark frames are taken. The CCD has an on-board thermometer, but a stable temperature at one point does not mean that the temperature as stabilized throughout the CCD. Direct measures displayed in Figure 4.8 show that it may take several hours to achieve reproducible dark frames. And the lower the requested temperature the longer it takes to achieve thermal equilibrium. For a requested temperature of $-20°$C, 1.8 hours are required for 95% (e^{-3} with $\tau \approx 0.6$ hour) of the excess counts to be eliminated. For a requested temperature of $-10°$C, 0.8 hours are required for 95% (e^{-3} with $\tau \approx 0.26$ hour) of the excess counts to be eliminated. When requesting particularly low CCD temperatures it is advisable to take dark frames at the end of the observing session or to cool down the camera well before sunset. (That is, the obvious time time to take dark frames—at dust between taking sky flats and dark-sky frames—may not be the best time to take the dark frames.)

Data Frames

The term 'object frame' here refers to a raw CCD frame of the sky as delivered by the camera-controlling software. If the object frame was not automatically dark-subtracted, there must be a matching (time & temperature) dark frame. A 'reduced frame' is an object frame that has been dark-subtracted and flat-field corrected. By adjusting for non-photon-induced charge and varying pixel sensitivity, we hope our reduced frame shows the actual distribution of photons. It should be noted that all of these corrections will be applied by making the proper requests of the CCD control software. In the end, you should retain both the fully processed reduced frame and the original object frame.

[28]For example: $(265, 742), (68, 97), (233, 866), (266, 166)$
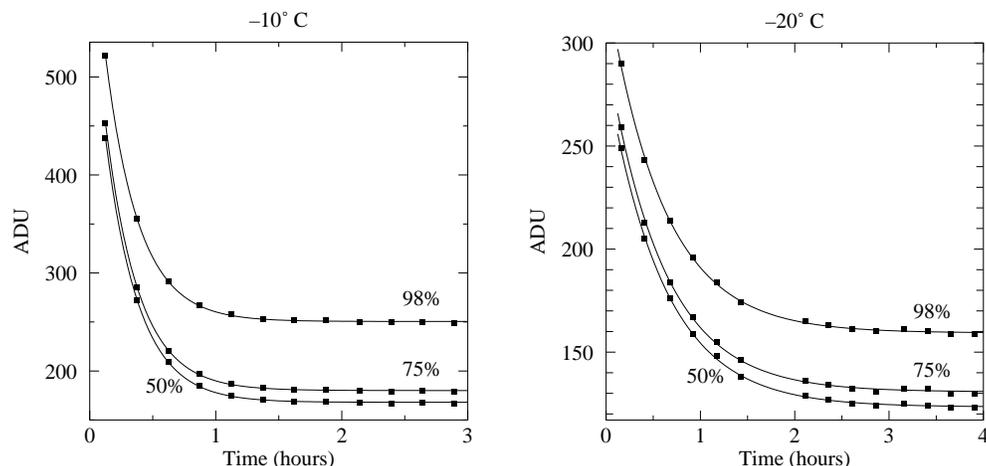
Figure 4.8: It can take hours to achieve a stable CCD temperature. The left plot displays the results of a sequence of 900 sec dark frames at an indicated (stable) temperature of $-10°$C. The count for the median (50 percentile), 75%, and 98% pixel is displayed as a function of time. In all cases the approach to equilibrium is exponential with $\tau \approx 0.26$ hours. Analogous results for $-20°$C show a much slower approach to equilibrium ($\tau \approx 0.8$ hours).

Basic Plan

Given a reduced frame, determining the number of counts/second from a star is a relatively straightforward[29] task which will be performed (with your help) by the software. This resulting rate must be converted to flux by multiplying (or dividing) by a variety of factors: collecting area of the telescope, reflectivity of the mirror, quantum efficiency of the CCD, CCD gain, ... However the main point is that there is a proportionality between this rate and the actual flux. By our theorem, this means that a magnitude calculated just using this rate (an 'instrumental magnitude') is just a constant off from the actual standardized magnitude. If we have a star in the reduced frame of known magnitude, that constant difference can be calculated just by taking the difference between the instrumental magnitude and the known magnitude. And then that calibrating constant can be applied to every star in the reduced frame, converting each instrumental magnitude to a standardized magnitude. In practice it is wise to use several calibrating stars and use the average value of magnitude difference as the calibration constant. (The standard deviation of the magnitude differences then gives us an estimate for the error.) The calculation of instrumental magnitudes (i.e., calculating $-2.5\log(\text{rate})$) is also a straightforward task best left to software. All you will need to do is calculate (probably using a spreadsheet) the magnitude differences of the calibrating stars, and then find their mean and standard deviation.

---

[29]Many important details are hereby hidden from view. A short list: (1) determining the background level to be subtracted from the star pixels (2) determining the proper aperture that encloses all the stars light or accounting for the star's light that has blended into the background (3) accounting for the star's light that is between pixels (4) dealing with pixels that are only fractionally within the aperture. In summary: background and point spread function (PSF) are the key issues.
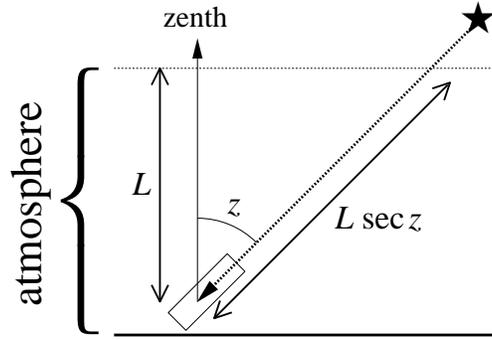
Figure 4.9: Stars that are near zenith go through the minimum amount of the Earth's atmosphere. Stars at a zenith angle z go through $\sec z$ times that minimum. This figure displays the atmosphere as a finite thickness slab, whereas in fact the atmosphere gradually thins out to nothing. Nevertheless, meter-by-meter a star at angle z must go through $\sec z$ times as much atmosphere as a star at zenith. 'Air mass' is defined as this multiplicative factor: $\sec z$.

Complications

Air Mass Corrections

Magnitudes are intended to be archivable numbers in which the particulars of the observation (e.g., telescope aperture, quantum efficiency, transparency of the atmosphere,...) have been removed. As describe above, this is relatively easy if the reduced frame contains calibrated sources. On the other hand, if we have separate reduced frames for the object and calibrated stars, we must arrange that the entire detection system (atmosphere included) acts the same (i.e., the same proportionality constant between rate and flux) for both frames. However, it is difficult to find calibrated stars that have the same angular altitude as the object; Thus one frame must be shot through a thicker slice of atmospheric than the other. Air mass corrections adjust for this path length difference and assume that the atmospheres (and optics) were otherwise identical.

The absorption of light is a multiplicative phenomenon; that is, if the first meter of the atmosphere allows fractional transmission of light T_1 and the next T_2, etc, then the intensity of light through a sequence of meters is:

$$F = T_N T_{N-1} \cdots T_2 T_1 \ F_0 = \prod_{i=1}^{N} T_i \ F_0 \qquad (4.30)$$

Since it is easier to deal with additive quantities rather than multiplicative quantities we define the optical depth τ by:

$$T \equiv e^{-\tau} \qquad (4.31)$$

so

$$F = T_N T_{N-1} \cdots T_2 T_1 \ F_0 = \exp\left(-\sum_{i=1}^{N} \tau_i\right) F_0 \qquad (4.32)$$

According the Beer-Lambert law, the absorption of light depends on atomic absorptivity (the 'cross section' σ, which generally depends on the wavelength of light), the number of atoms per volume ('number density' n), and the path length ℓ:

$$\tau = \sigma n \ell \tag{4.33}$$

We can think of the atmosphere as a sequence of filters, so the light intensity through the atmosphere is:

$$F = \exp\left(-\sum_{i=1}^{N} \sigma n \Delta \ell\right) F_0 = \exp\left(-\int \sigma n \, d\ell\right) F_0 \tag{4.34}$$

Of course the density of the atmosphere (and to some extent the composition of the atmosphere) depends on altitude. If we consider light that has taken the shortest path through the atmosphere (i.e., from a star directly overhead or at *zenith*), we call the above integral $\tau_0(\lambda)$. (The dependence of τ_0 on the wavelength of light is of course due to the dependence of σ on the wavelength of light.) As shown in Figure 4.9, if a star is at zenith angle z, the path length (and hence τ) in increased by the *air mass* factor $\sec z$. Thus:

$$F = \exp\left(-\tau_0 \, \sec z\right) F_0 \tag{4.35}$$

The magnitude that corresponds to this flux is:

$$m = 2.5 \tau_0 \log e \,\, \sec z + m_0 \tag{4.36}$$

where m is the magnitude measured at zenith angle z and m_0 is the magnitude as measured without any atmospheric absorption (i.e., the standardized magnitude). For any single star one can (assuming the atmosphere doesn't change over the few hours required) plot $(\sec z, m)$ data as the star rises, culminates and sets, and determine the y-intercept which is m_0. Alternatively for a set of stars each with known m_0 and measured m at varying $\sec z$, one can determine the correction factors A and B required to convert a star's $(\sec z, m)$ to m_0:

$$m - m_0 = B \sec z + A \tag{4.37}$$

Thus lacking in-frame calibrated stars, several calibration frames (at different $\sec z$) are required for each filter, and the reduction process becomes much more complicated.

Note that the zenith angle z can be calculated[30] from:

$$\cos z = \sin \delta \sin \varphi + \cos \delta \cos h \cos \varphi \tag{4.38}$$

where δ and h are the star's declination and hour angle and φ is the observatory's latitude. The above terms from astrometry (declination, hour angle, altitude, zenith, right ascension, . . .) are part of the everyday vocabulary of astronomers. If you are not familiar with these terms you should read the following online tutorials:

- http://www.physics.csbsju.edu/astro/CS/CSintro.html

- http://www.physics.csbsju.edu/astro/sky/sky.01.html

- http://www.physics.csbsju.edu/astro/terms.html

For example at SJU ($\varphi = 45°34.5'$) for stars on the celestial equator ($\delta = 0$), the minimum air mass (at $h = 0^h$) is $\sec z = 1.43$. The sequence of air masses: $\sec z$=1.5, 2.0, 2.5, 3.0 occurs at hour angles $h = \pm 1.18^h, \pm 2.96^h, \pm 3.68^h, \pm 4.10^h$. Figure 4.11 plots these air mass values for other declinations.

---

[30] `xephem` can do this calculation for you
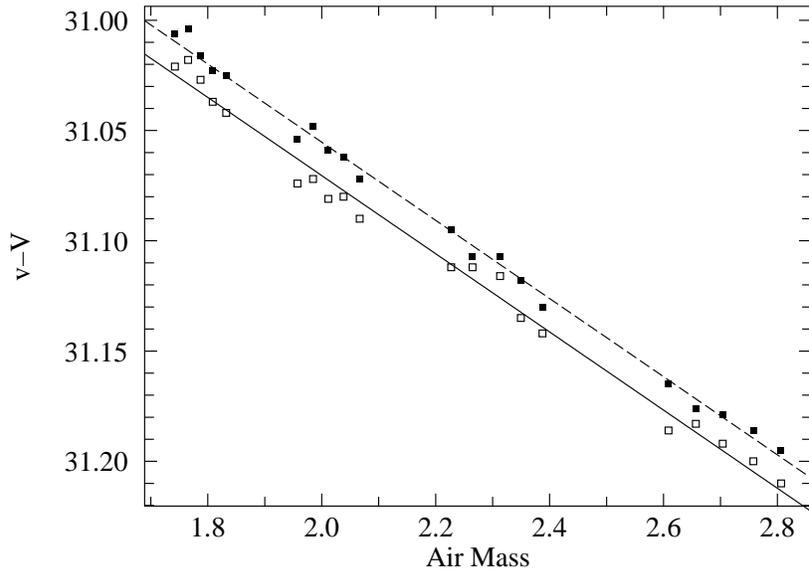
Figure 4.10: The instrumental magnitude minus the standardized magnitude $(v - V)$ for a red star ($\square$) and a blue star ($\blacksquare$) as a function of the air mass. Notice that the slopes are quite similar (i.e., $.177 \sec z$), but the intercepts are slightly different ($\Delta = 0.016(B - V)$), due to the difference in color. This color correction is explained in the following section. The rms deviation in these fits is 0.005 magnitudes which is about $\frac{1}{3}$ the intercept shift.
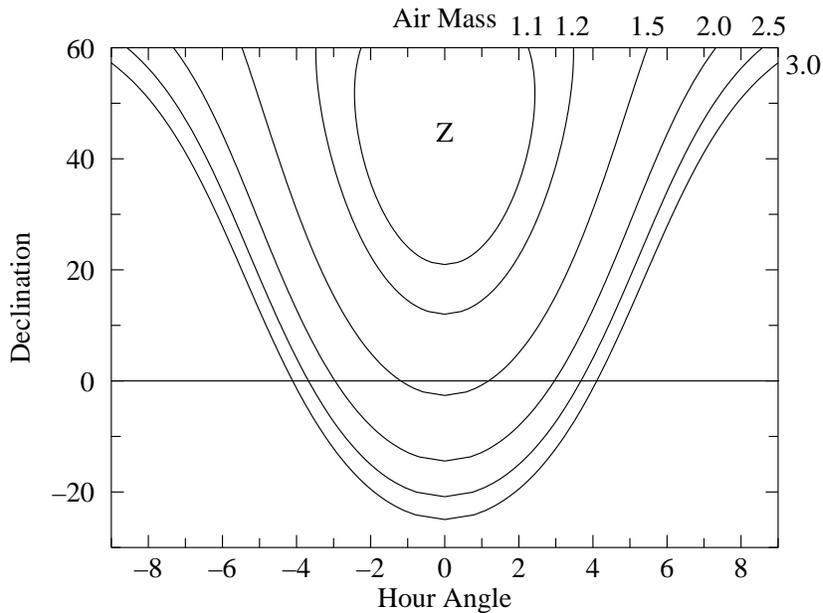
Figure 4.11: Air mass $(\sec z)$ depends on where a star is in the sky. Air mass is plotted here as a function of the star's declination and hour angle for the SJU observatory location. Z marks zenith, where the minimum (1) air mass is found.
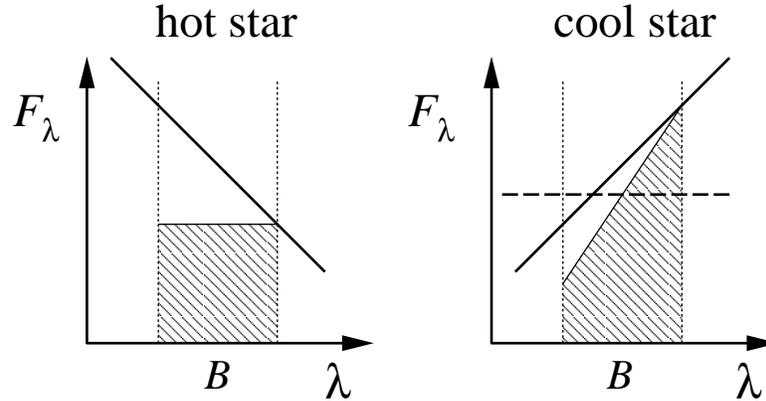
Figure 4.12: The atmosphere (in fact any optical component) acts as a colored filter on incident starlight. Typically short wavelength light is more strongly attenuated (i.e., reduced) than long wavelength light. Consider two stars that have the same B magnitude outside the atmosphere. Of course, the hot star has a larger fraction of its light in the short wavelength side of the B band. As a result more of its light will be affected and its measured magnitude will deviate from that of the cool star. In a tremendous exaggeration I've assumed above 50% absorption at the short-λ end of the B band, and 0% absorption at the long-λ end. As a result the total measured B-band flux (the shaded area) will be larger for the cool star even though both stars had the same flux outside the atmosphere. If narrow band filters are used there will be much less change in absorption over the band and so this effect is much reduced.

Color Corrections

We learned in the above section that the Earth's atmosphere acts as a non-negligible filter, attenuating (absorbing or scattering, generally reducing) starlight before it reaches our telescope. The fact that the atmosphere looks blue tells us that the atmosphere in fact acts as a colored filter, attenuating some wavelengths more than others. (That is, σ depends on λ, with short-λ (blue) light more attenuated than long-λ (red).) As discussed in Figure 4.12, this means that two stars which, outside our atmosphere have the same magnitude but different temperatures, will be measured as having different magnitudes from the ground. The upshot of this is the atmospheric optical depth parameter τ_0, depends (slightly) on the temperature, T, of the star:

$$B = 2.5 \, \tau_0(T) \, \log e \, \sec z + B_0 \qquad (4.39)$$

where B is the magnitude through the B filter at ground level and B_0 is the atmosphere-free B magnitude. The temperature of the star is related to the actual $B_0 - V_0$ color index, which in turn is related to the instrumental color index $b - v$. If we make a first order Taylor expansion of $\tau_0(T)$ in terms of $b - v$, we have (after lumping together all the unknown constants and assuming all the stars in the frame have the same z)

$$b \approx \alpha_B + \beta_B(b - v) + B_0 \qquad (4.40)$$

That is the offset between the instrumental magnitude and the atmosphere-free magnitude is not exactly a constant; instead we expect a slight dependence on the color of the star. A
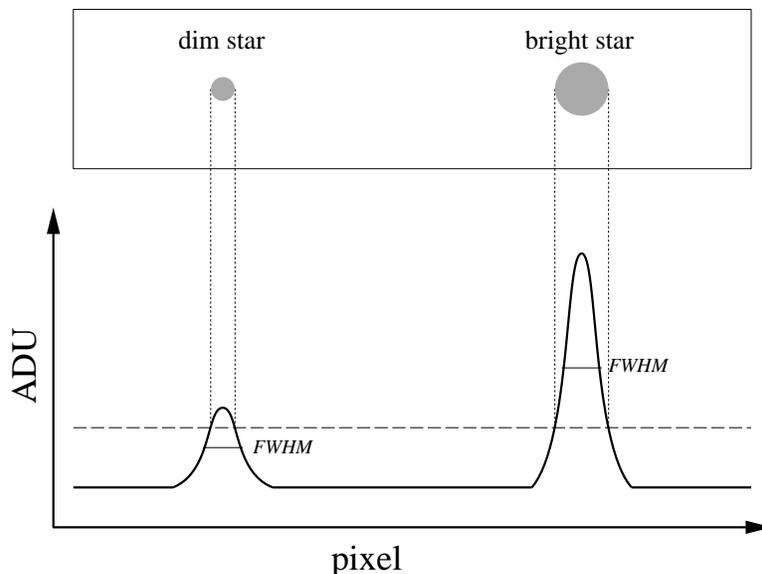
Figure 4.13: Every photon detected in a CCD frame has traversed (essentially) the same patch of atmosphere, so each star's image suffers the same distortion. Every stellar 'image' (really point spread function or PSF) has the same shape, only the total flux in each image varies (with the star's magnitude). However, a brief inspection of an image will show clearly larger spots for brighter stars. As shown above this is simply a matter of the level chosen as bright.

linear fit (using the data from the calibrated sources) can then be used to determine α and β, viz:

$$b - B_0 = \alpha_B + \beta_B(b - v) \tag{4.41}$$

and similarly for the V magnitudes:

$$v - V_0 = \alpha_V + \beta_V(b - v) \tag{4.42}$$

Subtracting these two equations gives us the relationship between the instrumental color index and the atmosphere-free color index:

$$(b - v) - (B_0 - V_0) = \alpha_B - \alpha_V + (\beta_B - \beta_V)(b - v) \tag{4.43}$$
$$[1 + (\beta_V - \beta_B)](b - v) + (\alpha_V - \alpha_B) = B_0 - V_0 \tag{4.44}$$

Thus we expect a linear relationship (with a nearly unit slope) between the instrumental color index and the atmosphere-free color index. The parameters of this linear relationship can be determined by a fit using the data from the calibrated stars.

Aperture Corrections

When you look at your CCD frames, you will not believe my statement that all the star images (really PSF) have the same size. However, both bright stars and dim stars are far enough away that their angular diameters are effectively zero (less than a thousandth of a single pixel). The photons from each star shoot through the Earth's atmosphere and are

deflected this way and that producing a multi-pixel disk in the frame. This is called atmospheric *seeing* (particularly for image size) or *scintillation* (particularly for image brightness). But since every photon is shot (basically) through the same patch of atmosphere, these random deflections must have the same distribution. The difference between a dim star and a bright star is simply the total number of photons in the images. (That is a long time exposure of a dim star should look exactly[31] like a short exposure of a bright star if the two cases produce the same number of photons.) To see that the distributions are identical you should look at the full width at half maximum (FWHM) of the PSF: First determine the peak count in an image, and then find the circle at which the count is half of this maximum. The diameter of the circle (typically a few arcsec) is the FWHM.

Because of the obvious difference in the image size, when totaling the counts in a stellar image, one is tempted to expand or reduce the circle of pixels counted as part of a star to match the apparent size of that star. *Do not do this!* Instead keep a constant aperture size for all frames. It may appear that counts will be missed as bright stars overflow the aperture, whereas dim stars have a too-large aperture. However recall that all stars actually have the same extent. Furthermore our theorem states that if we capture a consistent fraction of the counts in every star, all our instrumental magnitudes will just be a constant off from the standardized magnitudes. So a *consistent* fraction of the counts may be combined with all the other proportionality constants in the final difference between instrumental magnitude and standardized magnitude. Now there may be occasions where we must adjust the *aperture* used for magnitude calculation (for example to exclude a neighboring star), however then inconsistencies (and hence errors) are then being generated. The disadvantage of using a constant aperture is relatively more noise-prone 'background' pixels will be included in dim stars. There are a variety of solutions to this problems (generally under the heading of PSF fitting), however they are beyond the aims of this lab.

The blurring of star images (i.e., PSF) depends both on atmospheric seeing and the adjustment of the telescope. Clearly an out-of-focus telescope (see below) will distribute the same photons in a larger radius disk. It should also be noted that the telescope's 'optical aberrations' result in additional blurring ('coma') near the edge of the telescope's field of view.

Focus

The aim of focus is to achieve that smallest possible FWHM, so each star's light is concentrated in the fewest possible number of pixels. Same number of photons in fewer pixels, means a higher count in those central pixels. Since every pixel comes with a certain background 'noise', fewer pixels means less noise in the stellar magnitude. You should plan on spending a good bit of time (maybe a half hour) trying to achieve the best possible focus. There are two approaches to monitoring improved focus: (1) monitor the peak count in a bright star (~ 4 mag), (2) watch for the appearance of numerous dim stars as their peak gets above the 'bright' level. (You must learn to ignore the apparent size of bright stars: it is a poor measure of good focus.) A complicating factor is that both (1) and (2) vary independently of focus adjustments since atmospheric seeing varies from second to second, and since photon counts should be expected to vary $\sim \sqrt{N}$.

---

[31]We are assuming, of course, that the atmosphere acts the same during the long exposure as it does during the short exposure. No surprise: in fact the atmosphere is not time-invariant.

Software

xephem

`xephem` is planetarium software designed to display various types of night-sky maps in advance of observing. You will also use it to tabulate astrometry data relevant to your observations. The file `using.xephem.txt` describes how to use this program to produce the tables and maps required for the pre-observation phase of the lab. In addition you can use it for astrometry and photometry on `.fit` images (although I recommend `gaia` below for these reductions).

Aladin

`Aladin` is a front-end to various astronomy databases on the internet. You will use it to find images to use as finder maps and to identify (and obtain data on) the stars in those images. The Aladin server and SkyView+ are often a good sources of images; Simbad[32] is the usual source for stellar information. Sometimes just information from the SAO catalog is required; then VizieR[33] can be used to access just the SAO catalog. You will use `Aladin` mostly in the pre-observation phase. The file `using.Aladin.txt` briefly describes how to use this program to produce the maps and data required for the pre-observation phase of the lab. Note: NASA has a very handy web page[34] that also allows you to do many of these things. `gaia` (see below) is also and excellent way to make finder maps.

CCDops

CCDOps is software used to control the camera and create/save CCD frames. You will be using a small fraction of its capabilities. Typical commands will be: to set the temperature (cooling) of the CCD, focus, grab images (set the exposure time and perhaps take a dark frame), examine images (histogram, crosshair), save images (in both `.SBIG` and `.fit` formats), and reduce images by dark-subtracting, flat-fielding or averaging.

gaia

`gaia` is the recommended reduced-frame analysis program. In order to use `gaia` your images must be in either `.fit` or ndf (`.sdf`) format. You can convert your `.SBIG` frame into a ndf frame (`.sdf`) using the program `sbig2ndf` or the many-files script `batch_ndf`, but it is easier to directly save your images in `.fit` format. The file `using.gaia.txt` briefly describes how to use this program.

In addition to reporting a star's magnitude, `gaia` will report the error in that magnitude. This error is based on counting (Poisson statistics, $\sqrt{N}$). These errors will be smaller for brighter stars; a typical value might be ~ 0.003. Record one typical `gaia`-reported error

---

[32]`http://simbad.u-strasbg.fr/`

[33]`http://vizier.u-strasbg.fr/`

[34]`http://skyview.gsfc.nasa.gov/`: start with 1000×1000 pixel Digitized Sky Survey image with 0.5° Image Size, B-W Log Inverse Color Table, and a Grid

for your stars. Because of systematic errors, you will find the actual error is much larger ~ 0.02.

IRAF

IRAF is the Image Reduction and Analysis Facility, a general purpose software system for the reduction and analysis of astronomical data. It is the standard for professional astronomers: designed for the expert user—with no compromises for the beginner. While its use is encouraged for those thinking of becoming astronomers, I cannot recommend it for this lab.

Planning to Observe

Minnesota weather is is unpredictable[35] and we need an extraordinary, no-visible-clouds night. We must be prepared to use any/every of these rare nights on short notice. Of course, you have other commitments (your concert, the night before your math/physics exam, the St. Thomas game...), the aim here is to communicate those constraints to me and to your lab partners in well in advance. Thus you will need to carefully record your schedule a lunar month ahead. Note that Murphy's Law guarantees that if you forget to X-out some important night (your girlfriend's birthday, the day before your term paper is due, a 'free' day), that night will be clear, and somebody (I hope not me!) is going to be disappointed. My family commitments and intro astro labs are going X-out all Sunday nights and many Monday and Wednesday nights.

Before each full moon, mark on the class calendar (in the unix lab) the dates you and your lab partner(s) are committed to observe.

Since clear nights are precious, 'go' days need a well-thought-out plan. If you're following the standard project, your plan involves a target whose frame will include several stars with known[36] B, V, R, I magnitudes. Thus you need to: (A) Find a target that has the required stars. (Stars with recorded R magnitudes are relatively rare.) (B) Know where those standard stars are in your target (so you can aim so your CCD frames to include the stars you need). (C) Know where your target is among the neighboring stars (so if your first try at finding the target fails, you have a plan [not 'hunt-and-peck'] for getting to the right spot). (D) And of course, know that your target is well above the horizon when you are planning to observe. While the standard project does not require additional standard stars, you may want you to record alternative (out-of-frame) standards.

Thus, you must produce (and turn in) the pre-observation documents described below. Using the following *new* moon as a date, use `xephem`, to select your targets. The problem is to find a star cluster that has the required calibrated stars. For example, in early September `xephem` makes M39 or NGC 6871 obvious choices. However, I'm unable to find any

---

[35]`http://www.physics.csbsju.edu/astro/` has a link designed to report (guess) sky conditions at the SJU observatory. It cannot be relied on, but I believe it's the best available information. Note: black is good for this clear sky clock.

[36]Professionally, only very carefully measured 'standard' stars—for example those measured by Landolt or Stetson—would be used for calibration. For the purposes of this lab you may use any star that has a 'book' magnitude. It is not at all unusual for such literature magnitudes to have 10% errors

calibrating R magnitudes in those clusters. (Of course, correcting for air mass and using calibrated sources outside of the cluster is an option.) IC 4665 is often the default choice for early fall.

Pre-observation Checklist

From the class web site find `photometry/observatoryCCD3.txt`. Print out a copy. You will cut out and tape relevant sections of this file as part of your lab notebook Procedure.

The following should be prepared before each full moon (12-Sep-2011, 11-Oct-2011, 10-Nov-2011):

1. Sign up for 'go' days during the following lunar month.

2. Using `xephem` set for the date of the following new moon print out the following:

   (a) A Sky View showing the entire sky with the labeled location of the target.
   (b) A `.xephem/datatbl.txt` file recording basic data (RA, dec, altitude, air mass, . . .) for the target at the suggested observation time.
   (c) A copy&pasted list of basic data on the stars (e.g., from SAO_LX200.edb) used to locate the target.At least one nearby star should be a 'focus star' with mag $\sim$4–5. The file `using.xephem.txt` describes what data you need to have recorded.

3. Using `Aladin` (or web-based `skyview.gsfc.nasa.gov`) print out the following finding charts for your target cluster:

   (a) a low magnification (degree scale) image showing both a finder LX200 star and the desired CCD frame.
   (b) an image scaled to about the CCD frame size (0.5°) showing exactly the desired CCD frame. (Since the target CCD frame must include calibrated stars, clearly you must know where the calibrated stars are in the target.) Record the RA and dec for your target frame(s).

Learning to effectively use these pre-observation programs will be a bit of a challenge and take several hours, so do not procrastinate and feel free to seek my help. In addition to the usual problems of learning new software, you will need to learn how to make the web-based astronomy databases find the information you need. However, if you've failed to make these pre-observation finding charts, you can probably still make observations using one of the perennial targets (IC 4665 in the fall, M 67 in the spring).

Observation Evening

In order to take data the sky needs to be 90% cloud free: we must be able to aim the telescope at a star and not have a cloud go through that spot for an hour. Years of experience shows it is not possible to predict, even an hour before sunset, what the sky will be like at sunset. Thus the plan is for you to look at the sky 45–30 minutes before sunset and decide if the sky is suitable. Kirkman will do the same, and if we both reach the same

conclusion (not a sure outcome!) we will both arrive at the observatory 30 minutes before sunset. Since we need to take flats during sunset, if all parties don't arrive well before sunset, the observing session is canceled. Plan to be at the observatory for 2–3 hours (no eating in the observatory; no bathroom in the observatory). Perhaps an hour of that period (after flats, waiting for the sky to darken) will involve really monotonous tasks: feel free to bring homework or entertainment to the observatory to occupy that period. If you're not sure where the observatory is check out the links at `www.physics.csbsju.edu/astro`. Also on that page, check out the "Sky weather prediction for SJU Observatory". This link to the "Clear Sky Chart" has proven to be the best of many poor predictors of sky conditions. Look for the raw data frames to appear on the linux filesystem in the directiory `/ccd/`*YYMMDD*`lab`, e.g., pictures taken during the evening of 4th of July would be in `/ccd/120704lab`.

Example 1: Plan for Lunar Month starting 21-July-2005

Immediately on starting `xephem`, Set the location: SJU Observatory, click on the Calendar for the following <u>new</u> moon ("NM", in this case 4-August-2005) and enter a Local Time of 21:00 (9 P.M.). Hit Update so `xephem` is ready to display the sky at that time. View→Sky View then displays the night sky for that location/date/time. (Print a copy of this sky view as described in `using.xephem.txt` and record: UTC Time: 2:00, Sidereal: 16:37.) IC 4665 is the obvious target; its RA (listed in Table 4.2) is most nearly equal to the sidereal time and hence it is near the meridian, However, with a diameter of 70′, IC 4665 is much larger than the ST7 CCD frame, so only a small fraction of it will be imaged—a fraction that must include several calibrated stars. A visit to the IC 4665 cluster web page[37] finds a couple of dozen of VRIc observations. I copy and paste that data into a spreadsheet, and then seek corresponding UBV CCD data. The result is 12 stars with the full set of $BVRI$ data. Next we need to find a CCD frame that will include as many of these stars as possible. (It is of course possible to use a different set of stars to calibrate different filters, but it will be easier if one set of stars will serve for all calibrations.) Surprisingly, most of the 12 stars have SAO[38] identifications; only three stars must be located by RA and dec. `Aladin` can find a low magnification ($1.5° \times 1.5°$) image of IC 4665. (Print this image as a low magnification finder chart.) A request to VizieR for the SAO catalog objects near IC 4665 allows me to locate the calibrated sources found above. Since IC 4665 is much larger than our CCD frame, we aim for the largest possible subset of these stars. A CCD frame centered near SAO 122742 will pick up four fully calibrated stars (73, 82, 83, 89). Simbad gives $B-V$ data for several other bright stars in this frame (e.g., TYC 424-75-1, Cl* IC4665 P44), but I don't find additional $R-I$ data in Simbad. Returning to `obswww.unige.ch`, I find (depending on the exact position of the CCD frame) it may be possible to include $R-I$ calibrated stars 67, 76, 84, & 90. You should decide exactly how to place the CCD frame and locate the calibrated stars on a finder chart that is about $2\times$ the size of the CCD frame. For focus stars consider SAO 122671 (mag=3) and SAO 123005 (mag=5). Note that SAO 122723 would make a good (bright: mag=6.8) star to aim the LX200 near this object. (I call such aiming/finding stars 'LX200 stars'. The LX200 object library includes all SAO stars brighter than magnitude 7, but any bright star can help assure the telescope is aimed at what you intend (and that the telescope's reported RA/dec are accurate). You can load the SAO_LX200.edb database into `xephem` to display the LX200 stars; SAO_mag75.edb includes

---

[37]`http://obswww.unige.ch/webda/`
[38]Smithsonian Astrophysical Observatory—a standard catalog of 'bright' stars.

about twice as many stars (down to magnitude 7.5); SAO_full.edb includes about a quarter of a million stars (down below magnitude 9).

In such a large cluster, it would be wise to select additional fields, for example one centered near SAO 122709. A CCD frame there might include $R-I$ calibrated stars: 39, 40, 43, 44, 49, 50, 58, 62.

To select alternative standards... WIYN has a web page[39] with recommended standard star regions. (WIYN's stars are designed for a large telescope, and hence would require long exposures on our telescope). Peter Stetson's extensive list of standards is also online[40]. The `xephem` databases LONEOS.edb, Landolt83.edb and Buil91.edb contain shorter lists of brighter stars: < 13mag, ~ 9.5mag and < 8mag respectively.) Generally these good standard stars will be well-separated from the target, so air mass air mass correction would be required. (Since I have internal 'calibrated' sources, I'm not required to take CCD frames of these standards, however I've decided to 'be prepared' and hence have recorded the basic data I would need to observe them.) I select: #109 (PG1633+099, at RA=16:35, LX200 star: SAO 121671), #121 (110 506, at RA=18:43, LX200 star: SAO 142582), and #125 (111 1969, at RA=19:37, LX200 star: SAO 143678, also see: SAO 124878). These standard stars can be marked as `xephem` Favorites from the UBVRI.edb database. These sources should span a good range of air mass in the general direction of the target. Following the instructions in `using.xephem.txt`, I create a file of basic location data for my target and standard fields. By default, this file is: `.xephem/datatbl.txt` and can be printed (`% lp` *filename*) or edited (`% kwrite` *filename*).

Example 2: Plan for Lunar Month starting 19-August-2005

After Setting `xephem` for location: SJU Observatory, the new moon date 3-September-2005, and Local Time 21:00 (9 P.M.), I find and record: UTC Time: 2:00, Sidereal: 18:35. NGC 6633, whose RA is most nearly equal to the sidereal time and hence is near the meridian, is the obvious target. After adding NGC 6633 to `xephem` Favorites and loading the SAO database (Data→Files; Files→SAO_mag75.edb), it's easy to zoom in (lhs scroll bar) on NGC 6633 and find the nearby bright (5.7 mag) star SAO 123516[41], which would be a good star to steer the LX200. For focus stars consider SAO 122671 (mag=3) and SAO 123377 (mag=5). A visit to the NGC 6633 cluster web page[42] finds six VRIe observations for R. Cross-reference shows that all six are ~ 8 mag SAO stars[43], however only two (50 & 70) are in the central region of the cluster. (They could also be found using SAO_full.edb.) Stetson reports BVI data for NGC 6633, however the brightest ten of his stars are ~ 13 mag, which is $100\times$ dimmer than the ~ 8 mag R-mag stars. Thus different frames are required to properly expose the Stetson standards and the R 'standards'.

Using the Buil91.edb database, I can find bright standard stars: SAO 085402, SAO 141956, SAO 145050. These bright SAO stars are themselves LX200 stars.

---

[39] `http://www.noao.edu/wiyn/obsprog/images/tableB.html`

[40] `http://cadcwww.hia.nrc.ca/standards`

[41] VizieR reports $R = 5.703$, for this star, which may serve as an additional standard. FYI: this star is also known as: HD 170200 and HR 6928.

[42] Again: `http://obswww.unige.ch/webda/`

[43] These dim SAO stars are not in SAO_mag75.edb so I located them using `Aladin`

Example 3: Plan for Lunar Month starting 21-June-2005

After Setting `xephem` for location: SJU Observatory, the new moon date 6-July-2005, and Local Time 22:00 (10 P.M. — it's not dark at 9 P.M.), I find and record: UTC Time: 3:00, Sidereal: 15:43. Because of its larger declination, Upgren 1 has a bit less air mass than NGC 6633, so it becomes the target. Upgren 1 is not in `xephem`'s databases, so it must be added to Favorites following the procedure recorded in `using.xephem.txt`. A visit to the Upgren 1 cluster web page[44] finds seven stars with VRIc and UBV observations. However the color index disagreements are of order 0.05, so we can use this as an opportunity to find the correct values. Using the Landolt83.edb database, I find three neighboring standard stars with a range of colors: HD 102056 (LX200 star: SAO 81968, bluer than the Upgren 1 stars), HD 106542 (LX200 star: SAO 100009, redder than the Upgren 1 stars), and HD 107146 (LX200 star: SAO 100038, similar to the Upgren 1 stars). Using the Oja96.edb database, I find two very close stars with similar colors: BD+35 2356 and BD+34 2338. (For these dimmer, $\sim$10 magnitude, stars finder charts are required.) $*\#133$ (SAO 63257, with mag 5 partner 63256) is a neighboring mag 3 star, SAO 44230 is a mag 4 star — both can help focus and alignment. SAO 63118 is the nearest LX200 star to Upgren 1 (since Upgren 1 is not in the LX200 catalog of clusters, the final jump must be made based on RA/dec). All but one of the Upgren 1 seven stars is an SAO star; they are all easy to identify with `Aladin`. (FYI: spiral galaxies M94 and M63 might be worth a look.)

Target and standards were fairly close together and near the meridian during the measurements, with airmass varying from 1.01 to 1.14. This was not a sufficient range to detect airmass correction terms. Increasing airmass dims and reddens stars. As a result calibrated stars viewed through a larger air mass will have a smaller constant in the color calibration equation Eq. 4.44. In this case, stars at airmass 1.14 were used to determine the color calibration equation which was then applied to unknown stars at airmass 1. The result is a bias in the estimated color of the unknown stars: their actual $B - V$ is likely to be bigger than that calculated using the calibration equation (i.e., color calibration constant should be larger at airmass 1). The effect can probably be ignored for Δairmass$\sim$0.1; Adjusting for this effect (e.g., determining[45] the term B in Eq. 4.37) typically requires measuring stars with Δairmass$\sim$1.

Sample Data

The large open cluster IC 4665 was the target for two nights during August 2010. The results are reported in Figure 4.14.

Report Checklist

1. Write an introductory paragraph describing the basic physics behind this experiment. For example, why does smaller (even negative) $B - V$ correspond to higher temperature? Why do we expect $v - V$ to be approximately constant? (This manual has

---

[44] Again: `http://obswww.unige.ch/webda/`

[45] On a following night, stars from IC 4665 were used to provide high airmass data. The results were: $B - V = 1.30(b - v) - .59 - .09\sec z$, $R - I = .85(r - i) + .84 - .06\sec z$

Names	Constellation	Right Ascension (2000)	Declination (2000)	Diameter (arcmin)	Reddening (mag)
M34, NGC 1039	Per	$02^h42^m05^s$	$+42°45'42"$	25	0.07
IC 348	Per	03 44 30	+32 17 00	10	0.93
Collinder 69	Ori	05 35 06	+09 56 00	70	0.10
M35, NGC 2168	Gem	06 09 00	+24 21 00	25	0.26
NGC 2264, cone	Mon	06 40 58	+09 53 42	40	0.05
NGC 2301	Mon	06 51 45	+00 27 36	15	0.03
NGC 2420	Gem	07 38 24	+21 34 49	10	0.05
M67, NGC 2682	Cnc	08 51 18	+11 48 00	25	0.06
Upgren 1 (asterism)	CVn	12 35 00	+36 18 00	15	
M5, NGC 5904	Ser	15 18 33	+02 04 58	23	0.03
IC 4665	Oph	17 46 18	+05 43 00	70	0.17
NGC 6633	Oph	18 27 15	+06 30 30	20	0.18
NGC 6738 (asterism)	Aql	19 01 21	+11 36 54	30	
NGC 6823	Vul	19 43 12	+23 18 03	12	0.86
NGC 6940	Vul	20 34 32	+28 15 08	31	0.25
M52, NGC 7654	Cas	23 24 33	+61 36 58	13	0.65
NGC 7790	Cas	23 58 24	+61 12 30	17	0.53

Table 4.2: Candidate targets. `http://obswww.unige.ch/webda/` and *Allen's Astrophysical Quantities* p. 548 have lists of clusters. An ideal candidate would be within telescope range (Dec: $-5° < \delta$), a size that matches our camera ($\sim 15'$), have little reddening, and contain calibrated sources.

Label	Name	SpType	$b-v$	$r-i$	$B-V$	$R-I$	T_B (K)	T_R (K)
			16 August 2010					
A	TYC 424-648-1	G8	1.260	0.198	1.278	0.600	4054	4506
B	IC 4665 93	G5	0.922	0.024	0.858	0.447	5083	5089
C	TYC 428-1938-1	A	0.644	−0.151	0.505	0.305	6218	5894
D	TYC 428-691-1	A3V	0.518	−0.258	0.349	0.210	6900	6680
			17 August 2010					
A	TYC 424-648-1	G8	1.273	0.256	1.252	0.642	4112	4379
B	IC 4665 93	G5	0.951	0.007	0.844	0.425	5122	5189
C	TYC 428-1938-1	A	0.719	−0.165	0.553	0.258	6038	6252
D	TYC 428-691-1	A3V	0.571	−0.189	0.361	0.237	6839	6427

Table 4.3: The instrumental color indices: $b-v$ and $r-i$ of the 'unknown' stars in IC 4665 were measured using `gaia` and converted to standardized color indices: $B-V$ and $R-I$ using the calibration lines displayed in in Figures 4.14c and 4.14d. Eq. 4.28 (T_B) and Eq. 4.29 (T_R) were used to calculate the temperature of the stars from each color index. The night-to-night variation in the color indices is about .035 mag; almost certainly this variation is a measurement error not a real change in the stars. Note that this error is larger than that suggested by the statistical error reported by `gaia` (~ 0.005 mag) in the standard stars. (The dimmer 'unknown' stars had statistical errors ~ 0.01 mag.) The two temperatures T_B and T_R do not exactly agree; as discussed in the text systematic errors of a few percent are to be expected.
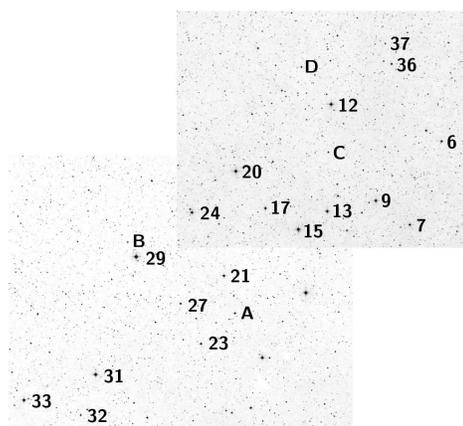
(a) This Digitized Sky Survey image shows the central $\sim 50' \times 45'$ region of the open cluster IC 4665. The standard stars of J.W. Menzies and F. Marang (MNRAS, 282, 313–316 (1996)) are labeled along with three 'unknown' stars: A, B, C, D.
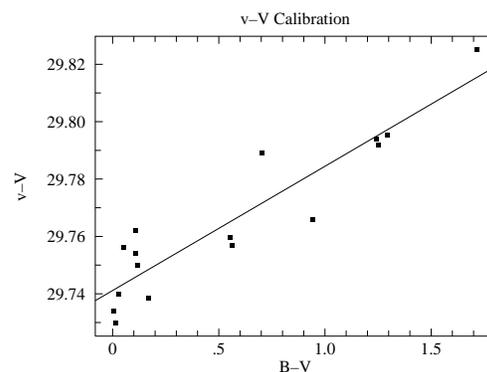
(b) The total color correction in this data set is approximately 0.1 mag. The deviation about this color corrections curve is about 0.01 mag. See Eq. 4.42.
$v - V = 29.74 + 0.04(B - V)$.
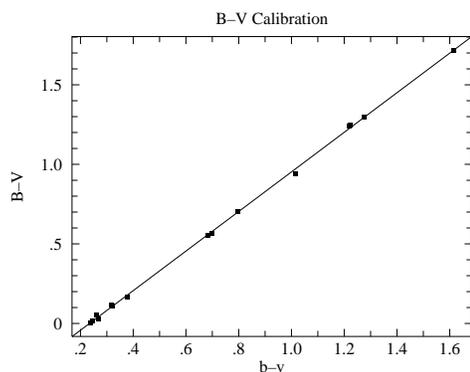
(c) Calibration of the standardize color index $B - V$ in terms of the instrumental color index $b - v$. See Eq. 4.44. The rms deviation from this calibration is about 0.014 mag.
$B - V = -0.29 + 1.24(b - v)$
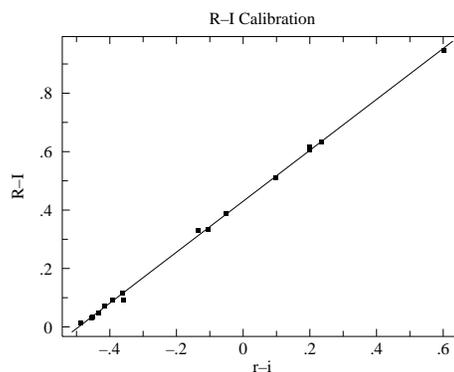
(d) Calibration of the standardize color index $R - I$ in terms of the instrumental color index $r - i$. The rms deviation from this calibration is about 0.01 mag.
$R - I = 0.43 + 0.87(r - i)$

Figure 4.14: Two frames were used to image the central region of the open cluster IC 4665 on 16-Aug-2010. This region includes 16 standard stars which were used to calibrate the relationships between standardized color indices and instrumental color indices. Using these relationships and the measured instrumental color indices, we can determine the standardized color indices of any star in the imaged region. The area was re-imaged the following night producing two sets of results for the unknown stars as reported in Table 4.3. Note: the reported 'rms deviation' is the square root of $WAPP^+$'s 'reduced chi-square' since neither x or y errors were entered.

many pages on these topics; your job is condense this into a few sentences and no equations.)

2. Files containing the reduced frames and object frames for your cluster in FIT format. (At least one for each filter.) Leave these in your linux account, but identify the filenames (and folder) in your notebook. While basic frame characteristics (filter, exposure time, date/time of exposure, ...) are contained in each file, you should also record this information in your notebook. (`/ccd/fitheader2` will display this information from the `.FIT` files in a format that allows easy copy & paste.)

3. Files containing the flat-field frames you used to flat-field your images. (One for each filter, but perhaps derived as an average of several frames.) Leave these in your linux account, but record the filenames and characteristics in your notebook.

4. Hardcopy finder charts, with your target, calibrated stars and unknown stars labeled. (Similar to Figure 4.14a.) Record an RA/dec grid on your finder charts.

5. `xephem` datafile containing basic data (RA, Dec, altitude, air mass for each observation night). Print this out and tape in your lab notebook.

6. Table containing the basic data on your calibrated stars, including standard names and magnitudes. (Perhaps included in the below file.)

7. Spreadsheet file, reasonably well documented, containing known and instrumental magnitudes. For each filter, calculate instrumental−standard magnitudes (e.g., $v-V$) and record the basic statistics (mean, standard deviation). (Note that aside from a smallish color correction, we expect $v - V$ to be constant.) You should use this spreadsheet to calculate instrumental color indices and prepare the data for the following plots. Self document these calculations before you print out the data table for inclusion in your notebook.

8. Hardcopy plot of $B - V$ vs. instrumental $b - v$ with a fitted line. (Similar to Figure 4.14c.) (As usual: include a hardcopy fit report for each fit!)

9. Hardcopy plot of $R - I$ vs. instrumental $r - i$ with a fitted line. (Similar to Figure 4.14d.)

10. Hardcopy plot of $v - V$ vs. known $B - V$ with a fitted line. (Similar to Figure 4.14b.)

11. Calculation of the temperature (T_B, see Eq. 4.28) of several unknown stars based on $b - v$. (See Table 4.3.) The value of $B - V$ for the unknown stars will be based the relationship you calibrated between standard stars' $B - V$ and $b - v$ (see #8 above). Given an error for $B - V$, the 'high-low' method will allow you to propagate that $B - V$ error into a T_B error. What should you use for the error in $B - V$? Note that neither $B - V$ nor $b - v$ errors were entered when you fit your calibrating line, so computer-reported errors are some sort of nonsense. The error in $B - V$ for the unknown stars can be estimated from the typical deviation-from-fit observed for the standard stars. These deviations (deviation from fit: $\Delta y_i = y_i - (Bx_i + A)$) are displayed by $WAPP^+$. Alternatively, when no errors are supplied to $WAPP^+$, the reduced chi-square it reports is simply the average deviation squared:

$$\text{reduced } \chi^2 = \frac{1}{N-2} \sum (\Delta y_i)^2 \qquad (4.45)$$

Thus the square root of the reduced chi-square immediately gives you something like the average (rms) deviation.

12. Calculation of the temperature (T_R, see Eq. 4.29) with error of those same stars based on $r - i$.

13. A properly formatted (units, errors, sig figs) final results table displaying the above temperatures with errors. Are T_B and T_R consistent? Comments on the uncertainty in temperature. What do you believe are the major sources of uncertainty?

Other Projects

The lab is aimed at one fairly simple project: measuring stellar magnitudes using in-frame calibrated stars. However, the techniques discussed here can be applied to a variety of other projects, some of which are listed below.

1. Imaging of extended objects (galaxies, nebulae, planets,...)

2. Orbit determination of Solar System objects (moons, comets, asteroids,...)

3. Magnitude as a function of time

   (a) Pulsing stars: Delta Scuti (`http://www.astro.univie.ac.at/~dsn/`), RR Lyrae, Cepheids
   (b) Cataclysmic variables (`http://cba.phys.columbia.edu/`)
   (c) Eclipsing binaries (`http://www.rollinghillsobs.org/`)
   (d) Novae and supernovae
   (e) Rotation period of an asteroid
   (f) American Association of Variable Star Observers (`www.aavso.org`) projects

4. Careful determination of any significant parameter of the telescope system (e.g., S/N issues, atmospheric extinction, color corrections, tracking, ...)

References

1. Santa Barbara Instrument Group (`www.sbig.com`), *Operating Manual* for STL-1001E, `STL_Operating_Manual.pdf`

2. Santa Barbara Instrument Group (SBIG), *Users Guide* for CCDOps Version 5 `CCDops_v5.pdf`

3. Meade (`www.meade.com`), *Instruction Manual* for 12" LX200 `LX200_Classic_Manual.pdf`

4. Howell, Steve B., *Handbook of CCD Astronomy*, Cambridge, 2000

5. Howell, Steve B., *Astronomical CCD Observing and Reduction Techniques*, ASP Vol. 23 1992

6. Buil, Christian, *CCD Astronomy*, Willmann-Bell, 1991

7. Romanishin, W. *An Introduction to Astronomical Photometry Using CCDs*, 2006
    `http://observatory.ou.edu`

8. J. Palmer & A.C. Davenhall [starlink], *The CCD Photometric Calibration Cookbook*,
    `starlink_ccd.pdf`

9. ed. Cox, Arthur N., *Allen's Astrophysical Quantities*, AIP/Springer, 2000

10. American Association of Variable Star Observers (`www.aavso.org`), *CCD Manual*
    `aavso_ccdmanual.pdf`

Appendix: Available Sloan Filters: $g'r'i'$

The Sloan Digital Sky Survey is an ambitious project to deeply image and reduce a large fraction of the sky. The imaging is done in 5 new filters[46] $u'g'r'i'z'$; our camera has available the central filters: $g'r'i'$. Figure 4.15 displays the characteristics of these filters.

Most any deeply exposed frame in the region imaged by SDSS will include stars[47] with cataloged[48] photometry accurate to approximately $\pm.03^m$. While this error is $\sim 10\times$ larger than the usual $UBVRI$ standard stars, it is adequate for this lab. Unfortunately, many open clusters are not in the region imaged by SDSS (because the focus of the project is objects beyond our Galaxy.) These Sloan magnitudes are based on the AB_ν system rather than the Vega system. Instead of the star Vega defining magnitude=0, it is intended that, for each filter, magnitude=0 corresponds to $F_\nu = 3631$ Jy. The results are not hugely different from Vega-based systems: $g = -.08$, $g - r = -0.25$, $r - i = -0.23$ for Vega.

Various transformations of $BVRI \iff g'r'i'$ for normal stars have been published[49], I reproduce a few of these results below:

$$B - g = 0.313(g - r) + 0.219 \tag{4.46}$$
$$V - g = -0.565(g - r) - 0.016 \tag{4.47}$$
$$R - r = -0.153(r - i) - 0.117 \tag{4.48}$$
$$i - I = 0.247(R - I) + 0.329 \tag{4.49}$$
$$r - i = 1.007(R - I) - 0.236 \tag{4.50}$$
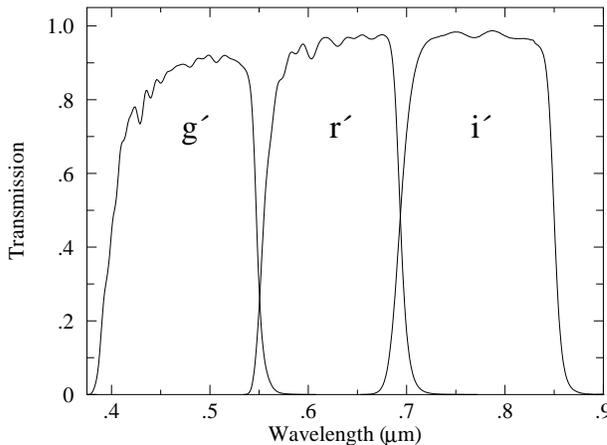
Filter Name	λ (μm)	$\Delta\lambda$ (μm)	F_0 (W/m^2 μm)
g'	.46	.13	4.71×10^{-8}
r'	.61	.12	2.81×10^{-8}
i'	.74	.12	1.87×10^{-8}

Figure 4.15: The characteristics of the Sloan filters: g' (green), r' (red), i' (infrared) from the Gemini Observatory: http://www.gemini.edu and Annual Review Astronomy & Astrophysics (2005, Bessell)

---

[46] Fukugita, M., Ichikawa, T., Gunn, J. E., et al. 1996 AJ, 111, 1748

[47] The brightest stars recorded will have mag=14, which in the context of our telescope, are dim. For r' and i' use exposure times ≥ 120 s. For g' use exposure times ≥ 300 s.

[48] http://www.sdss.org/

[49] Jordi, K.; Grebel, E. K.; Ammon, K. A&A **460** 339 (arXiv:astro-ph/0609121)
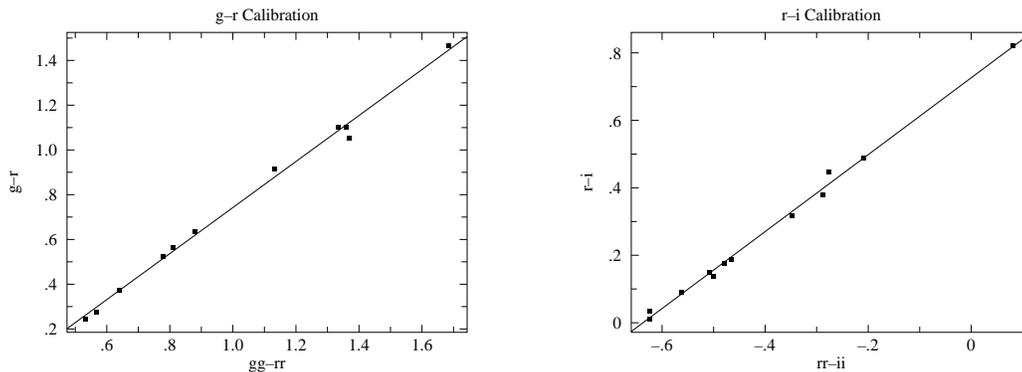
(a) Calibration of the standardize color index $g - r$ in terms of the instrumental color index $gg - rr$. See Eq. 4.44. The rms deviation from this calibration is about 0.03 mag.

$g - r = -0.29 + 1.03(gg - rr)$

(b) Calibration of the standardize color index $r - i$ in terms of the instrumental color index $rr - ii$. The rms deviation from this calibration is about 0.015 mag.

$r - i = 0.73 + 1.14(rr - ii)$

Figure 4.16: Two identical frames (taken 16-Aug-2010) of the central region of the open cluster NGC 6791 were used to calibrate the relationships between standardized color indices and instrumental color indices . This region includes 12 standard stars ($BVRIgri$).

Appendix: ST-7E Camera

In June 2009 we 'retired' the old but fully functional SBIG ST-7E camera. I've dumped the specifications for this camera in this Appendix as it is possible that it might be returned to service. Flaws in the CCD are most evident in the bias frame[50]. Measurements on our CCD shows that the difference between two bias frames shows an approximately normal distribution (aside from the flaws) with a standard deviation ~ 6 ADU. Also note that the system artificially introduces an offset of 100 into each pixel, so $\frac{2}{3}$ of the pixels in a bias frame are expected in the range 100 ± 6 ADU. The ST-7 uses a thermoelectric cooler which can cool the CCD $\sim 30°$C below ambient temperature, reducing the dark currents by about a factor of 10. The resulting dark currents are *not* negligible for exposures longer than a few minutes.

---

[50]Our CCD has a flaw at pixel $(592, 194)$ which affects the entire 592 column

Telescope	Meade LX200		CCD	Kodak KAF-0401E
aperture	12"		pixel	9 μm $\times$ 9 μm
focal ratio	$f/10$		full-well	$\sim 50,000\ e^-$/pixel
focal length	120"		read noise	15 e^- rms
Focal Reducer	Meade f/3.3 FR/FF		gain	2.3 e^-/ADU
effective f	50"			
Filter Wheel	SBIG CFW-8A		λ	Quantum Efficiency
Camera	SBIG ST-7E ABG		.65 μm	60%
pixels	765×510		.55 μm	50%
image	$18' \times 12'$		.45 μm	40%
scale	1.45"/pixel		.40 μm	30%

Table 4.4: Older CCD photometry equipment available at the SJU observatory. Note: When combined with a Celestron f/6.3 FR/FF this camera yielded 0.97"/pixel; when used on the 10" Meade telescope without any FR/FF it yielded 1.07"/pixel.

The large open cluster IC 4665 was the target for two nights during August 2005. The results are reported in Figures 4.19 and 4.20.

During the first night an overlapping mosaic of five frames was used to image the central region of the cluster. During a 2 hour period $BVRI$ frames were taken at each pointing. (During this time the air mass to IC 4665 changed from 1.30 to 1.46, but the data show no sign of a change in the transparency of the atmosphere, so no air mass corrections were applied.) Dome flat-field frames (taken 12 hours earlier) were used to produce reduced frames. `gaia` was used to measure the instrumental magnitude of 18 standard stars and three 'unknown' stars. The magnitudes were pasted into a `gnumeric` spreadsheet which was used to calculate the color indices: $b - v$ and $r - i$. (Since overlapping fields were used typically a standard star appeared in a couple of frames. When this occurred average instrumental color indices were used.) A calibrating line relating the standard stars' known $B - V$ and the measured $b - v$ was fit using regression in `gnumeric`. An identical process was used to find the relationship between $R - I$ and $r - i$. The results are displayed in Figures 4.19c and 4.19d. Using these relationships and the measured instrumental color indices, we can determine the standardized color indices of any star in the imaged region. As shown in Table 4.5, Eq. 4.28 or Eq. 4.29 can then be used to find the temperature of each 'unknown' star.

During the second night six (generally not overlapping) frames were selected with the aim of capturing a diverse (hot and cool) set of standard stars. The data collected during this night *did* show a change in the transparency of the atmosphere (there was a larger variation in air mass: 1.32 to 1.63; see Figure 4.20a), so air mass corrections were applied.
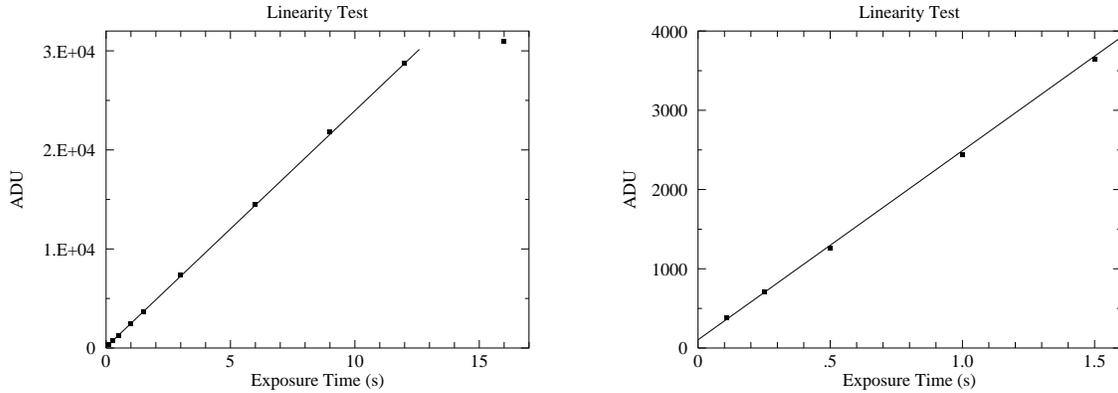
Figure 4.17: Below $\sim 30{,}000$ ADU (12 second exposure) the response of our CCD seems to be linear, with saturation quite evident in the 16 second exposure. In fact, the CCD should not be trusted to be linear above $\sim 25{,}000$ ADU.

The linear fit may look good, but it is in fact not great: reduced χ^2=8. Systematic error in the shutter speed control is probably a problem for the short exposures. (Elimination of the two shortest exposures results in a reduced χ^2=1.6.) Also note that the y intercept is not exactly zero: with this camera zero-time exposures are designed to produce an output of about 100 ADU.

Label	Name	SpType	$b-v$	$r-i$	$B-V$	$R-I$	T_B (K)	T_R (K)
A	BD+05 3486	A2	0.404	−0.708	0.2822	0.1826	7255	6966
B	GSC 00428-00981	G5	0.833	−0.387	0.8399	0.4625	5133	5015
C	TYC 424-517-1	K5	1.047	−0.151	1.1181	0.6683	4421	4306

Table 4.5: The instrumental color indices: $b-v$ and $r-i$ of the 'unknown' stars in IC 4665 were measured using `gaia` and converted to standardized color indices: $B-V$ and $R-I$ using the calibration lines displayed in in Figures 4.19c and 4.19d. Eq. 4.28 (T_B) and Eq. 4.29 (T_R) were used to calculate the temperature of the stars from each color index. The two temperatures do not exactly agree; as discussed in the text systematic errors of a few percent are to be expected.
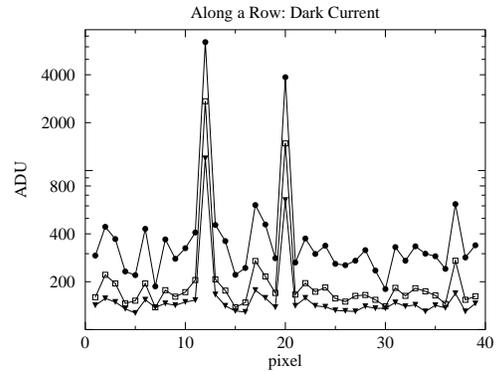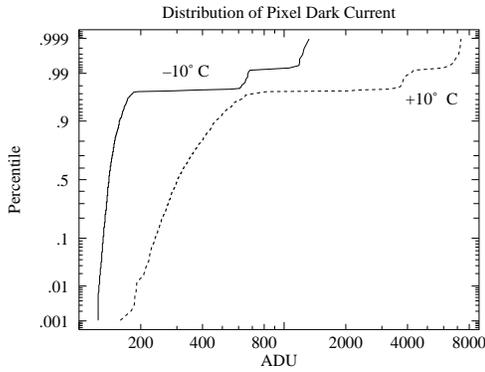
(a) The distribution of pixel dark current at $+10°$C and $-10°$C. All but a few of the pixels have a (log)normal distribution of dark current. A few percent of the pixels show much larger dark current: these are the 'hot' pixels. Note that both classes of pixels respond to temperature: lowering the temperature reduces every pixel's dark current.

(b) Part of a row of pixels is displayed at three temperatures: ● = $+10°$C, □ = $0°$C, ▼ = $-10°$C. Isolated hot pixels are randomly sprinkled throughout the image (here at columns 12 and 20). The data here and in (a) are from 15 minute dark frames.
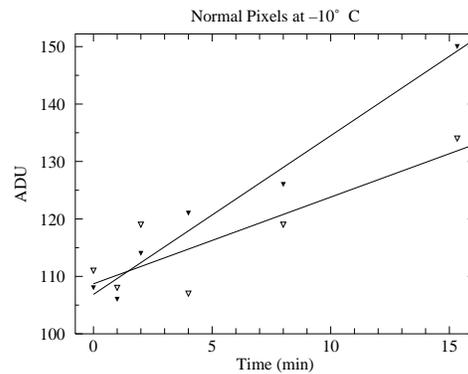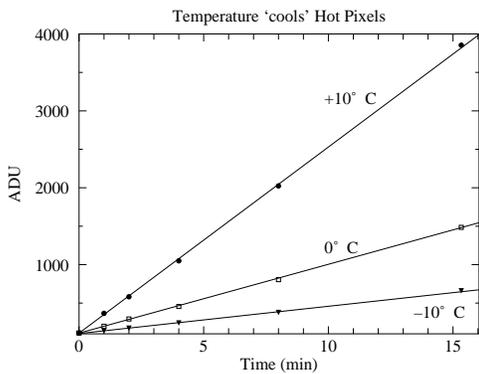
(c) Dark current produces ever larger stored charge in a pixel; of course reducing the temperature reduces the dark current.

(d) While it is more evident in this low count data, there is always deviation in counts. While we can subtract the average dark charge, deviations from the average produce 'noise' in our images. Reducing the temperature reduces this noise.

Figure 4.18: Dark frames are 'exposures' with the shutter closed. The source of these counts is not starlight, rather a thermally induced current allows charge to accumulate in each pixel. This 'dark current' can be exceptionally large in 'hot pixels'.
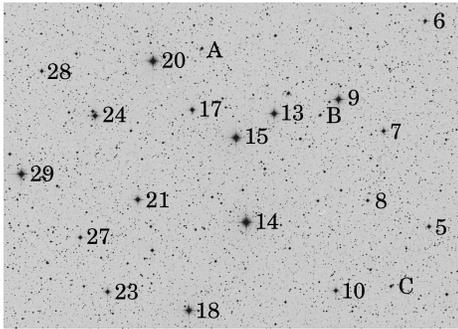
(a) This Digitized Sky Survey image shows the central $\sim 50' \times 25'$ region of the open cluster IC 4665. The standard stars of J.W. Menzies and F. Marang (MNRAS, 282, 313–316 (1996)) are labeled along with three 'unknown' stars: A, B, C.
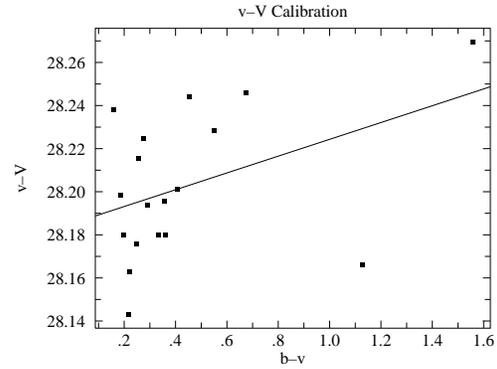
(b) There is little evidence for a significant color correction in this data set. See Eq. 4.42.
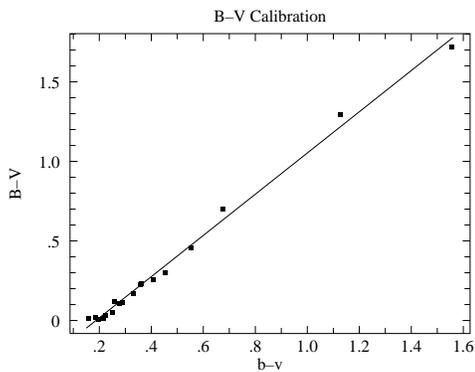
(c) Calibration of the standardize color index $B - V$ in terms of the instrumental color index $b - v$. See Eq. 4.44.
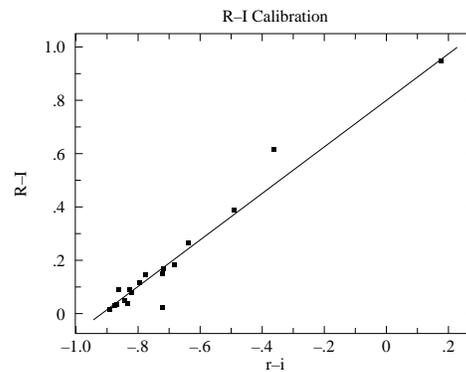
(d) Calibration of the standardize color index $R - I$ in terms of the instrumental color index $r - i$.

Figure 4.19: An overlapping mosaic of five frames was used to image the central region of the open cluster IC 4665. This region includes 18 standard stars which were used to calibrate the relationships between standardized color indices and instrumental color indices. Using these relationships and the measured instrumental color indices, we can determine the standardized color indices of any star in the imaged region.
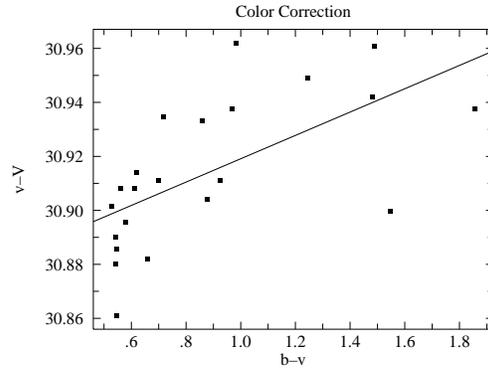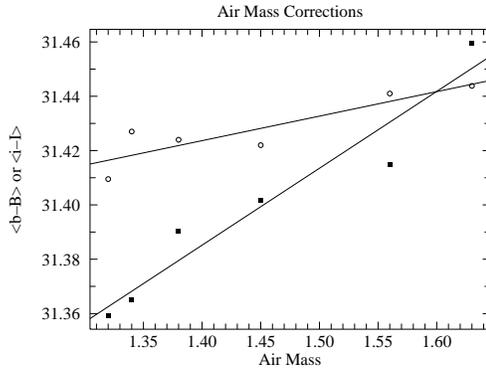
(a) Short wavelengths are more affected by atmospheric absorption than long; the above shows the effect on B magnitudes (filled squares) and I magnitudes (open circles). The y axis displays the difference between instrumental and standardized magnitudes averaged over all the blue stars in the frame.

(b) Color correction is evident in this data set, but both air mass and color correction should be applied to this data. This full result was: $v - V = 30.611 + 0.179 \sec z + 0.048(b - v)$. The rms variation from this line was less than 0.02 mag.
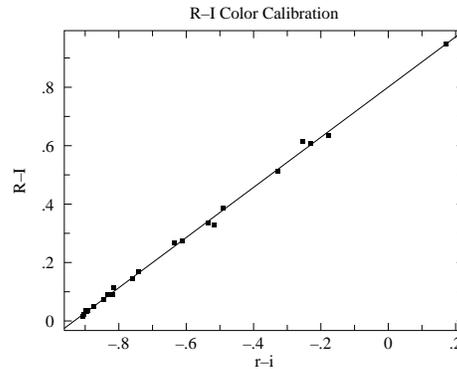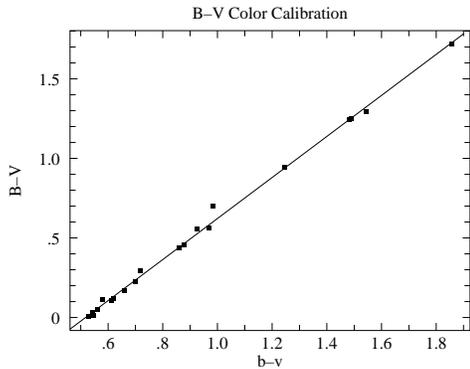
(c) Calibration of the standardize color index $B - V$ in terms of the instrumental color index $b - v$.
$B - V = -0.67 + 1.29(b - v)$

(d) Calibration of the standardize color index $R - I$ in terms of the instrumental color index $r - i$.
$R - I = 0.80 + 0.86(r - i)$

Figure 4.20: Six frames were used to image 22 standard stars in the open cluster IC 4665. The selected stars have a wide variation in color in order to to achieve robust color calibrations. During the 2^+ hours of data collection the air mass increased from 1.32 to 1.63 and the resulting increase in atmospheric absorption increased (dimmed) B, V, R, and I magnitudes in proportion to the air mass (see Figure 4.20a).

5: Thermionic Emission

What Leon [Cooper] did ...was to make a brilliant oversimplification which allowed its essential nature to be comprehended immediately. This process of "model-building", essentially that of discarding all but essentials and focusing on a model simple enough to do the job but not too hard to see all the way through, is possibly the least understood — and often the most dangerous — of all the functions of a theoretical physicist...Actually, in almost every case where I have been really successful it has been by dint of discarding almost all of the apparently relevant features of reality in order to create a "model" which has the two almost incompatible features:

(1) enough simplicity to be solvable, or at least understandable;

(2) enough complexity left to be interesting, in the sense that the remaining complexity actually contains some essential features which mimic the actual behavior of the real world, preferably in one of its as yet unexplained aspects.

Philip W. Anderson (1977 Nobel Laureate) *More and Different: notes from a thoughtful curmudgeon* (2011) p. 37

Purpose

While we think of quantum mechanics being best demonstrated in processes that show discontinuous change, historically quantum mechanics was first revealed in systems where a large number of particles washed out the jumps: blackbody radiation and thermionic emission. In this lab you will investigate these two phenomena in addition to classical space-charge limited electron emission: Child's Law.

Introduction

Metals, as demonstrated by their ability to conduct an electric current, contain mobile electrons. (Most electrons in metals, particularly the "core" electrons closest to the nucleus, *are* tightly bound to individual atoms; it is only the outermost "valence" electrons that are somewhat "free".) These free electrons are generally confined to the bulk of the metal. As you learned in E&M, an electron attempting to leave a conductor experiences a strong force
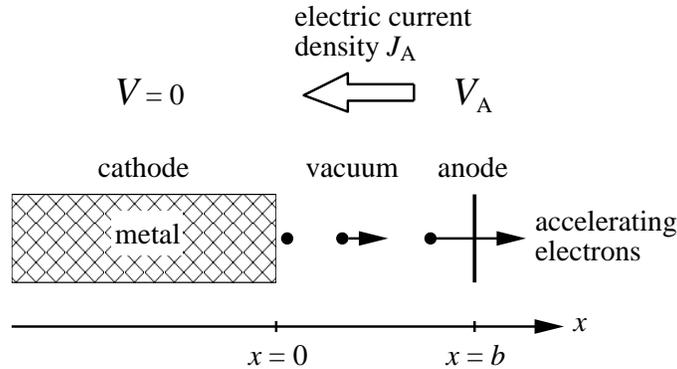
Figure 5.1: A planar cathode and a planar anode are separated by a distance b. A positive potential difference V_A attracts electrons from the cathode to the anode, so the speed of the electrons $v(x)$ increases as they approach the anode. The moving electrons constitute an electric current from anode to cathode. The resulting steady current density is called J_A.

attracting it back towards the conductor due to an image charge:

$$F_x = -\frac{e^2}{4\pi\epsilon_0(2x)^2} \tag{5.1}$$

where x is the distance the electron is from the interface and e is the absolute value of the charge on an electron. Of course, inside the metal the electric field is zero so an electron there experiences zero (average) force. You can think of these valence electrons as bouncing around inside a box whose "walls" are provided by the image-charge force. (Odd to think: the "walls" are non-material force fields; the "inside" of the box is filled with solid metal.) Since temperature is a measure of random kinetic energy, if we increase the temperature of the metal, the electrons will be moving faster and some will have enough energy to overcome the image-charge force (which after all becomes arbitrarily small at large distances from the interface) and escape. This is electron "evaporation". The higher the temperature the larger the current of escaping electrons. This temperature induced electron flow is called thermionic emission. Starting in 1901, Owen Richardson studied this phenomenon and in 1929 he received the Nobel prize in Physics for his work.

A hot wire will be surrounded by evaporated electrons. An electric force can pull these electrons away from the wire — the larger the electric force, the larger the resulting current of electrons. The precise relationship between the voltage and the resulting current flow is called Child's law[1] (or the Child-Langmuir law, including Langmuir who independently discovered it while working at G.E.). In this experiment you will measure both Child's Law and the Richardson Effect.

Child's Law

Consider a planar interface between a metal ($x < 0$) and "vacuum" ($x > 0$). Vacuum is in quotes because this region will contain escaped electrons—a 'space charge'—rather than

---

[1]Clement Dexter Child (1868–1933) Born: Madison, Ohio, A.B. Rochester, Ph.D. Cornell

being totally empty[2]. The number of electrons per volume (i.e., the number density) is denoted by n.

In this experiment, the metal will be heated (i.e., its a 'hot cathode' or filament) which will result in a supply of electrons 'evaporated' from the metal into the vacuum. An additional conducting sheet (the anode) is located at $x = b$. A positive potential difference, V_A, between the cathode and the anode plane provides a force pulling these electrons from the vicinity of the cathode towards the anode. The result is a stream of moving electrons (a current); the number density $n(x)$ and speed $v(x)$ of these electrons will depend on location, x, between the plates. The negatively charged electrons moving to the right constitute a steady electric current density to the left, i.e., a steady conventional electric current from the anode to the cathode:

$$J = -en(x)v(x) = -J_A \tag{5.2}$$

Since the electrons leave the metal with (nearly) zero speed at zero potential, we can calculate their speed along the path to the anode using conservation of energy:

$$\frac{1}{2}mv^2 - eV(x) = 0 \tag{5.3}$$

$$v = \sqrt{\frac{2e}{m} V(x)} \tag{5.4}$$

where $V(x)$ is the potential difference ("voltage") at x and m is the mass of an electron. Because the accelerating electrons constitute a steady current (i.e., J_A doesn't depend on position), $n(x)$ must decrease as the electrons speed toward the anode. The varying space charge density affects the electric potential in the "vacuum" according to Poisson's equation[3]:

$$\frac{\partial^2 V}{\partial x^2} = -\frac{\rho(x)}{\epsilon_0} = \frac{en(x)}{\epsilon_0} \tag{5.5}$$

Putting these pieces together with have the differential equation:

$$\frac{d^2V}{dx^2} = \frac{J_A}{\epsilon_0 v(x)} = \frac{J_A}{\epsilon_0 \sqrt{\frac{2e}{m} V(x)}} \tag{5.6}$$

Since the electric field will be zero at the interface, we have a pair of initial conditions:

$$\left. \frac{\partial V}{\partial x} \right|_{x=0} = 0 \tag{5.7}$$

$$V|_{x=0} = 0 \tag{5.8}$$

This differential equation looks a bit like Newton's second law:

$$\frac{d^2x}{dt^2} = \frac{1}{m} F(x(t)) \tag{5.9}$$

as you can see if in Newton's second law you substitute:

$$t \longrightarrow x$$
$$x(t) \longrightarrow V(x)$$
$$\frac{1}{m} F(x(t)) \longrightarrow \frac{J_A}{\epsilon_0 \sqrt{\frac{2e}{m} V(x)}}$$

---

[2]In fact a perfect vacuum is not possible, so the word "vacuum" actually refers simply to a region with relatively few particles per volume

[3]Poisson's equation is derived in the Appendix to this lab.

Recall that force problems are often most simply solved using conservation of energy and that conservation of energy was proved using an integrating factor of dx/dt. If we try the analogous trick on our voltage problem, we'll multiply Poisson's equation by dV/dx:

$$\frac{dV}{dx} \times \frac{d^2V}{dx^2} = \frac{J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}} \, V^{-\frac{1}{2}} \times \frac{dV}{dx} \tag{5.10}$$

$$\left(\frac{1}{2} \left[\frac{dV}{dx} \right]^2 \right)' = \frac{J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}} \left(\frac{V^{\frac{1}{2}}}{\frac{1}{2}} \right)' \tag{5.11}$$

$$\frac{1}{2} \left[\frac{dV}{dx} \right]^2 = \frac{J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}} \frac{V^{\frac{1}{2}}}{\frac{1}{2}} + \text{constant} \tag{5.12}$$

The initial conditions require the constant to be zero, so

$$\frac{1}{2} \left[\frac{dV}{dx} \right]^2 = \frac{J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}} \frac{V^{\frac{1}{2}}}{\frac{1}{2}} \tag{5.13}$$

or

$$\frac{dV}{dx} = \sqrt{\frac{4J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}}} \, V^{\frac{1}{4}} \tag{5.14}$$

This differential equation is separable:

$$\frac{dV}{V^{\frac{1}{4}}} = \sqrt{\frac{4J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}}} \, dx \tag{5.15}$$

$$\frac{V^{\frac{3}{4}}}{\frac{3}{4}} = \sqrt{\frac{4J_A}{\epsilon_0 \sqrt{\frac{2e}{m}}}} \, x \tag{5.16}$$

where again the initial conditions require the constant of integration to be zero. Finally:

$$V(x) = \left[\frac{9J_A}{4\epsilon_0 \sqrt{\frac{2e}{m}}} \right]^{\frac{2}{3}} x^{\frac{4}{3}} \tag{5.17}$$

Of course, $V(b)$ is the anode voltage V_A, so we can rearrange this equation to show Child's law:

$$J_A = \left[\frac{4\epsilon_0}{9b^2} \sqrt{\frac{2e}{m}} \right] V_A^{\frac{3}{2}} \tag{5.18}$$

Much of Child's law is just the result of dimensional analysis, i.e., seeking any possible dimensionally correct formula for J_A. Our differential equation just involves the following
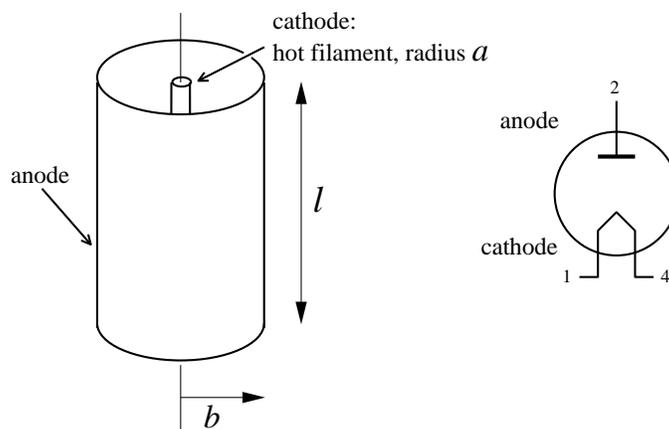
Figure 5.2: Coaxial cylinders: an inner wire (radius a) and outer cylindrical anode (radius b), form a vacuum tube diode. The cathode is heated so electron evaporation is possible, and a potential difference V_A attracts electrons from the cathode to the anode. The speed of the electrons $v(r)$ increases as they approach the anode. The moving electrons constitute a steady electric current from anode to cathode. Since the same current is spread out over larger areas, the current density, J, between the cylinders must be proportional to $1/r$.

constants with dimensions (units) as shown:

$$b \quad : \quad L \tag{5.19}$$

$$V_A \quad : \quad \frac{E}{Q} = \frac{ML^2/T^2}{Q} \tag{5.20}$$

$$\epsilon_0\sqrt{\frac{2e}{m}} \equiv k \quad : \quad \frac{Q^2}{EL}\frac{Q^{\frac{1}{2}}}{M^{\frac{1}{2}}} = \frac{Q^{\frac{5}{2}}}{M^{\frac{3}{2}}L^3/T^2} \tag{5.21}$$

$$J_A \quad : \quad \frac{Q/T}{L^2} \tag{5.22}$$

where the dimensions are: L=length, T=time, M=mass, E=energy, and Q=charge. To make a dimensionally correct formula for J_A, we just eliminate the M dimension which we can only do with the combination:

$$V_A k^{\frac{2}{3}} \quad : \quad \frac{Q^{\frac{2}{3}}}{T^{\frac{2}{3}}} \tag{5.23}$$

We can then get the right units for J_A with:

$$\frac{\left(V_A k^{\frac{2}{3}}\right)^{\frac{3}{2}}}{b^2} = \frac{k}{b^2}\,V_A^{\frac{3}{2}} \quad : \quad \frac{Q/T}{L^2} \tag{5.24}$$

Thus the only possible dimensionally correct formula is

$$J_A \propto \frac{k}{b^2}\,V_A^{\frac{3}{2}} \tag{5.25}$$

The exact proportionality constant, found from the differential equation, is (as usual) is not hugely different from 1.

We have derived Child's law for the case of infinite parallel plates, but you will be testing it in (finite length) coaxial cylinders. The inner wire (radius a) is the cathode; the outer cylinder (radius b) is the anode. Your cylinder with have some length ℓ, but we will below consider infinite length coaxial cylinders. Note that dimensional considerations require that the anode current per length should be given by a formula like:

$$I/\ell \equiv j \propto \frac{k}{b} \, V_A^{\frac{3}{2}} \tag{5.26}$$

although we could have an arbitrary function of the radius ratio: b/a on the right-hand-side.

From Poisson's equation[4] we have:

$$\nabla^2 V = \frac{J}{\epsilon_0 v(r)} = \frac{I}{2\pi r \ell \epsilon_0 v(r)} = \frac{j}{2\pi r \epsilon_0 \sqrt{\frac{2e}{m}}} \, V^{-\frac{1}{2}} \tag{5.27}$$

Using the Laplacian in cylindrical coordinates we find:

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \frac{\partial V}{\partial r} = \frac{j}{2\pi r \epsilon_0 \sqrt{\frac{2e}{m}}} \, V^{-\frac{1}{2}} \tag{5.28}$$

There is no known formula for the solution to this differential equation, but we can make considerable progress by writing the differential equation in terms of dimensionless quantities:

$$r/a \;=\; \rho \tag{5.29}$$

$$V \;=\; \left[\frac{ja}{2\pi\epsilon_0 \sqrt{\frac{2e}{m}}} \right]^{\frac{2}{3}} f(\rho) \tag{5.30}$$

yielding:

$$\frac{\partial^2 f}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial f}{\partial \rho} = f''(\rho) + \frac{1}{\rho} \, f'(\rho) = \frac{1}{\rho} \, f^{-\frac{1}{2}} \tag{5.31}$$

with initial conditions:

$$f(1) \;=\; 0 \tag{5.32}$$
$$f'(1) \;=\; 0 \tag{5.33}$$

We can numerically solve this differential equation using *Mathematica*:

```
NDSolve[{f''[p]+f'[p]/p==1/(p Sqrt[f[p]])}, f[1]==0, f'[1]==0, {f},{p,1,200}]
```

It's actually not quite that simple. The cathode, at $\rho = 1$, is actually a singular point of the differential equation (i.e., $f''(1) = \infty$). However the situation very near the cathode is well approximated by the planar case, where we've shown:

$$
\begin{aligned}
V(x) \;&=\; \left[\frac{9 J_A}{4\epsilon_0 \sqrt{\frac{2e}{m}}} \right]^{\frac{2}{3}} x^{\frac{4}{3}} = \left[\frac{9 I}{2\pi a \ell 4\epsilon_0 \sqrt{\frac{2e}{m}}} \right]^{\frac{2}{3}} (r-a)^{\frac{4}{3}} = \left[\frac{9 ja}{2\pi 4\epsilon_0 \sqrt{\frac{2e}{m}}} \right]^{\frac{2}{3}} \left(\frac{r-a}{a} \right)^{\frac{4}{3}} \\[2mm]
&=\; \left[\frac{9}{4} \right]^{\frac{2}{3}} \left[\frac{ja}{2\pi\epsilon_0 \sqrt{\frac{2e}{m}}} \right]^{\frac{2}{3}} \left(\frac{r-a}{a} \right)^{\frac{4}{3}}
\end{aligned}
\tag{5.34}
$$

---

[4]Poisson's equation is derived in the Appendix to this lab.
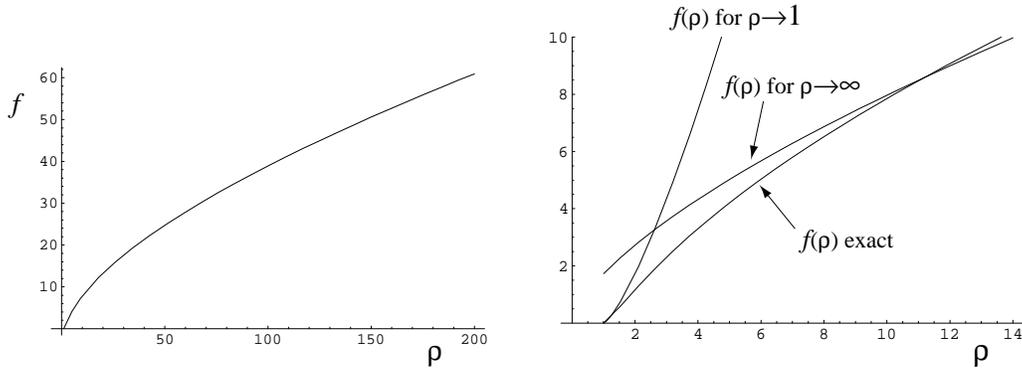
Figure 5.3: The plot on the left displays the dimensionless voltage f obtained by numerical solution to the differential equation. The plot on the right compares various approximations for f to this numerical solution.

So, near the cathode (i.e., ρ slightly larger than 1):

$$f(\rho) \approx \left[\frac{9}{4}\right]^{\frac{2}{3}} (\rho - 1)^{\frac{4}{3}} \tag{5.35}$$

We can use this approximation to start our numerical differential equation solution at a non-singular point (like $\rho = 1.00001$).

Real devices are designed with $b/a = \rho_{\text{anode}} \gg 1$. The behavior of f for large ρ can be determined by finding constants A and α for which $f = A\rho^{\alpha}$ is a solution to the differential equation. One finds:

$$f = \left(\frac{9}{4}\rho\right)^{\frac{2}{3}} \tag{5.36}$$

A useful approximation for the range: $100 < b/a < 1000$ is:

$$f = \left(\frac{9}{4}\rho\right)^{\frac{2}{3}} + 2 \tag{5.37}$$

(For example, the device used in lab has $b/a = 121.5$. For this value, the differential equation gives $f = 44.136$; the above approximation gives: $f = 44.130$.)

We recover Child's law by rearranging (5.30):

$$\frac{2\pi\epsilon_0 \sqrt{\frac{2e}{m}}}{a} \left[\frac{V_A}{f(b/a)}\right]^{\frac{3}{2}} = j = I/\ell \tag{5.38}$$

Note: Langmuir's original work (Phys. Rev. **22**, 347 (1923)) on this subject is expressed in terms of β where:

$$\beta^2(\rho) \equiv \frac{4}{9}\frac{f^{\frac{3}{2}}}{\rho} = \begin{cases} \xrightarrow{\rho \to 1} & (\rho - 1)^2 \\ \xrightarrow{\rho \to \infty} & 1 \end{cases} \tag{5.39}$$

So:

$$\boxed{\frac{8\pi\epsilon_0 \ell \sqrt{\frac{2e}{m}}}{9b\beta^2} V_A^{\frac{3}{2}} = I} \tag{5.40}$$

$\beta^2 = 1.072$ for the device used in lab.

Richardson's Law

Most any thermal process is governed by the Boltzmann factor:

$$\exp\left(-\frac{\Delta E}{kT}\right) = e^{-\Delta E/kT} \tag{5.41}$$

where k is the Boltzmann constant. Approximately speaking the Boltzmann factor expresses the relative probability for an event requiring energy ΔE in a system at (absolute) temperature T. Clearly if $\Delta E \gg kT$, the probability of the event happening is low. If an electron requires an energy W (called the work function) to escape from the metal, The Boltzmann factor suggests that this would happen with relative probability $e^{-W/kT}$. Thus you should expect that the current emitted by a heated metal would follow:

$$I \sim e^{-W/kT} \tag{5.42}$$

Clearly you should expect different elements to have different work functions, just as different atoms have different ionization potentials. What is surprising is that the proportionality factor in the above equation includes a universal constant — that is, a constant that just depends on the properties of electrons (and, most importantly, Planck's constant, h) and does *not* depend on the type of material. (This situation is similar to that of blackbody radiation, in which photons rather than electrons are leaving a heated body, which was Planck's topic in discovering his constant. We will take up this topic on page 121.) Thermionic emission probes the quantum state of the electrons statistically, whereas the photoelectric effect probes much the same physics electron by electron. (The next level problem is to explain why this universal constant (the Richardson constant, A) in fact does depend a bit on the material.) To show:

$$J = AT^2 e^{-W/kT} \tag{5.43}$$

where

$$A = \frac{4\pi emk^2}{h^3} = 1.2 \times 10^6 \,\text{A/m}^2\text{K}^2 \tag{5.44}$$

Quantum Theory: Free Electron Gas

Instead of thinking about electron particles bouncing around inside a box, de Broglie invites us to consider standing waves of electron probability amplitude:

$$\psi = N \exp(ik_x x) \exp(ik_y y) \exp(ik_z z) = N e^{i\mathbf{k}\cdot\mathbf{r}} \tag{5.45}$$

Recall [5] that vector $\hbar\mathbf{k}$ is the momentum, $\mathbf{p} = m\mathbf{v}$, of the electron and $\hbar = h/2\pi$. Periodic boundary conditions on the box (which we take to be a cube with one corner at the origin and the diagonally opposite corner at the point $\mathbf{r} = (L, L, L)$) require each component k_i to satisfy:

$$k_i = \frac{2\pi \, n_i}{L} \tag{5.46}$$

---

[5]For a review see: `http://britneyspears.ac/physics/dos/dos.htm`
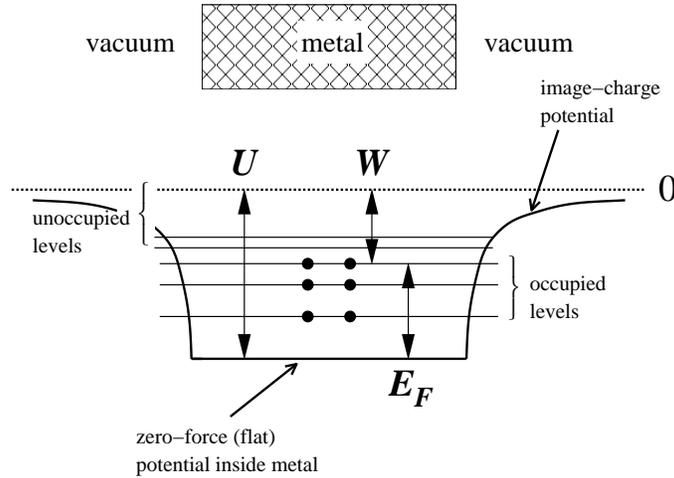
Figure 5.4: Electrons in the metal experience a constant confining potential of depth U. Possible quantum mechanical states for these electrons are displayed as horizontal lines. Electrons fill all the available states up to the Fermi energy, E_F. The work function, W, is defined at the minimum energy needed to remove an electron from the metal. As shown above: $W = U - E_F$.

where each n_i is an integer. Thus each wave function is specified by a triplet of integers: $\mathbf{n} = (n_x, n_y, n_z)$, the $\mathbf{n}$-vector. Applying Schrödinger's equation, we find that this wavefunction has energy:

$$E(\mathbf{n}) = \frac{\hbar^2 \mathbf{k}^2}{2m} = \frac{(2\pi\hbar)^2 \mathbf{n}^2}{2mL^2} = \frac{(2\pi\hbar)^2 (n_x^2 + n_y^2 + n_z^2)}{2mL^2} \tag{5.47}$$

Notice that there is a definite relationship between the velocity vector of the electron and the $\mathbf{n}$-vector.

$$\mathbf{v} = \frac{2\pi\hbar}{mL} \, \mathbf{n} \tag{5.48}$$

Another way of saying the same thing is that allowed quantum-state velocities form a cubic lattice with cube-side $2\pi\hbar/mL$. The number of states with electron velocities in some specified region (for example a velocity-space parallelepiped with sides: $\Delta v_x \Delta v_y \Delta v_z$) can be found from the number of $2\pi\hbar/mL$ sided cubes that fit into the region, which is the volume of that velocity-space region divided by $(2\pi\hbar/mL)^3$. Hence:

$$\text{number of states with velocity between } \mathbf{v} \text{ and } \mathbf{v} + \Delta\mathbf{v} = \frac{\Delta v_x \Delta v_y \Delta v_z}{(2\pi\hbar/mL)^3} \tag{5.49}$$

$$\text{number of states } \underline{\text{per volume}} \text{ with velocity between } \mathbf{v} \text{ and } \mathbf{v} + \Delta\mathbf{v} = \frac{\Delta v_x \Delta v_y \Delta v_z}{(2\pi\hbar/m)^3}$$

$$= \left(\frac{m}{2\pi\hbar}\right)^3 \Delta v_x \Delta v_y \Delta v_z = \mathcal{N} \Delta v_x \Delta v_y \Delta v_z \tag{5.50}$$

where $\mathcal{N}$ is the (constant) density of states in velocity space.

Quantum Theory: Fermi Energy

Fermions (half-integer spin particles), in contrast to bosons (integer spin particles), cannot group together. Since the electron is "spin $\frac{1}{2}$", each of the above states can hold at most 2
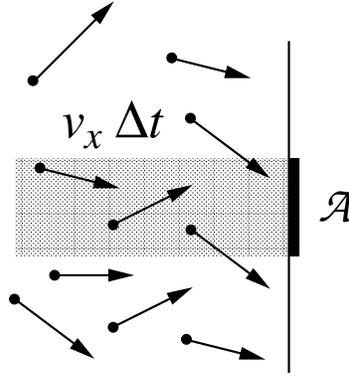
Figure 5.5: Consider just those electrons with some particular x-velocity, v_x. In order to hit the wall during the coming interval Δt, an electron must be sufficiently close to the wall: within $v_x \Delta t$. The number of such electrons that will hit an area $\mathcal{A}$ will be equal to the number of such electrons in the shaded volume (which is the perpendicular extension of $\mathcal{A}$ a distance $v_x \Delta t$ into the volume). Note that many electrons in that volume will *not* hit $\mathcal{A}$ because of large perpendicular velocities, but there will be matching electrons in neighboring volumes which will hit $\mathcal{A}$. To find the total number of hits, integrate over all possible v_x.

electrons: one spin up and one spin down. The probability that a particular fermion state with energy E will be occupied is given by a generalization of the Boltzmann factor called Fermi-Dirac statistics:

$$f(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)} \tag{5.51}$$

where E_F is called the Fermi energy. The Fermi energy is basically a disguise for the number of electrons, as, approximately speaking, it is the dividing line between occupied states and unoccupied states. (If the Fermi energy is high, there must be lots of occupied states and hence lots of electrons.) Note that if $E \gg E_F$, the exponential factor is huge and we can neglect the "+1" in the denominator so

$$f(E) \approx \exp\left(-\frac{E - E_F}{kT}\right) \tag{5.52}$$

that is, if $E \gg E_F$ Fermi-Dirac statistics approximate the Boltzmann factor.

Classical Theory: Electron Escape

The density of states combined with the Boltzmann factor gives us the number of free electrons per unit volume with a particular velocity. In order for an electron to escape during some time Δt, it must have v_x sufficient to overcome the image-charge barrier and it must be sufficiently close to the wall. All the electrons with $v_x > \sqrt{2U/m}$ within a distance $v_x \Delta t$, will escape, where U is the depth of the potential well for the electrons. Thus the

number of electrons escaping through area $\mathcal{A}$ during Δt is:

$$\int_{\sqrt{2U/m}}^{\infty} dv_x \int_{-\infty}^{\infty} dv_y \int_{-\infty}^{\infty} dv_z \, 2\mathcal{N} f(E) \, \mathcal{A} v_x \Delta t$$

$$= 2\mathcal{N} e^{E_F/kT} \mathcal{A} \Delta t \int_{\sqrt{2U/m}}^{\infty} e^{-mv_x^2/2kT} \, v_x dv_x \int_{-\infty}^{\infty} e^{-mv_y^2/2kT} dv_y \int_{-\infty}^{\infty} e^{-mv_z^2/2kT} dv_z$$

$$= \frac{4\pi m (kT)^2}{(2\pi\hbar)^3} \, \mathcal{A} \Delta t \, \exp\left(-\frac{(U - E_F)}{kT}\right) \tag{5.53}$$

where we have used the Gaussian integral:

$$\int_{-\infty}^{\infty} e^{-\alpha z^2} dz = \sqrt{\frac{\pi}{\alpha}} \tag{5.54}$$

The electric current density is the electric *charge* escaping per time per area:

$$J = \frac{e \times \text{number escaping}}{\mathcal{A} \Delta t} = \frac{4\pi e m (kT)^2}{h^3} \, \exp\left(-\frac{W}{kT}\right) \tag{5.55}$$

which is Richardson's equation, with work function W given by $U - E_F$.

Experiment: Richardson's "Constant" is Material Dependent!

Experimentally it is found that Richardson's constant depends on the material[6] Why?

1. The work function depends on temperature (due to, for example, thermal expansion of the lattice of atoms). If the data analysis assumes it's constant, the resulting A will be grossly in error.

2. Classically reflection requires a turning point (where $v_x = 0$), whereas quantum mechanical reflections are possible just due to sharp changes in potential. Quantum mechanical reflection at the metal boundary was not included in our calculations; we assumed every energetic electron headed toward the wall went through the wall.

3. Surface contamination can affect emission probability. In fact, it was originally thought that thermionic emission was 100% due to surface contamination. (You can think of surface contamination as a locally varying work function.)

4. Even in the absence of surface contamination, in typical experiments, the metal is polycrystalline and different crystal surfaces have different work functions.

---

[6]This should remind you a bit of the material dependent emissivity, ϵ_T, for blackbody radiation to be discussed on page 121.

Table 5.1: G.E. FP-400 Specifications

Filament (W) length	3.17 cm	(1.25")
Filament diameter	0.013 cm	(0.005")
Anode (Zr coated Ni) I.D.	1.58 cm	(0.620")
Maximum filament voltage	4.75 V	
Maximum filament current	2.5 A	
Maximum anode voltage	125 V	
Maximum anode current	55 mA	
Maximum anode dissipation	15 W	
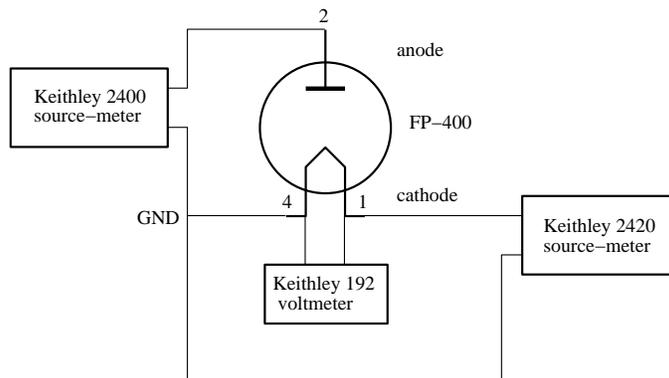
Figure 5.6: Circuit diagram (note tube pin labels) for thermionic emission experiment.

Experiment

This experiment involves thermionic emission from the hot tungsten filament of a G.E. FP-400 vacuum tube.

Temperature Determination

Often temperature measurement in physics experiments is easy. In the "normal" range of temperatures there are many types of transducers which convert temperature to an electrical quantity (e.g., Pt resistance thermometers, thermocouples, thermistors, ICs). However at the extremes of high and low temperature, measurement becomes tricky. Questions like "What exactly defines *temperature*?" must then be answered. This experiment requires "high" temperatures in a vacuum, so we do not have direct contact with the material whose temperature we seek. In addition the FP-400 tube was not built by us, so we have limited direct information about the device.

One common way to measure temperature is by using the temperature dependence of resistance. The resistance of metals is approximately proportional to temperature. Jones and Langmuir[7] have published a table of resistance vs. temperature for tungsten, from which

---
[7]GE Rev **30**, 310 (1927)

Kirkman has found an approximating formula:

$$T_r = 112 + 202x - 1.81x^2 \tag{5.56}$$

where x is the ratio of the hot resistance to that at 293 K.

A problem with this approach is that the measured resistance, R_{measured}, will include both the resistance of the tungsten[8] filament, R_W and the wires supporting it in the vacuum tube, R_{support}. Thus the quantity we seek (tungsten filament resistance, R_W) must be calculated as the small difference between two numbers:

$$R_W = R_{\text{measured}} - R_{\text{support}} \tag{5.57}$$

a situation that produces big relative errors. Additionally, we have no independent way of measuring R_{support} (we can't take the tube apart); In the end you will measure R_{support} at room temperature and then assume it is constant[9].

There is a further problem with any measurement of *voltage* when parts of the system are at different temperatures: thermally induced emfs (thermocouples). If the ends of the tungsten filament are at different temperatures, there will be a voltage induced approximately proportional to the temperature difference between the two ends. This additional voltage source confuses the resistance determination. The phenomena can be detected and corrected by reversing the direction of current flow (which reverses the Ohm's law voltage, but does not affect the sign of the thermal voltage.) Thermal voltages are generally less than a mV, and so are negligible once our measured voltages approach one volt.

Another approach is suggested by Jones & Langmuir. In a vacuum the temperature of a current-carrying wire is the result of an equilibrium between electrical power dissipated in the wire and energy lost in the form of radiation. (We assume that energy lost through conduction — either through the wire-supports or residual air in the "vacuum" — is negligible.) According to the Stefan-Boltzmann law, the power radiated from the surface of a hot black-body is given by:

$$P = \sigma T^4 \mathcal{A} \tag{5.58}$$

where σ is the Stefan-Boltzmann constant, T is the temperature of the body, and $\mathcal{A}$ is the surface area of the body. (In fact tungsten is not a black-body, so when applied to tungsten the above should be multiplied by a "fudge factor", the total emissivity ϵ_T, about 0.3.) Using just crude approximations, we equate the electrical power dissipated to the power radiated:

$$I^2 T \frac{\ell}{d^2} \sim I^2 T \frac{\ell}{\frac{\pi}{4}d^2} \sim I^2 R = \epsilon_T \sigma T^4 \pi d\ell \sim T^4 d\ell \tag{5.59}$$

where d is the diameter of the wire and ℓ is the length of the wire. On the right hand side we've assumed that the resistivity of tungsten is proportional to temperature, and on the left hand side we've assumed ϵ_T doesn't depend on temperature. We conclude:

$$I^2 \sim T^3 d^3 \tag{5.60}$$

$$\left[\frac{I}{d^{\frac{3}{2}}} \right]^{\frac{2}{3}} \sim T \tag{5.61}$$

---

[8]The chemical symbol for tungsten is W from the German Wolfram
[9]See "Assuming Away Variation" page 18 in Chapter 0
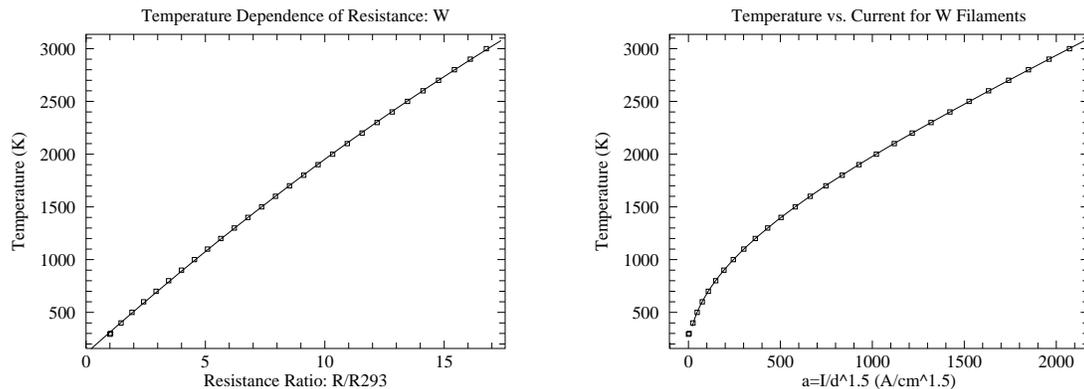
(a) The temperature of a tungsten filament as a function its resistance with an approximating curve. The x-axis is the ratio of the hot tungsten resistance to that at 293 K.

(b) The temperature of a tungsten filament as a function of the current flowing through it with an approximating curve. The x-axis is the current divided by the diameter of the wire (in cm) raised to the $\frac{3}{2}$ power.

Figure 5.7: From the data of Jones & Langmuir, two ways to determine temperature of a tungsten filament.

Absent the above approximations we can hope that temperature is a function of $a \equiv I/d^{\frac{3}{2}}$.

Once again Jones & Langmuir provide us with calibrating data for this expected relationship. For temperatures 400 K $< T <$ 3000 K, Kirkman finds:

$$T_i = 117 + 56a^{0.5} + 0.00036a^{1.8} \tag{5.62}$$

Finally, attaining thermal equilibrium[10] is a problem that affects most any temperature measurement. The balance between electrical power in and heat lost is not immediately achieved. The parts of the system with large heat capacities (particularly the filament supports and other large structures in the tube), will only gradually approach equilibrium. Thus "soak" time is required following each jump in heating power. The effect of this "thermal inertia" is seen in "hysteresis": temperatures obtained while increasing the temperature disagree with those found while decreasing the temperature. This will be an important source of uncertainty.

Hands-on Electrical Measurements

Support Resistance: R_{support}

As shown in Figure 5.6, the tungsten filament (cathode) is powered by a Keithley 2420 source-meter. The filament+support voltage is measured directly at the socket with a Keithley 192 voltmeter. By combining the current through the filament (measured from

---

[10]See "Special Problem: Temperature" page 17 in Chapter 0

the 2420) with the voltage across the socket (from the 192), the series resistance, $R_{\mathrm{measured}} = R_W + R_{\mathrm{support}}$, can be determined. At room temperature, the filament resistance, R_W, can be calculated from the filament geometry (see Table 5.1) and the resistivity of W-filament material at room temperature: $\rho_{293} = 5.49 \ \mu\Omega \cdot \mathrm{cm}$. (You should calculate: $R_W \sim .1 \ \Omega$.) Then R_{support} can be calculated from known quantities:

$$R_{\mathrm{support}} = R_{\mathrm{measured}} - R_W \tag{5.63}$$

Because room temperature R_{measured} is 'small' (and hence error prone), you will make three distinct measurements of it. Begin by sourcing 1 mA and then 10 mA into the room temperature filament (using the 2420), reading the resulting voltages on the 192, and calculating the corresponding R_{measured}. Follow up those two measurements with a four-terminal resistance measurement just using the 192. (If R_{measured} is substantially larger than $\sim .2 \ \Omega$, confirm that you have firm, low-resistance contacts between the socket and the tube. Working the socket-tube contact may be required to find the lowest resistance spot.)

Maximum Filament Current, 2420 Voltage Compliance Limit

Following your determination of R_{support} in a room temperature tube, check out the conditions required to stay just below the tube's maximum ratings (4.75 V, 2.5 A). Using the 2420, successively source filament currents of 2.0 A, 2.1 A, 2.2 A, . . . to directly determine the maximum current you can use without exceeding the 4.75 V limit across the tube's filament. Note that at just below maximum conditions, the 2420 will probably need to produce voltages above 4.75 V (because of the resistance of the external wires: the voltage drop across the connecting wires is not zero). Record the maximum 2420 voltage and tube current allowed by the tube's ratings; you will need these numbers in step #2 of your computer program.

Data Collection Plan

You will be collecting two types of data at the same time: thermal characteristics of the filament and the thermionic properties of the tube (anode current/voltage relationship). Starting at a filament current of 0.9 A, increase the current flowing through the filament in steps of 0.1 A to a maximum current (found as described above, about 2.4 A) and then reverse those steps down to a filament current 1.0 A. The up-sweep in filament current followed by the down-sweep will allow you to test for hysteresis. At each step in current, allow the filament to approach thermal equilibrium (wait, say, 15 seconds) and then measure the voltage across and current through the cathode/anode. Calculate filament temperature two ways (Equations (5.56) and (5.62)). Average the two to estimate *the* temperature, and use half the absolute value of the difference to estimate the uncertainty.

You see above a classic example of systematic error. The temperature is measured two different ways. Direct application of error propagation formulas to these temperatures calculated from 6-digit meter values would suggest small uncertainties. However the two temperatures in fact disagree. If only one method had been used to measure temperature, we would have badly underestimated the error.

T^4 vs. Power: Testing Stefan-Boltzmann

By conservation of energy we know that the power dumped into the filament (mostly from electrical heating, but also from other sources like radiation from the room temperature environment to the filament) should equal the power out of the filament (from black-body radiation and other factors like conduction down the supports). Thus:

$$\epsilon_T \sigma \mathcal{A} T^4 = I^2 R_W + \text{ constant} \tag{5.64}$$

$$T^4 = \frac{1}{\epsilon_T \sigma \mathcal{A}} I^2 R_W + \text{ constant} \tag{5.65}$$

$$y = bx + a \tag{5.66}$$

A graph of T^4 vs. power should be a straight line from which you will determine ϵ_T. (Note that the error in power is quite small, so we have properly used it as an x variable.) In order to test this relationship you will want to make a file containing the filament power, T^4 (use the average of the two temperatures: $(T_i^4 + T_r^4)/2$), and the error in T^4 (use half the difference from the two temperatures: $|T_i^4 - T_r^4|/2$).

I_A vs. V_A: Testing Child and Richardson

You will collect anode current vs. voltage curves for each different filament (cathode) temperature. Use the Keithley 2400 to sweep the anode voltage logarithmically from 2 V to 120 V. (Note the maximum limits for the anode: 0.055 A or 125 V. Do not exceed either!) According to Child's law, the anode current, I_A, should increase in proportion to $V_A^{\frac{3}{2}}$. Of course, at sufficiently high voltage the current will be limited by the maximum electron evaporation rate, and a current plateau forms at a level given by Richardson's law. At the maximum filament current (corresponding to the maximum filament temperature and evaporation rate), plateau formation occurs at very high voltage and you have the longest run of data following Child's law. Make a file containing V_A, I_A, and δI_A which you can use to fit to the Child's law functional form:

$$I_A = k_1 \left(V_A - k_2\right)^{\frac{3}{2}} \tag{5.67}$$

In addition, you will want to make a big continuous file containing: V_A, I_A at every temperature tested. The current plateaus can be extracted from this file and fit to the Richardson relationship:

$$I_A = k_1 \, \mathcal{A} T^2 \, e^{-k_2/T} \tag{5.68}$$

Computer Data Collection

As part of this experiment you will write a computer program to control the experiment. *Plagiarism Warning*: like all lab work, this program is to be your own work! Programs strikingly similar to previous programs will alarm the grader. I understand that this will often be a difficult and new experience. Please consult with me as you write the program, and test the program (with tube disconnected!) before attempting a final data-collecting run.

Your program will control all aspects of data collection. In particular it will:

0. Declare and define all variables.

1. Open (i.e., create integer nicknames—i.e., `iunit`—for) the enets `gpib1` and `gpib2`.

2. Initialize meters—requires knowing the gpib primary address—i.e., `iadd`—of the meter and the `iunit` it is attached to. Get the status of each meter after you have initialized it.

   (a) Each source-meter must be told the maximum voltage and current it may produce during the experiment. Initialize the 2400 (anode voltage/current) for the tube maximum ratings.

   (b) Initialize the 2420 (filament voltage/current) for the near tube-maximum ratings found above[11]. In the following I assume the current maximum is 2.4 A, but it may be different for your tube.

   (c) Initialize the 192 for autorange DC voltage measurements.

3. Open the files:

   (a) `filament.dat` (intended for: I_f, V_f, T_r, T_i of filament).

   (b) `stefanB.dat` (intended for: power, T^4, δT^4 of filament).

   (c) `VI.dat` (intended for: all V_A, I_A of anode, with comments ('!') for filament I_f, T_r, T_i).

   (d) `child.dat` (intended for: V_A, I_A, δI_A of anode at maximum filament current).

   (e) `child-.dat` (like above but intended for a downsweep of anode voltage).

   (f) `rich.dat` (intended for: T_r, T_i, I_A — i.e., the estimated temperatures and the corresponding maximum anode current for Richardson's Law).

4. Tell the 2420 source-meter to source a filament current of 0.9 A.

5. Let the system sleep for 60 seconds to approach thermal equilibrium.

6. Do a sequence of measurements where the filament temperature is sequentially *increased* (i.e., a temperature up-sweep) with filament currents ranging from 0.9 A to some maximum (e.g., 2.4 A) current in steps of 0.1 A.

   (a) Tell the 2420 source-meter to source a filament current (I_f).

   (b) Let the system sleep for 15 seconds to approach thermal equilibrium.

   (c) Request a logarithmic sweep of the anode voltage (controlled by the 2400 source-meter) from 2 V to 120 V. Receive the resulting arrays: V_A and I_A.

   (d) Turn off the anode voltage.

   (e) Repeat (a) thus receiving an updated version of the filament current (it will be very close to the requested current).

   (f) Read the 192 to get the filament voltage (V_f).

---

[11] See Hands-on Electrical Measurements, p. 122. Recall: the 2420 maximum voltage will need to be a bit above 4.75 V. If you have not yet completed those measurements, temporarily initialize with 4.75 V.

(g) Calculate R_{measured} from the voltage across the the filament and the current through the filament. Calculate[12] the resistance of the (hot) tungsten filament, R_W, from R_{measured} and R_{support}. Finally following Eqs. (5.56), calculate T_r using x (the ratio of R_W to the previously calculated filament resistance at 293 K). Remark: typically you do not know the numerical value of R_{support}, etc. when you are writing this code. Simply define needed variables with dummy values to be replaced when you complete the 'Hands-on Electrical Measurements' described on page 122.

(h) Calculate T_i from Eq. (5.62)

(i) Write a line to the file `filament.dat` reporting: I_f, V_f, T_r, T_i.

(j) Write a line to the file `stefanB.dat` reporting: filament power $(R_W I_f^2)$, T^4, and δT^4 (see p. 124).

(k) Write a comment line (i.e., starts with '!') to the file `VI.dat` reporting filament data: I_f, T_r, T_i.

(l) Write the anode sweep data (one data-pair per line) in the file `VI.dat`.

(m) Write a line to the file `rich.dat` reporting T_r, T_i, I_A. Use I_A at V_A=120 V as the estimated plateau current. (When the experiment is over, you will need estimate δI_A based on hysteresis, and may need to delete I_A values if, for example, the current did not plateau or if cold emission substantially added to the plateau current.)

(n) Increment the filament current by 0.1 A and return to (a).

7. Collect data for Child's Law. Begin by repeating all the steps (a)–(m) outlined in 6 above using the maximum filament current. In addition to (l), write the anode sweep data $(V_A, I_A, \delta I_A)$ in the file `child.dat` (one data-triplet per line). In this case, δI_A will be calculated from the manufacturer's specs: percent+digits; the fortran function `eAk2400` can do this error calculation automatically. Now check for hysteresis by doing a reverse anode sweep: from 120 V down to 2 V. Write this reverse anode sweep data $(V_A, I_A, \delta I_A)$ in the file `child-.dat`.

8. Do a sequence of measurements where the filament temperature is *decreased* (i.e., a temperature down-sweep) by sequentially sourcing filament currents from one step down from maximum (e.g., 2.3 A) back to 1.0 A. Follow steps (a)–(m) used in the temperature up-sweep (part 6 above) and then:

(n) decrement the filament current by 0.1 A and return to (a).

9. Turn off the output of the 2420.

10. Close all files.

Note that the 0.9 A filament current data is just trash collected so the 1.0 A filament current data is taken on a pre-warmed filament.

---

[12]See page 122
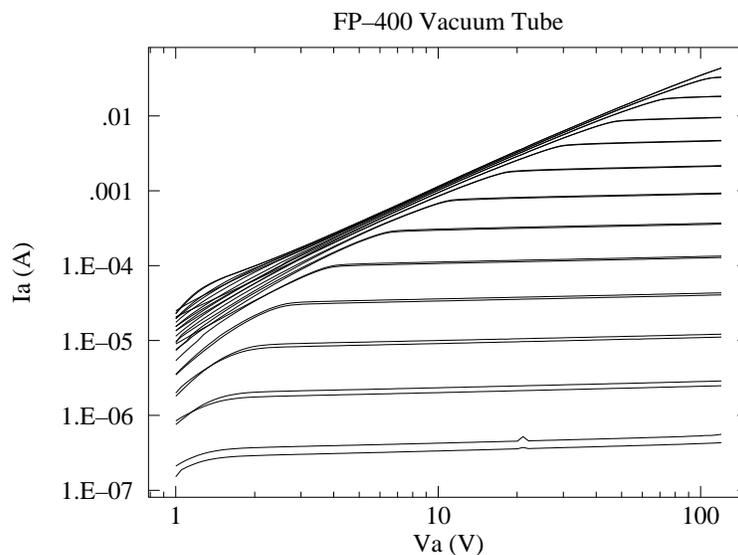
FP–400 Vacuum Tube

Figure 5.8: The temperature dependence of thermionic emission in a FP-400 vacuum tube. Each curve shows the anode current-voltage relationship at a particular filament temperature. At high filament temperatures and low anode voltages the anode current follows Child's law: an upward sloping straight line on this log-log plot. At sufficiently high anode voltage, the filament current plateaus at a value given by Richardson's law. At low filament temperatures and high anode voltages, "cold emission" may augment (distort) the plateau.

Observations

While the computer collects the data observe the light from the filament. (There is a $\frac{1}{16}$" diameter hole in the anode allowing light from the mid-point of the filament to escape.) Initially the filament will not appear to be incandescent (i.e., not a source of light at all: dark) so it may help to turn off the lab lights to better observe the beginning of visible incandescence. According to the Stefan-Boltzmann law the light energy radiated depends on T^4, so small changes in T produce big changes in light intensity. In addition to changes in light intensity, you should look for the more subtle change in light color: from a dull red to a brilliant yellow. Record your observations! At what filament current/temperature did you first see the filament producing light?

Data Analysis

Beginnings

The main result of this experiment is a plot similar to Figure 5.8 showing the anode current-voltage relationship at various filament temperatures. Production of such a multi-plot is a

bit complex, and you will almost certainly need further detailed instructions[13] from your instructor on using the program `Nplot`, so delay making this plot until your instructor is available. Each curve in the multi-plot represents a anode voltage sweep, which for small anode voltage V_A follows the nearly power-law relationship discovered by Child (power laws look linear in the log-log plot) but then levels out ('plateaus') at a current given by Richardson's Law. The curves are paired: each filament current was done once when the temperature was being increase and once when the temperature was being decreased.

The first check you should make is: is a proper plateau achieved for every filament current? If the high-temperature VI sweep reaches a plateau, then Child's Law will not apply at high V_A so `child.dat` will require editing; if it does not reach a plateau, then Richardson's Law does not apply to that sweep so `rich.dat` will require editing.

The anode current, I_A at the maximum anode voltage $V_A = 120$ V has been stored in the file `rich.dat` as a candidate plateau along with temperature estimates T_r and T_i (all currently lacking error estimates). While the actual Richardson's Law relationship (Eq. 5.69) is not on the list of $WAPP^+$ functions, we can arrange a quick but useful approximation.

According to Richardson's law, these plateau currents should satisfy:

$$I_A = \mathcal{A}AT^2\, e^{-W/kT} = k_1\, \mathcal{A}T^2\, e^{-k_2/T} \tag{5.69}$$

where $\mathcal{A}$ is the tungsten filament surface area. Clearly the largest (if unknown) error is in T, so the standard approach would be to put T on the y-axis and I on the x-axis: the opposite of what is implied by the above equation. However, we can't simply solve the above equation for T without some approximations. It turns out that $e^{-k_2/T}$ is the significant factor in the above equation, so we start by ignoring the T^2 and assume:

$$I = K\, e^{-k_2/T} \tag{5.70}$$

for some constant K. Now if we take $\log_e$ of both sides:

$$\log(I) = \log(K) - k_2\, \frac{1}{T} \tag{5.71}$$

or

$$\frac{1}{T} = \log(K)/k_2 - \frac{1}{k_2}\, \log(I) \tag{5.72}$$

This equation is now in a form[14] known to $WAPP^+$. Thus you can quickly $WAPP^+(I,T)$ data from `rich.dat` (of course still lacking x-errors and y-errors) and find an approximating curve. Retain hardcopy[15] of the fit results. Request (and hardcopy) a linearized plot of your data with x-scale: log and y-scale: inverse as in Fig. 5.9. This should allow you to check for aberrant data points. (Usually the high temperature curve has not plateaued, and so the high temperature data point must be removed. Occasionally the low temperature data points are also aberrant.)

---

[13]Documentation on the `plot` program can be found in Appendix B. It is also worth noting that since a log-log plot is requested, negative values—commonly negative anode currents—must be removed from the file `VI.dat`. As a general rule: do not destroy your original data, in this case simply rename the edited file something like `VI2.dat`.

[14]Inverse-Natural Log: $1/y = A + B\log(x)$

[15]You should retain a hard and/or soft copy of every successful fit. Hardcopy ends up taped into your lab notebook. I retain softcopy by copy&paste into a generic text file, e.g., using `kwrite`.

Work Function Systematic Error
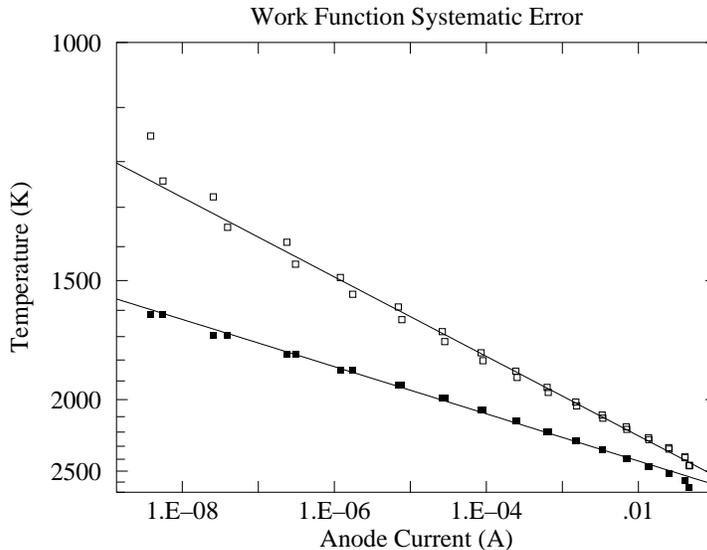
Figure 5.9: Simplified analysis suggests Richardson's Law data can be fit to the Inverse-Natural Log relationship of Eq. 5.72. The filled-square points use $y = T_i$, the unfilled-square points use $y = T_r$. Systematically different temperature measurements yield systematically different B (slopes in the above plot) and hence systematically different work functions $W = -k/B$. Note the inexact pairing of the data due to temperature hysteresis. Note that the rightmost datapoint is aberrant: a non-plateau.

But what should be used for T?... You have two values (T_i and T_r) and they are not the same! The answer is both: fit/plot once with T_i and then again with T_r. The systematically different temperature scales will result in different fit parameters— in particular you will get an upper-limit and lower-limit for the work function from the two B values of your two fits, as $W/k = k_2$ and $k_2 \approx -1/B$. The work function is an atomic quantity, and it is usually expressed in the atomic scale unit eV[16]. δW can then be taken as half the difference between these two extreme values. (Note an oddity: you've determined a value for δW, but not yet found the best value for W.) Take this opportunity to delete the aberrant data from `rich.dat` (and, say, save as `rich2.dat`).

We can also a quick but useful check of the Child's law data using $WAPP^+$. Once again the proper functional form (Eq 5.76) is not available in $WAPP^+$, but a power law is somewhat similar. In this case the files `child.dat` and `child-.dat` include a y-error estimate based on the k2400 specifications. (In fact these book-based errors very much underestimate the deviations you will experience: expect a horrendous reduced χ^2). Check that $WAPP^+$'s B is nearly 1.5, record the A value (you will need it as an initial estimate for k_1 when you do the fit to the proper functional form), make a log-log plot (no hardcopy yet!) to check for aberrations (like a plateau at large V_A).

Finally, the Stefan-Boltzmann Law *can* be properly evaluated using $WAPP^+$, as a line is the expected function. The file `stefanB.dat` has the required data with y-errors (x-errors are small). Produce a plot (and hardcopy fit report) similar to Fig. 5.10. I expect you'll find

---

[16] $1 \, eV = 1.6022 \times 10^{-19}$ J is the energy an electron gains in going through a potential difference of 1 V.

Stefan–Boltzmann Law
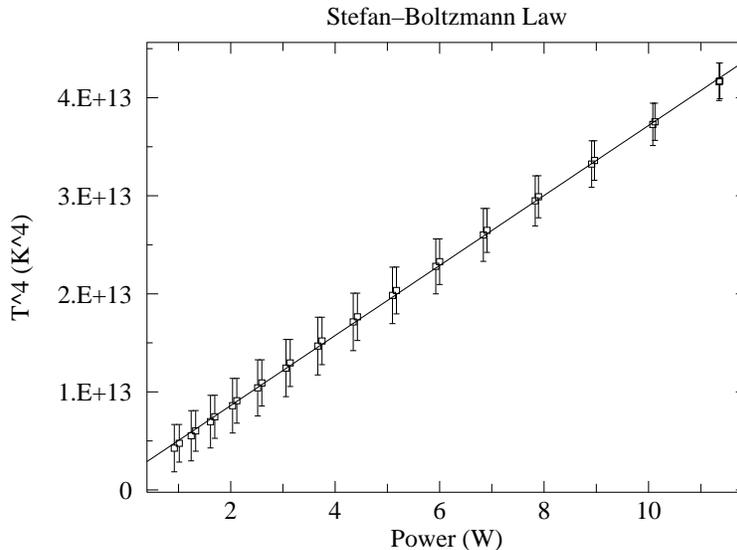
Figure 5.10: A test of the Stefan-Boltzmann law: that power radiated is proportional to T^4. Note that the fit line hits well within each error bar. The χ^2 for this fit will be "small". Evidently the average temperature is a better measure of temperature than you might expect from the deviations between T_i and T_r.

a small reduced χ^2 due to the large systematic error in temperature, so a bit of additional work will be required to estimate $\delta\epsilon_T$.

Stefan-Boltzmann Law II

You should have already checked for an approximately linear relationship between electrical power in and T^4:

$$\text{Power} = \epsilon_T \sigma \mathcal{A} T^4 \tag{5.73}$$

and found a reduced χ^2 indicative of large systematic uncertainty in temperature. We now seek an estimate for ϵ_T (with error) at the highest temperature. The best value for ϵ_T can be found by plugging in our best estimates for T^4 and the measured power (found in the file `stefanB.dat`), and $\mathcal{A}$ (calculated based on the dimensions recorded in Table 5.1). Alternatively ϵ_T could be calculated based on slope as suggested by Eq. 5.66. But how should we incorporate the large systematic errors in T^4 and the unknown systematic error in $\mathcal{A}$? For the surface area $\mathcal{A}$, all we know is the 'book' values for the dimensions of the filament. Based on the sigfigs in the reported values, I estimate:

$$\frac{\delta\mathcal{A}}{\mathcal{A}} \approx 10\% \tag{5.74}$$

(mostly due to the filament diameter, where a 'small' uncertainty leads to a large fractional uncertainty). We can then use the 'high-low' method or use the proper formula in Appendix E to estimate the range of possible values for ϵ_T, given the range of possible values for T^4 and $\mathcal{A}$ (assume the error in power is negligible).
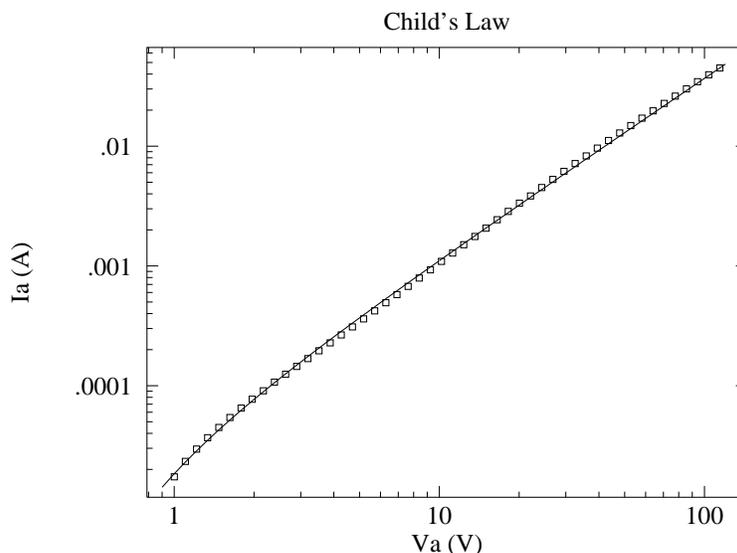
Figure 5.11: A plot of the space-charge limited thermionic emission in a FP-400 vacuum tube (Child's law). The data was taken at a filament current of 2.4 A. Every other data point has been eliminated so the fit line is not obscured. Note that the fit line systematically misses the data, sometimes a bit high others a bit low. The measurement errors are tiny, so these small misses do result in a "too-large" χ^2. Nevertheless, the law provides an excellent summary of the data over a huge range of variation.

Child's Law II

At sufficiently high filament temperature (and low anode voltage), Child's law governs the anode current-voltage relationship:

$$I_A = \frac{8\pi\epsilon_0 \ell \sqrt{\frac{2e}{m}}}{9b\beta^2} \; V_A^{\frac{3}{2}} \tag{5.75}$$

(see vicinity of Eq. 5.40, page 115, for a definition of symbols) At the highest filament temperature (e.g., filament current of 2.4 A) you have saved the $(V_A, I_A, \delta I_A)$ data in the file `child.dat`. Now `fit`[17] this data to the functional form:

$$I_A = k_1 \; (V_A - k_2)^{\frac{3}{2}} \tag{5.76}$$

(where k_2 represents a small offset between ground and the actual average voltage of the filament). The program `fit` will require an initial guess for k_1; use the $WAPP^+$ value for A found above. (In `fit` the default initial guess for parameters is zero; that is reasonable for k_2.) Do not be surprised if you get a huge reduced χ^2. Find an estimate for k_1 error either by a 'fudged fit' or a bootstrap. As always retain a copy of your full fit report, including the covariance matrix. Plot your data and best-fit function using the program `plot`, producing a result similar to Fig. 5.11.

---

[17]The program `fit` is documented in Appendix A. Note that plotting and fitting now are handled by two different programs, not a combined program like $WAPP^+$.
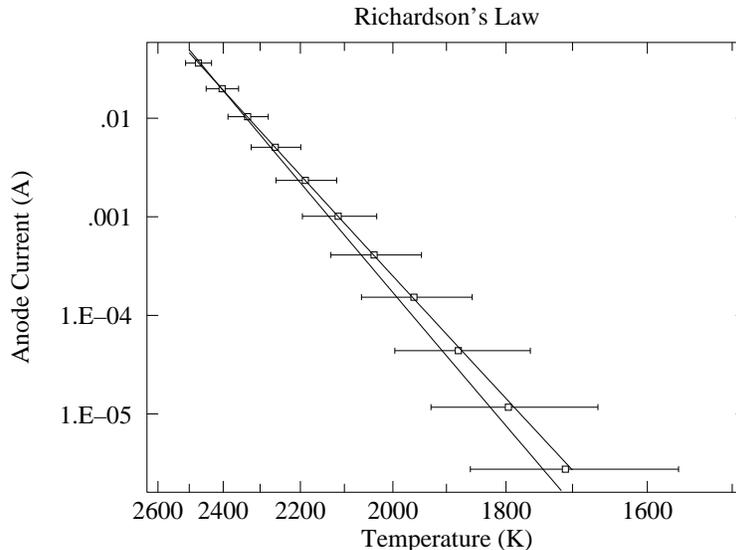
Figure 5.12: A Richardson Plot of thermionic emission in a FP-400 vacuum tube. Each data point displays the plateau anode current at a particular filament temperature. The curve through the data fits for the work function; the slightly steeper curve uses the book value for the work function.

Follow exactly the same process for the downsweep data in the file `child-.dat`. Often you will find that the k_1 values for the two sweeps differ by more than computer-based value of δk_1. Systematic error (here a type of hysteresis) is responsible. Note that the usual reduced χ^2 alerts us to a problem, but measuring twice (in different ways) provides an estimate (perhaps still an under-estimate) for δk_1: half the difference between the two values of k_1.

We expect that

$$k_1 = \frac{8\pi\epsilon_0 \ell \sqrt{\frac{2e}{m}}}{9b\beta^2} \tag{5.77}$$

so using the tube dimensions in Table 5.1, the electron charge-mass ratio e/m can be calculated. Since the FP-400 is a finite length cylinder (rather than the infinite cylinder discussed in our theory) use the effective length[18] $= 0.7 \times \ell$ as the length of the cylinder. But what can we use as errors for the 'book' values for b and ℓ? Thermal expansion of the filament should, by itself, change ℓ by about 1% (notice the spring-tensioned support for the filament in the FP-400). Based on the sigfigs in the reported values, I estimate:

$$\frac{\delta(b/\ell)}{(b/\ell)} \approx 3\% \tag{5.78}$$

Calculate e/m and its error. Compare to the 'known' value (citation required!).

---

[18]This effective length corrects for current loss through the ends of the cylinder under space-charge situations. A smaller correction should be applied when large anode voltages are designed to sweep up all evaporated electrons, i.e., for Richardson's Law, where 90% electron collection seems appropriate. The details of handling such corrections to theory may be found in reference 1.

Richardson-Dushman Law II

You should have already checked for an approximately exponential-inverse relationship between T and I_A, edited out aberrant data (e.g., not yet plateaued), and have a range of possible values (due to systematic error in temperature) for k_2 in the expected relationship:

$$I_A = k_1 \, \mathcal{A}T^2 \, e^{-k_2/T} \tag{5.79}$$

We now seek a treatment incorporating the hysteresis error in I_A and the proper functional form, to obtain the best possible value for k_2. We will need to manipulate[19] the data in `rich2.dat` to bring together equivalent data from the temperature upsweep and downsweep. An easy way to get the data into the `gnumeric` spreadsheet, is to type it to the screen using the linux command[20] `cat`:

`cat rich2.dat`

You can then copy and paste[21] this data into `gnumeric`. Our aim to put together data with the same filament current, i.e., to combine the first line of data with the last; to combine the second line of data with the next-to-last; etc. This is easily accomplished by typing the data bottom-to-top to the screen using the linux command[22] `tac`:

`tac rich2.dat`

The results can be copy and pasted next to the previous data so that the data we aim to combine (e.g., both $I_f = 1$ A) is on the same line. The best estimate for T is the average of the four Ts (the two T_i should be nearly identical); the best estimate for I_A is the average of the two I_As; for δI_A use half the difference between the two values of I_A ($|I_{A1} - I_{A2}|/2$); for δT use half the difference between T_r and T_i (average the two differences). The result of this 'data reduction' is half as many data points, but with a value for δI_A based on hysteresis. (The alternative for δI_A would be the meter manufacturer's specifications, and we know they are way to small from analysis of Child's Law.)

Report the relative importance of hysteresis and calibration in temperature uncertainty determination. For $I_f = 1.2$ A record the difference between the two T_r due to hysteresis and the difference between T_r and T_i which is a temperature calibration uncertainty.

Copy and paste this reduced data into a new file, and fit it to Eq. 5.79. (As always retain a copy of the full fit report.) You should have an estimate for k_2 from *WAPP+*; use the book[23] value for A, 0.72×10^6 A/m^{2}K^2, as an initial guess for k_1. Produce a plot similar to Fig. 5.12; include lines both for your best fit and the 'book' values of k_1 and k_2.

The work function is an atomic quantity, and it is usually expressed in the atomic scale unit eV[24]. Calculate the work function from your value of k_2 in joules and eV and compare to the book[23] value of 4.5 eV.

Since our temperatures are so uncertain, particularly at the low end, the best estimate for

---

[19]This processing could very easily have been done within the program itself. I've instead opted to make the program as simple as possible at a cost of additional 'by-hand' processing in a spreadsheet.

[20]from con<u>cat</u>enate—commonly this command is used to combine several files

[21]Note use of "See two two separators as one": Alt-e

[22]clever or what?

[23]Blakemore, *Solid State Physics*, p. 191

[24]1 eV = 1.6022×10^{-19} J is the energy an electron gains in going through a potential difference of 1 V.

the Richardson constant A comes from substituting the book value of the work function and the plateau (T, I_A) measurement for the highest valid filament temperature into Eq. 5.69. We can then estimate the systematic uncertainty in A by using the 'high-low' method of 191 $(A^+ \to I^+, T^-, \mathcal{A}^-)$.

Report Checklist

1. Write an introductory paragraph describing the basic physics behind this experiment. Explain why higher cathode currents produced ever higher plateaus of anode current. Explain why higher anode voltages produced higher anode current and why the curves of Fig. 5.8 seem to bunch up (coincide) on the left hand side.. (This manual has many pages on these topics; your job is condense this into a few sentences and no equations.)

2. Calculations (no errors) of room temperature R_W. Measurements (4-wire ohmmeter and direct voltage/current) of R_{measured} at room temperature. Calculation of R_{support}.

3. Observations of the light intensity and color as a function of filament temperature. Comments on temperature hystersis/calibration (see p. 133).

4. Data files and computer program: Leave them in your linux account. Print out a copy of your program, the file `filament.dat`, and the data you used to fit Richardson's Law; Tape them into your lab notebook. Document how your Richardson's Law data was calculated in your spreadsheet. (I would self-document the spreadsheet, then File▶Print Area▶Set Print Area to select just the relevant 4 columns, including the self-documentation of those columns, and then print.)

5. Plots similar to Figures 5.8, 5.12 (with `fit` curve Eq. 5.68 and also the Richardson function using "book" values for k_1 and k_2), 5.11 (two plots: `child.dat` and `child-.dat` each with `fit` curve Eq. 5.67), 5.10 (with line Eq. 5.66) and 5.9 (two separate $WAPP^+$plots one with T_r data and the other with T_i data). Note that Figure 5.8 is complex to produce. Use a file of `Nplot` commands and feel free to talk to me about how to go about doing this. Carefully note the use of log and inverse scales (which requires positive data—edit to achieve this!). <u>Include a fit report for each fit curve.</u>

6. Experimental values (with error range) for: W (in eV) and A. See page 133 for calculating W, A, and δA; see page 129 for calculation of δW. Experimental values (with error range) for: e/m and ϵ_T.

7. Show the steps connecting Equations (5.28)–(5.30) to Equation (5.31). Show that $f = A\rho^\alpha$ is a solution to (5.31) if and only if A and α are as reported in Equation (5.36). Substitute the $\rho \to 1$ approximation (Eq. 5.35) into the differential equation (5.31). Show that while we do not have an exact solution to the differential equation, the singular parts (i.e., those that approach infinity as $\rho \to 1$) cancel.

8. Make a final results table, recording your numerical results from #6 (with proper units and sigfigs) adjacent to the corresponding 'book' values. Don't forget to record your tube's identifying letter!

Comment: Classical vs. Quantum

I said above that the presence of an $\hbar$ in Richardson's constant shows that the process is governed by quantum mechanics. It is not quite so simple. The evaporation of water is not fundamentally different from that of electrons, while the former (due to the larger mass of a water molecule) is well-approximated by a classical calculation. Classical evaporation can be calculated much as quantum evaporation using the Maxwell-Boltzmann speed distribution and the number density (of water molecules) rather than the disguised version of this: Fermi energy ($E_F^{\frac{3}{2}} \propto$ number density). We can write the classical rate in terms of the quantum with no $\hbar$ visible:

$$\text{classical flux} = \frac{4}{3\sqrt{\pi}} \left[\frac{E_F}{kT}\right]^{\frac{3}{2}} \times \text{quantum flux} \tag{5.80}$$

The different temperature dependence[25] for the classical flux ($T^{\frac{1}{2}}e^{-W/kT}$ vs. $T^2 e^{-W/kT}$) cannot be detected experimentally: the Boltzmann factor overwhelms all other temperature dependencies. On the other hand, since $E_F \gg kT$, the expected classical rate is much larger than the quantum rate. This played a role in the mistaken idea that thermionic emission was due to surface contamination: the experimental rate seemed too small to be classical evaporation. On the other hand a more fruitful interpretation of the "low" rate was that only a fraction ($\sim kT/E_F$) of the electrons were thermally active. This connected with other observations (like the "small" specific heat of the electron gas) and provided a link to the idea of a degenerate Fermi gas.

Comment: Uncertainty

Inspection of Figures 5.8–5.12 shows that something "funny" is going on. (I can't say "abnormal" or "unusual" as it is neither.) Figure 5.10 shows the unmistakable signs of "small reduced χ^2": The fitted line goes nearly dead-center through all 30 error bars, never coming even close to an edge. For error bars sized at one standard deviation (σ), you should expect total misses of the error bar about $1/3$ of the time. In addition recall that each data point is really a double: the same filament current was sourced as part of a temperature up-sweep and as part of a temperature down-sweep. These repeated measurement should also frequently miss by a standard deviation, but here they are so close that the two points often look like just one. The answer to this puzzle is that the error bars are not displaying statistical ('random') error. Instead the temperature was measured two different ways (T_i and T_r), and the error bar represented the deviation between these two measurement methods. When different methods of measurement produce different answers for the same quantity, we have a textbook example of *systematic error* (in contrast to statistical error). Notice that if we had used the statistical deviation of just one measure of temperature, we would seriously underestimated the error. Furthermore since quite accurately measured electrical quantities were used to calculate the temperature (via Equation 5.62 or Equation 5.56), application of the usual error propagation methods would also have produced very small errors in T. The full range of our uncertainty in temperature is made evident only by measuring T two different ways. (This is typically how systematic errors are detected.)

---

[25]Saul Dushman (Phys. Rev. **21**, 623–636 (1923)), while working at G.E., provided a general thermodynamic argument for the T^2 dependence and the universal nature of A. The resulting relationship is therefore sometimes known as the *Richardson-Dushman* relationship.

Having detected systematic errors, we should seek an explanation...In addition to the problems previously cited (particularly use of book values for filament dimensions and the problems associated with R_{support}), nonuniform filament temperature due to filament supports may be the problem. Koller (p. 89) reports the filament temperature 0.5 cm from a support is reduced by 15% (and of course the effect is greater closer to the support). Thermionic emission is reduced a similar amount 1.3 cm from a support. Thus the quantity we are seeking (<u>a</u> filament temperature) does not even exist, so it is hardly surprising that different measurements give different results. (It turns out the T_r is something like the average temperature; whereas T_i is something like the central temperature.) These effects have been investigated, and Koller gives the theory to correct such measurements, but such considerations are beyond the scope of this experiment.

Figure 5.8 shows the opposite problem, "large reduced χ^2": The fitted line systematically misses almost every error bar. In this case, the miss might be called "small", but the error bar is smaller still. Once again, errors were not calculated statistically (manufacturer's specifications were used), so "reduced $\chi^2 = 1$" should not really be expected. In this case, my guess is that the problem is with our simplified theory (we assumed: infinite cylinders, no random component to the electron velocity [zero electron temperature], uniform filament [temperature, voltage, emissivity, ...], perfect vacuum, no incipient current plateau). We could of course test these hypotheses by further experiments with different tubes, but such work is beyond the scope of this experiment. (Indeed you most likely have detected V_A-sweep hysteresis; correction for this dramatically reduces reduced χ^2, but not all the way to ~ 1.)

Summary: Very large or very small reduced χ^2 suggests significant non-statistical errors, a very common —perhaps even the usual— situation. Computer generated errors are some sort of none sense in this circumstance. Presumably your theory and/or measurement methods are less than perfect. That is, of course, No Surprise. Science next requires you to guess and perhaps further investigate which imperfections are the leading source of the problems, i.e., what changes to the experiment would ameliorate the problem.

References

1. Jones & Langmuir *GE Review*, The Characteristics of Tungsten Filaments as Functions of Temperature **30** (1927) Part I pp. 310–19, Part II pp. 354–61, Part III pp. 408–12

2. Melissinos *Experiments in Modern Physics*, 1966, pp. 65–80

3. Preston & Dietz *The Art of Experimental Physics*, 1991, pp. 141–47, 152–61

4. Blakemore *Solid State Physics*, 1974, pp. 188–95

5. Koller *The Physics of Electron Tubes*, 1937, Ch. I, VI, VIII

6. G.E. *FP-400 Description and Rating* FP400.pdf found at `http://www.tubedata.org`

YESTERDAY IS HISTORY
words by Dr. James D. Livingston
tune: Yesterday, by Paul McCartney

Yesterday — when old classical ideas held sway
Drude and Maxwell-Boltzmann had their say
And we believed them yesterday

Suddenly — quantum free-electron theory
Introduced the Fermi energy
And yesterday was history

Fer-mi theory you'll agree has much to say
We see Drude was wrong — he belongs to yesterday

Density — of electron states provides the key
Multiplied by probability
Reveals electron properties
Yesterday is history
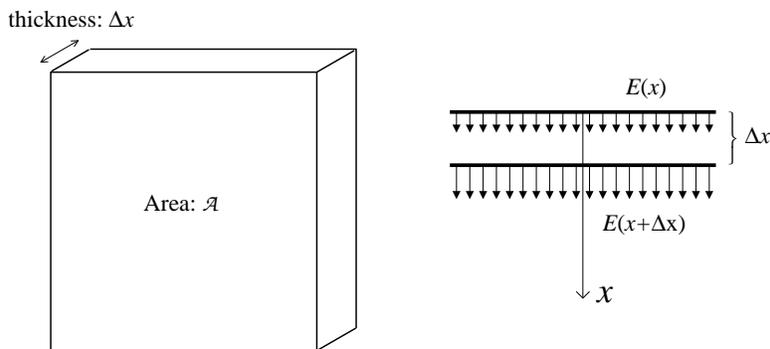
Figure 5.13: Gauss' Law is used to calculate the charge between two plates of area $\mathcal{A}$ separated by a distance Δx. Since (by assumption) the potential just depends on x, the electric field is in the x direction and is given by $E = -dV/dx$.

Appendix—Poisson's Equation

Equation 5.5 and Equation 5.28 made reference to "Poisson's Equation", which is really a topic for Physics 341, rather than part of Physics 200. In this appendix, Poisson's Equation is derived starting from two Physics 200 results: Gauss' Law: that the electric flux leaving a region depends just on the charge enclosed by that region:

$$\oint \vec{\mathbf{E}} \cdot \hat{\mathbf{n}} \, dA = Q_{\text{enclosed}}/\epsilon_0 \tag{5.81}$$

and the relationship between electric field and electric potential (voltage):

$$E_x = -\frac{dV}{dx} \tag{5.82}$$

Poisson's Equation is a differential equation equivalent to Gauss' Law. It is usually written in terms of the Laplacian (∇^2), which in turn can most easily be expressed in terms of second derivatives w.r.t. x, y, and z:

$$\nabla^2 V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2} = -\rho/\epsilon_0 \tag{5.83}$$

where ρ is the electric charge density. We need Poisson's Equation only in cases where the electric potential depends on just one variable (x or cylindrical r), which simplifies the required proofs considerably.

As shown in Figure 5.13, if V is a function of x alone: $V(x)$, we can find the charge between two plates of area $\mathcal{A}$ using Gauss' Law:

$$Q_{\text{enclosed}} = \epsilon_0 \mathcal{A} \left(E(x + \Delta x) - E(x) \right) \approx \epsilon_0 \mathcal{A} \, \frac{dE}{dx} \, \Delta x \tag{5.84}$$

Thus the charge density between the plates is given by:

$$\rho = \frac{Q_{\text{enclosed}}}{\text{volume}} = \frac{\epsilon_0 \mathcal{A} \frac{dE}{dx} \Delta x}{\mathcal{A}\Delta x} = \epsilon_0 \, \frac{dE}{dx} = -\epsilon_0 \, \frac{d^2 V}{dx^2} \tag{5.85}$$
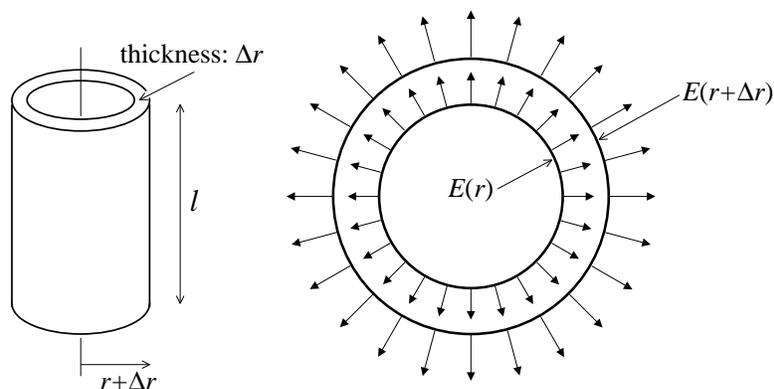
Figure 5.14: Gauss' Law is used to calculate the charge between two coaxial cylinders of length l separated by a distance Δr. Since (by assumption) the potential just depends on r, the electric field is in the r direction and is given by $E = -dV/dr$.

which provides what is needed for Equation 5.5.

As shown in Figure 5.14, if V is a function of r alone: $V(r)$, we can find the charge between two coaxial cylinders using Gauss' Law:

$$
\begin{aligned}
Q_{\text{enclosed}} &= \epsilon_0 l \ \{2\pi(r + \Delta r)E(r + \Delta r) - 2\pi r E(r)\} \\
&= \epsilon_0 l \ \{2\pi r(E(r + \Delta r) - E(r)) + 2\pi \Delta r E(r + \Delta r)\} \\
&\approx \epsilon_0 l \ \left\{2\pi r \frac{dE}{dr} + 2\pi E(r)\right\} \Delta r
\end{aligned}
\tag{5.86}
$$

Thus the charge density between the cylinders is given by:

$$
\rho = \frac{Q_{\text{enclosed}}}{\text{volume}} = \frac{\epsilon_0 l \ \{2\pi r \ dE/dr + 2\pi E(r)\} \Delta r}{2\pi r l \Delta r} = \epsilon_0 \left\{\frac{dE}{dr} + \frac{1}{r} E\right\} \tag{5.87}
$$

$$
= -\epsilon_0 \left\{\frac{d^2V}{dr^2} + \frac{1}{r} \frac{dV}{dr}\right\} \tag{5.88}
$$

which provides what is needed for Equation 5.28.

6: Scientific Communication

Poster Topics:

Presentation of a lab project as a poster is the final component of this course. I will be assigning you a topic; feel free to drop me email with a list of your preferences! (First come, first served. Choose a topic ASAP and avoid the crunch at finals!) Topics:

1. Photometry: T_B, T_R

2. Bubble Chamber: m_μ

3. Bubble Chamber: m_π

4. Thermionic Emission: Stefan-Boltzmann Law

5. Thermionic Emission: Richardson-Dushman Law

6. Thermionic Emission: Child-Langmuir Law

7. Langmuir Probe: T_e

8. Langmuir Probe: n_e

9. Langmuir Probe: V_f, V_p

Poster Basics:

size: maximum width: 6 feet; maximum height: 4 feet

fonts: title: 30–45 mm (160 pt)
authors and affiliations: 25–30 mm (96 pt)
main headings: 10 mm (48 pt)
subheadings: 8 mm (32 pt)
text: 5 mm (24 pt)

sections: Title, Author(s) with affiliation[1], Abstract, Introduction, Methods (not required for standard techniques; a block diagram of the apparatus is often

---

[1]I.e., school/location: Saint John's University, Collegeville, MN or College of Saint Benedict, St. Joseph, MN; a logo is also nice

helpful), Theory, Results and/or Discussion (if possible, use figures to display results; never display raw data or fit reports), Conclusions, References, Acknowledgments

first steps: Assemble your figures. Make sure that they convey your main points and can be understood from a distance. For each main section of your poster, try to boil the essential points down to about 350 words (which will, in the proper font, approximately fill an $8\frac{1}{2}$" $\times$ 11" sheet of paper). These section pages can then be assembled along with the figures to form a rough version of your poster. See the references below for book-length instructions on creating an effective poster.

prepare: a 1–2 minute 'talk' that explains how you collected you data and why the data you are displaying behaves as it does. (If your data displays 'unexpected' behavior, explain what was expected and how your data differs from the expected.) This talk is to serve as an introduction to your experiment for folks who are too lazy to read your poster or who have questions about your work.

hints:
- Focus on the main points as poster space is limited. Eliminate everything but the essentials.
- Posters are visual, so your figures must convey your main messages.
- There should be a clear path through the poster: typically left to right and top to bottom.
- Avoid unusual fonts and unnecessary font changes.
- Size fonts for viewing at a distance. (This is particularly important for visual elements like plot axes labels.)
- The focus of this course has been systematic error, so discussion of the systematic error in your project should be part of your poster.
- Properly report numbers: sigfigs, error, units!
- Do not plagiarize! Sentences/diagrams/photos grabbed from the web (or this manual) must be *cited* (not just listed at the end under References).
- Tables of raw data, fit reports, etc. do not generally belong on a poster: use graphs to display data.

warning: Two years ago Dr. Steck flustered many students with the question: "What is the basic physics behind your experiment?" Inability to explain simply, in a few sentences (and no equations), what is going on in your experiment puts your entire work in jeopardy as listeners figure if you don't understand the basics, you cannot be trusted on the details. You really must be fully prepared for such questions. Feel free to practice your explanation with me. Here, for example, are ∼200 words about Child's law:

> Child's law states that the current through the vacuum tube is proportional to the voltage drop across the tube to the three-halves

power. The current here consists of electrons in the 'vacuum' moving from cathode to anode because of the electric field caused by the voltage difference. It makes perfect sense that larger voltages (and hence larger forces on the electrons) result in faster electrons and hence more current. The fact that the power is exactly 1.5 is a result from dimensional analysis. The physics behind that dimensional analysis involves current conservation (that the same current must pass through any cylinder between cathode and anode) and 'space charge': the charge density (due to the presence of electrons) in the 'vacuum' modifies the electric field (which can be calculated via Poisson's equation). Current conservation means that the electron density is lowest where the electrons are moving fastest, so the highest space charge is near the cathode (i.e., before the electrons have moved very far 'downhill'). Child's Law is limited by the maximum possible rate of electron evaporation from the cathode: a temperature dependent effect known as Richardson's Law.

References

1. Briscoe, Mary Helen *Preparing Scientific Illustrations* 1996, Springer-Verlag, Q222.B75

2. Gosling, Peter J. *Scientist's Guide to Poster Presentations* 1999, Kluwer

3. `http://www.aspb.org/education/poster.cfm`

4. `http://www.stanford.edu/group/blocklab/dos and donts of poster presentation.pdf`

5. `http://www.sigmaxi.org/meetings/student/hints.shtml`

6. `http://www.asp.org/education/howto_onPosters.html`

7. `http://www.physics.csbsju.edu/370/poster_fonts.pdf`

7: Langmuir's Probe

Purpose

The purpose of this lab is to measure some basic properties of plasmas: electron temperature, number density and plasma potential.

Introduction

When you think of electrical conductors, you probably think first of metals. In metals the valence electrons are not bound to particular nuclei, rather they are free to move through the solid and thus conduct an electrical current. However, by far the most common electrical conductors in the Universe are *plasmas*: a term first applied to hot ionized gases by Irving Langmuir (see below). In conditions rare on the surface of the Earth but common in the Universe as a whole, "high" temperatures[1] provide enough energy to eject electrons from atoms. Thus a plasma consists of a gas of freely flying electrons, ions, and yet unionized atoms. It should come as no surprise that during the extraordinary conditions of the Big Bang, all the matter in the Universe was ionized. About 380,000 years after the Big Bang, the Universe cooled enough for the first atoms to form. Surprisingly about 400 million years after that the Universe was re-ionized, and the vast majority of the matter in the universe remains ionized today (13.7 billion years after the Big Bang). Some of this ionized matter is at high density (hydrogen gas more dense than lead) and extremely high temperature at the center of stars, but most of it is believed to be at extremely low density in the vast spaces between the galaxies.

Perhaps the most obvious characteristic of conductors (metals and plasmas) is that they are shiny; that is, they reflect light. A moment's thought should convince you that this is

---

[1] What does "high temperature" mean? When you are attempting to make a Bose condensation at less than a millionth of a degree, liquid helium at 4 K would be called hot. When you are calculating conditions a few minutes after the Big Bang, a temperature of a billion degrees Kelvin would be called cool. An important insight: Nothing that has units can be said to be big or small! Things that have units need to be compared to a "normal state" before they can be declared big or small. Here the normal state refers to conditions allowing normal solids and liquids to exist. Tungsten, which is commonly used in the filaments of light bulbs, melts at about 3700 K; Carbon does a bit better: 3800 K. The "surface" of the Sun at 6000 K has no solid bits. At temperatures of about 5000 K most molecules have decomposed into elements which in turn have partially "ionized": ejecting one or more electrons to leave a positively charged core (an ion) and free electrons. I'll pick 5000 K as my comparison point for "hot", but of course some elements (e.g., sodium) begin to ionize a lower temperatures, and others (e.g., helium) ionize at higher temperatures. The key factor determining the ionized fraction in the Saha equation is the "first ionization energy".

not an "all-or-nothing" property. While metals reflect radio waves (see satellite TV dishes), microwaves (see the inside of your microwave) and visible light, they do not reflect higher frequency light like X-rays (lead, not aluminum, for X-ray protection). The free electron number density (units: number of free electrons/m^3), n_e, determines the behavior of light in a plasma. (Almost always plasmas are electrically neutral; i.e., the net electric charge density ρ is near zero. If the atoms are all singly ionized, we can conclude that the ion number density, n_i, equals n_e. In this lab we will be dealing with partially ionized argon with a neutral atom density $n_n \gg n_e \approx n_i$.) The free electron number density determines the *plasma frequency*, ω_p:

$$\omega_p = 2\pi f_p = \sqrt{\frac{n_e e^2}{\epsilon_0 m}} \tag{7.1}$$

where $-e$ is the charge on an electron and m is the mass of an electron. If light with frequency f is aimed at a plasma: the light is reflected, if $f < f_p$; the light is transmitted, if $f > f_p$. Thus conductors are shiny only to light at frequencies below the plasma frequency. In order to reflect visible light, the plasma must be quite dense. Thus metals ($n_e \sim 10^{28}$ m^{-3}) look shiny, whereas semiconductors ($n_e \sim 10^{24}$ m^{-3}) do not. The plasma at "surface" of the Sun (with ionized fraction less than 0.1% and $n_e \sim 10^{20}$ m^{-3}) would also not look shiny. You will find the plasma used in this lab has even lower n_e; it will look transparent.

The defining characteristic of conductors (metals and plasmas) is that they can conduct an electric current. Since conductors conduct, they are usually at a nearly constant potential (voltage). (If potential differences exist, the resulting electric field will automatically direct current flow to erase the charge imbalance giving rise to the potential difference.) At first thought this combination (big current flow with nearly zero potential difference) may sound odd, but it is exactly what small resistance suggests. In fact the detection of big currents (through the magnetic field produced) first lead to the suggestion[2] of a conductor surrounding the Earth—an ionosphere. Edward Appleton (1924) confirmed the existence and location of this plasma layer by bouncing radio waves (supplied by a B.B.C. transmitter) off of it. In effect the ionosphere was the first object detected by radar. Appleton's early work with radar is credited with allowing development of radar in England just in time for the 1941 Battle of Britain. Appleton received the Nobel prize for his discovery of the ionosphere in 1947. (Much the same work was completed in this country just a bit later by Breit & Tuve.)

The plasma frequency in the ionosphere is around 3–10 MHz (corresponding to $n_e \sim 10^{11}$–10^{12} m^{-3}). Thus AM radio (at 1 MHz) can bounce to great distances, whereas CB radio (at 27 MHz) and FM (at 100 MHz) are limited to line-of-sight. (And of course when you look straight up you don't see yourself reflected in the ionospheric mirror, as $f_{\text{light}} \sim 5 \times 10^{14}$ Hz. On the other hand, extra terrestrials might listen to FM and TV, but we don't have to worry about them listening to AM radio.) The actual situation is a bit more complex. In the lowest layer of the ionosphere (D region), the fractional ionization is so low that AM radio is more absorbed than reflected. Sunlight powers the creation of new ions in the ionosphere, so when the Sun does down, ionization stops but recombination continues. In neutral-oxygen-rich plasmas like the D region, the plasma disappears without sunlight. Higher up in the ionosphere (the F region, where n_e is higher and n_n lower) total

---

[2]Faraday (1832), Gauss (1839), Kelvin (1860) all had ideas along this line, but the hypothesis is usually identified with the Scot Balfour Stewart (1882).

recombination takes much more than 12 hours, so the plasma persists through the night. Thus AM radio gets big bounces only at night.

I have located a plasma (a *hot* ionized gas) high in the Earth's atmosphere, yet folks climbing Mt. Everest think it's cold high up in the Earth's atmosphere. First, the ionosphere starts roughly 10 times higher than Mt. Everest, and in the F Region (about 200 km up) "temperature" is approaching 1000 K, warm by human standards if not by plasma standards. But the electrons are hotter still... up to three times hotter (still not quite hot by plasma standards). This is an odd thought: in an intimate mixture of electrons, ions, and neutral atoms, each has a different temperature. As you know, in a gas at equilibrium the particles (of mass M) have a particular distribution of speeds (derived by Maxwell and Boltzmann) in which the average translational kinetic energy, $\langle E_K \rangle$ is determined by the absolute temperature T:

$$\langle E_K \rangle = \frac{1}{2} M \langle v^2 \rangle = \frac{3}{2} kT \tag{7.2}$$

where k is the Boltzmann constant and $\langle \rangle$ denotes the average value. Thus, in a mixture of electrons (mass m) and ions (mass M_i) at different temperatures (say, $T_e > T_i$), you would typically find the electrons with more kinetic energy than the ions. (Note that even if $T_e = T_i$, the electrons would typically be moving much faster than the ions, as:

$$\frac{1}{2} m \langle v_e^2 \rangle = \frac{1}{2} M_i \langle v_i^2 \rangle \tag{7.3}$$

$$v_e\big|_{rms} = \sqrt{\langle v_e^2 \rangle} = \sqrt{\frac{M_i}{m}} \; v_i\big|_{rms} \tag{7.4}$$

that is the root-mean-square (rms) speed of the electrons will be $\sqrt{M_i/m} \approx \sqrt{40 \cdot 1827} \approx$ 270 times the rms speed of the Argon ions, in the case of an ^{40}Ar plasma).

How can it be that the collisions between electrons and ions fail to average out the kinetic energy and hence the temperature? Consider a hard (in-line) elastic (energy conserving) collision between a slow moving (we'll assume stopped, $u_i = 0$) ion and a speeding electron (v_i).
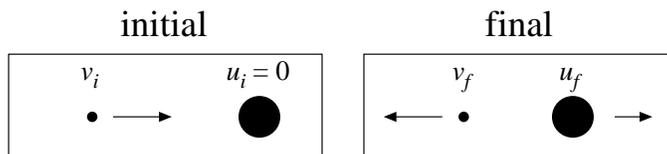
We can find the final velocities (v_f & u_f) by applying conservation of momentum and energy. The quadratic equation that is energy conservation is equivalent to the linear equation of reversal of relative velocity:

$$v_i = u_f - v_f \tag{7.5}$$

$$mv_i = mv_f + Mu_f \tag{7.6}$$

with solution:

$$u_f = \frac{2m}{m + M} \; v_i \tag{7.7}$$

You can think of the ion velocity as being built up from a sequence of these random blows. Usually these collisions would be glancing, but as a maximum case, we can think of each

blow as adding a velocity of $\Delta u = 2mv_i/(m + M)$ in a random direction. Consider the ion velocity vector before and after one of these successive blows:

$$\mathbf{u}_f = \mathbf{u}_i + \Delta\mathbf{u} \tag{7.8}$$
$$u_f^2 = u_i^2 + (\Delta u)^2 + 2\,\mathbf{u}_i \cdot \Delta\mathbf{u} \tag{7.9}$$

Now on average the dot product term will be zero (i.e., no special direction for $\Delta\mathbf{u}$), so on average the speed-squared increases by $(\Delta u)^2$ at each collision. Starting from rest, after N collisions we have:

$$u_f^2 = N\,(\Delta u)^2 \tag{7.10}$$
$$\frac{1}{2}\,M\,u_f^2 = N\,\frac{1}{2}\,M\,(\Delta u)^2 \tag{7.11}$$
$$= N\,\frac{1}{2}\,M\,\left[\frac{2m}{m + M}\,v_i\right]^2 \tag{7.12}$$
$$= N\,\frac{4m}{m + M}\,\frac{M}{m + M}\,\frac{1}{2}\,mv_i^2 \tag{7.13}$$

Thus for argon, $N \approx 18,000$ hard collisions are required for the ion kinetic energy to build up to the electron kinetic energy. Note that in nearly equal mass collisions (e.g., between an argon ion and an argon atom), nearly 100% of the kinetic energy may be transferred in one collision. Thus ions and neutral atoms are in close thermal contact; and electrons are in close contact with each other. But there is only a slow energy transfer between electrons and ions/atoms. In photoionization, electrons receive most of the photon's extra energy as kinetic energy. Slow energy transfer from the fast moving electrons heats the ions/atoms. When the Sun goes down, the electrons cool to nearly the ion temperature.

Note that the hottest thing near you now is the glow-discharge plasma inside a fluorescent bulb: $T_e > 3 \times 10^4$ K... hotter than the surface of the Sun, much hotter than the tungsten filament of an incandescent light bulb. The cool surface of the bulb gives testimony to the low plasma density ($n_e \sim 10^{16}$–10^{17} m^{-3}) inside the bulb. Containing this hot but rarefied electron gas heats the tube hardly at all — when the plasma's heat gets distributed over hugely more particles in the glass envelope, you have a hugely reduced average energy, and hence temperature.

Plasma People

Irving Langmuir (1881–1957)

Born in Brooklyn, New York, Langmuir earned a B.S. (1903) in metallurgical engineering from Columbia University. As was common in those days, he went to Europe for advanced training and completed a Ph.D. (1906) in physical chemistry under Nobel laureate Walther Nernst at University of Göttingen in Germany. Langmuir returned to this country, taking the job of a college chemistry teacher at Stevens Institute in Hoboken, New Jersey. Dissatisfied with teaching, he tried industrial research at the recently opened General Electric Research Laboratory[3] in Schenectady, New York. Langmuir's work for G.E. involved the

---

[3]G.E. calls this lab, which opened in 1900, the "first U.S. industrial laboratory devoted to research, innovation and technology", but Edison's Menlo Park "invention factory" (1876) would often claim that honor. Bell Labs (founded 1925), with six Nobel prizes in physics, would probably claim to be the world's preeminent industrial research lab, but the break up of the "Ma Bell" monopoly has also reduced Bell Labs.

then fledgling[4] electric power industry. He begin with improving the performance of incandescent electric light bulb. (Langmuir is in the inventors hall of fame for patent number 1,180,159: the modern gas-filled tungsten-filament incandescent electric light.) His work with hot filaments naturally led to thermionic emission and improvements in the vacuum triode tube that had been invented by Lee de Forest[5] in 1906. Working with glow discharge tubes (think of a neon sign), he invented diagnostic tools like the Langmuir probe to investigate the resulting "plasma" (a word he coined). "Langmuir waves" were discovered in the plasma. Along the way he invented the mercury diffusion pump. In high vacuum, thin films can be adsorbed and studied. As he said in his 1932 Nobel prize lecture:

> When I first began to work in 1909 in an industrial research laboratory, I found that the high-vacuum technique which had been developed in incandescent lamp factories, especially after the introduction of the tungsten filament lamp, was far more advanced than that which had been used in university laboratories. This new technique seemed to open up wonderful opportunities for the study of chemical reactions on surfaces and of the physical properties of surfaces under well-defined conditions.

In 1946, Langmuir developed the idea of cloud seeding, which brought him into contact with meteorologist Bernard Vonnegut, brother of my favorite author Kurt Vonnegut. That's how Langmuir became the model for Dr. Felix Hoenikker, creator of "ice-nine" in the novel *Cat's Cradle*. In fact Langmuir created the ice-nine idea (a super-stable form of solid water, with resulting high melting point, never seen in nature for want of a seed crystal) for H.G. Wells who was visiting the G.E. lab.

Lyman Spitzer, Jr (1914–1997)

Lyman Spitzer was born in Toledo, Ohio, and completed his B.A. in physics from Yale in 1935. For a year he studied with Sir Arthur Eddington at Cambridge, but that did not work out so he returned to this country and entered Princeton. He completed his Ph.D. in 1938 under Henry Norris Russell, the dean of American astrophysics. Following war work on sonar, he returned to astrophysics. His 1946 proposal for a large space telescope earned him the title "Father of the Hubble Space Telescope".

Because of the bend in *The Curve of Blinding Energy*[6], the lowest energy nuclei are of middle mass (e.g., ^{56}Fe). Thus nuclear energy can be released by breaking apart big nuclei (like ^{235}U and ^{239}Pu): fission as in the mis-named atomic bomb or by combining small nuclei (like ^{2}H deuterium and ^{3}H tritium): fusion as in the hydrogen bomb. In 1943 Edward Teller and Stanislaw Ulam started theoretical work on bombs based on thermonuclear fusion then called "super" bombs. The end of WWII slowed all bomb work, but the explosion of the first Russian atomic bomb, "Joe 1", in 1949, rekindled U.S. bomb work. This history of renewed interest in H-bomb work is mixed up with Russian espionage—real and imagined,

---

[4]Although not germane to my topic, I can't resist mentioning the famous AC (with Tesla and Westinghouse) vs. DC (with Edison, J.P. Morgan, and G.E.) power wars just before the turn of the century. The battle had a variety of bizarre twists, e.g., each side inventing an electric chair based on the opposite power source aiming to prove that the opponent's source was more dangerous than theirs. Easy voltage transformation with AC guaranteed the victory we see today in every electrical outlet worldwide. Unfortunately the AC frequency did not get standardized so its 60 Hz here and 50 Hz in Europe.

[5]1873–1961; "Father of Radio", born Council Bluffs, Iowa, B.S & Ph.D. in engineering from Yale

[6]Title of an interesting book by John McPhee; ISBN: 0374515980; UF767 .M215 1974

"McCarthyism", and the removal of Robert Oppenheimer's[7] security clearance in 1954.[8]

Our piece of the fusion story starts with Spitzer's 1951 visit with John Wheeler[9] in Los Alamos at just the right moment. The building of the Super, in response to Joe 1, had been failing: difficult calculations showed model designs would not ignite. Energy lost by thermal radiation and expansion cooled the "bomb" faster than nuclear energy was released.[10] But just before Spitzer arrived, Ulam and Teller had come up with a new method (radiation coupling for fuel compression) that Oppenheimer called "technically sweet". Meanwhile, in a story Hollywood would love, Argentine president Juan Perón announced that his protégè, Ronald Richter an Austrian-German chemist, working in a secret island laboratory had achieved controlled fusion. The story (of course) fell apart in less than a year, but it got both Spitzer and funding agencies interested. Spitzer's idea (the "Stellarator") was a magnetically confined plasma that would slowly fuse hydrogen to helium, releasing nuclear energy to heat steam and turn electrical generators. Spitzer and Wheeler hatched a plan to return to Princeton with a bit of both projects (of course funded by the government[11]): Project Matterhorn B would work on bombs and Project Matterhorn S would work on stellarators. Matterhorn B made calculations for the thermonuclear stage of the test shot Mike (1 Nov 1952)—the first H-bomb. The device worked even better than they had calculated.

From 1951 until 1958 stellarator research was classified. Optimistic projections[12] for fusion reactors were believed by all—after all physicists had completed key projects (atomic bomb, radar, H-bomb) like clockwork. Why should controlled fusion be much different from the carefully calculated fusion of an H-bomb? Early hints of fusion success (neutron emission) turned out to be signs of failure: "instabilities" or disturbances that grew uncontrollably in the plasma. Turbulence—the intractable problem in hydrodynamics[13] from the 19th century—came back to bite physics in the 1950s. Instead of the hoped-for "quiescent" plasma, experiment found large amplitude waves: a thrashing plasma that easily escaped the magnetic field configurations physicists had thought should confine it. In 1961 Spitzer turned directorship of the Princeton Plasma Physics Laboratory (PPPL) over to Melvin Gottlieb, and largely returned to astrophysical plasmas.

---

[7]J. Robert Oppenheimer (1904–1967): born New York, NY, Ph.D. (1927) Göttingen. Directed atomic bomb work at Los Alamos during WWII; 'father of the atomic bomb'.

[8]See: *Dark Sun*, by Richard Rhodes, UG1282.A8 R46 1995

[9]John Archibald Wheeler (1911–2008): born Jacksonville, FL, B.S.+Ph.D. (1933) Johns Hopkins, Feynman's major professor at Princeton in 1942. Famous book: *Gravitation*. Coined the word "black hole".

[10]*Curve of Binding Energy* p. 64: One day, at a meeting of people who were working on the problem of the fusion bomb, George Gamow placed a ball of cotton next to a piece of wood. He soaked the cotton with lighter fuel. He struck a match and ignited the cotton. It flashed and burned, a little fireball. The flame failed completely to ignite the wood which looked just as it had before—unscorched, unaffected. Gamow passed it around. It was petrified wood. He said, "That is where we are just now in the development of the hydrogen bomb."

[11]The U.S. Atomic Energy Commission (AEC) initiated the program for magnetic fusion research under the name Project Sherwood. In 1974 the AEC was disbanded and replaced by the Energy Research and Development Administration (ERDA). In 1977, ERDA in turn was disbanded and its responsibilities transferred to the new Department of Energy (DOE). Since 1977, DOE has managed the magnetic fusion research program.

[12]An August 1954 report on a theoretical Model D Stellarator (only Model B, with a 2" tube, had actually been built), using assumptions that proved false, projected a power output approximately four times that of Hoover Dam. The usual joke is that controlled fusion will always be just ten years away.

[13]Turbulence in hydrodynamics is one of the Clay Millennium Prize Problems (essentially Nobel + Hilbert for mathematics in this century): 1 million dollars for a solution!

The current focus for magnetically confined plasma research is the "tokamak": a particular donut-shape (torus) configuration that confines the plasma in part using a large current flowing through the plasma itself. Designed by Russians Igor Tamm and Andrei Sakharov, the T-3 Tokamak surprised the plasma physics world when results were made public in 1968. In 1969, PPPL quickly converted the C-Stellarator into the ST tokamak.

Returning to astrophysics, Spitzer's influential books demonstrate his connection to plasma physics: *Physics of Fully Ionized Gases* (1956 & 1962), *Diffuse Matter in Space* (1968), *Physical Processes in the Interstellar Medium* (1978), and *Dynamical Evolution of Globular Clusters* (1988).

Summary

Almost everywhere in time and space, plasmas predominate. While present in some natural phenomena at the surface of the Earth (e.g., lightning), plasmas were "discovered" in glow discharges. In the first half of the 1900s, plasma physics was honored with two Nobels[14] Langmuir worked to understand "industrial" plasmas in things that are now considered mundane like fluorescent lights. Appleton located the ionosphere: an astronomical plasma surrounding the Earth. Both Nobels were connected to larger historical events (the rise of radio and radar in time to stop Hitler at the English channel).

In the second half of the 1900s, plasma physics was connected to unpleasant problems: politics (McCarthyism), espionage, and turbulence. While H-bombs worked as calculated, controlled fusion proved difficult and only slow progress has been achieved. Astrophysical plasmas (for example, around the Sun) have also proved difficult to understand.

In this century, "industrial" plasmas are again newsworthy with plasma etching for computer chip manufacture and plasma display screens for HDTV.

Since the calculation of plasma properties has proved so difficult, measurements of plasma properties ("plasma diagnostics") are critical. In all sorts of plasmas (astrophysical, thermonuclear, industrial), the primary plasma diagnostic has been that first developed by Langmuir. The purpose of this lab is to measure basic plasma properties (T_e, n_e) using a Langmuir probe.

Glow Discharge Tube

In a glow (or gas) discharge tube, a large voltage ($\sim$100 V) accelerates free electrons to speeds sufficient to cause ionization on collision with neutral atoms. The gas in the tube is usually at low pressure ($\sim$1 torr), so collisions are not so frequent that the electrons fail to reach the speed required for ionization.

Making an electrical connection to the plasma is a more complicated process than it might seem:

   A. The creation of ions requires energetic collisions (say, energy transfer $\sim$10 eV). Kinetic

---

[14]An additional Nobel for plasma physics: Hannes Alfvén (1970). Note that Tamm and Sakharov (named in the context of the Tokamak) also received Nobels, but not for plasma physics work.
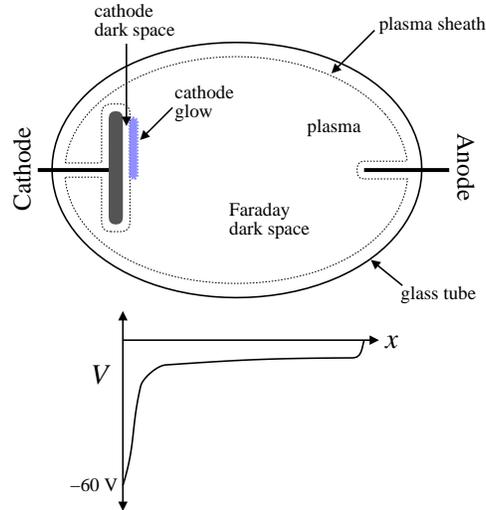
Figure 7.1: When a current flows between the anode (+) and cathode (−), the gas in the tube partially ionizes, providing electrons and ions to carry the current. The resulting plasma is at a nearly constant potential. Electric fields (from potential differences) exist mostly at the edge of the plasma, in the plasma sheath. The largest potential drop is near the cathode. The resulting cathode glow is the region of plasma creation. Increased discharge current I_c results in expanded coverage of the cathode by the cathode glow, but not much change in the cathode voltage V_c. Note that if the anode/cathode separation were larger, a *positive column* of excited gas would be created between the Faraday dark space and the anode.

energy for the collision must in turn come from potential differences of $\gtrsim$ 10 V. However, we've said that conductors (like the plasma) are at approximately constant potential. Thus ion creation must occur at the edge of the plasma.

B. It turns out that attempts to impose a potential difference on a plasma fail. Typically potential differences propagate only a short distance, called the Debye length λ_D, into the plasma:

$$\lambda_D = \sqrt{\frac{\epsilon_0 k T_e}{e^2 n_e}} \tag{7.14}$$

Thus we expect the "edge" of the plasma to be just a few λ_D thick.

C. The small electron mass (compared to ions), guarantees high electron mobility in the plasma. Thus we expect electrons to carry the bulk of the current in the plasma. But this cannot be true in the immediate vicinity of the cathode. The electrons inside the cold cathode (unlike the heated cathode in thermionic emission) are strongly held— they will not spontaneously leave the cathode. Thus near the cathode the current must be carried by ions which are attracted to the negatively charged cathode. Once in contact with the cathode an ion can pick up an electron and float away as a neutral atom. Note particularly that there is no such problem with conduction at the anode: the plasma electrons are attracted to the anode and may directly enter the metal to continue the current. Thus we expect the active part of the discharge to be directly adjacent to the cathode.

D. If you stick a wire into a plasma, the surface of the wire will be bombarded with

electrons, ions, and neutrals. Absent any electric forces, the impact rate per m^2 is given by

$$J = \frac{1}{4} \, n \, \langle v \rangle = \frac{1}{4} \, n \, \sqrt{\frac{8kT}{\pi M}} \tag{7.15}$$

where n is the number density of the particles and $\langle v \rangle$ is their average speed. If the particles follow the Maxwell-Boltzmann speed distribution, the average speed is determined by the temperature T and mass M of the particles. (Recall: $v_{rms} = \sqrt{3\pi/8} \, \langle v \rangle \approx 1.085 \, \langle v \rangle$ for a Maxwell-Boltzmann distribution.) Since the electron mass is much less than the ion mass ($m \ll M_i$) and in this experiment the temperatures are also different ($T_e \gg T_i$), the average electron speed is much greater than the average ion speed. Thus an item placed in a plasma will collect many more electrons than ions, and hence will end up with a negative charge. The over-collection of electrons will stop only when the growing negative charge (repulsive to electrons, attractive to ions) reduces the electron current and increases the ion current so that a balance is reached and zero net current flows to the wire. The resulting potential is called the *floating potential, V_f*.

The upshot of these considerations is that objects immersed in a plasma do not actually contact the plasma. Instead the plasma produces a "sheath", a few Debye lengths thick, that prevents direct contact with the plasma. We begin by demonstrating the above equations.

The starting point for both equations is the Boltzmann factor:

$$\text{probability} = \mathcal{N}e^{-E/kT} \tag{7.16}$$

which reports the probability of finding a state of energy E in a system with temperature T, where $\mathcal{N}$ is a normalizing factor that is determined by the requirement that the total probability adds up to 1. The energy of an electron, the sum of kinetic and potential energy, is

$$E = \frac{1}{2} \, mv^2 - eV \tag{7.17}$$

where V is the voltage at the electron's position. (See that for an electron, a low energy region corresponds to a high voltage region. Thus Boltzmann's equation reports that electrons are most likely found in *high* voltage regions.) To find $\mathcal{N}$ add up the probability for all possible velocities and positions:

$$
\begin{aligned}
1 &= \mathcal{N} \int_{-\infty}^{+\infty} dv_x \int_{-\infty}^{+\infty} dv_y \int_{-\infty}^{+\infty} dv_z \quad \int d\mathcal{V} \, \exp\left(\frac{-\frac{1}{2} \, mv^2 + eV(\mathbf{r})}{kT} \right) \\
&= \mathcal{N} \int_{-\infty}^{+\infty} e^{-mv_x^2/2kT} dv_x \int_{-\infty}^{+\infty} e^{-mv_y^2/2kT} dv_y \int_{-\infty}^{+\infty} e^{-mv_z^2/2kT} dv_z \quad \int d\mathcal{V} \, e^{eV(\mathbf{r})/kT} \\
&= \mathcal{N} \left[\frac{2\pi kT}{m} \right]^{3/2} \mathcal{V} \, e^{eV(\mathbf{r_0})/kT}
\end{aligned}
\tag{7.18}
$$

where we have used the Gaussian integral:

$$\int_{-\infty}^{+\infty} e^{-\alpha x^2} dx = \sqrt{\frac{\pi}{\alpha}} \tag{7.19}$$
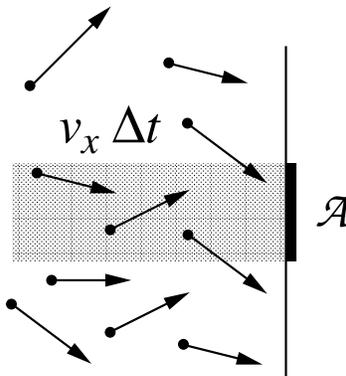
Figure 7.2: Seeking the number of electrons to hit the area $\mathcal{A}$, focus on just those electrons with some particular x-velocity, v_x. Electrons within $v_x \Delta t$ of the wall will hit it sometime during upcoming time interval Δt. The number of such electrons that will hit an area $\mathcal{A}$ will be equal to the number of such electrons in the shaded volume. Note that many electrons in that volume will *not* hit $\mathcal{A}$ because of large perpendicular velocities, but there will be matching electrons in neighboring volumes which will hit $\mathcal{A}$. To find the total number of hits, integrate over all possible v_x.

and the mean value theorem to replace the integral over the volume $\mathcal{V}$ of the electron gas with $\mathcal{V}$ times some value of the integrand in that domain. Thus:

$$\text{probability} = \frac{e^{-E/kT}}{\left[\frac{2\pi kT}{m}\right]^{3/2} \mathcal{V} e^{eV(\mathbf{r_0})/kT}} = \frac{1}{\mathcal{V}} \left[\frac{m}{2\pi kT}\right]^{3/2} \exp\left(-\frac{\frac{1}{2}mv^2 - e(V - V_0)}{kT}\right) \quad (7.20)$$

So, if we have N (non-interacting) electrons in $\mathcal{V}$, the expected distribution of electrons (w.r.t. position and velocity) can be expressed as:

$$f = n_0 \left[\frac{m}{2\pi kT}\right]^{3/2} \exp\left(-\frac{\frac{1}{2}mv^2 - e(V - V_0)}{kT}\right) \quad (7.21)$$

where $n_0 = N/\mathcal{V}$ is the bulk electron density which is also the electron density at potential V_0. If we don't care about the distribution of electron speed, we can add up (integrate) over all possible speeds (which just reproduces the Gaussian factors from $\mathcal{N}$), resulting in the electron number density, n:

$$n = n_0 \exp\left(\frac{e(V - V_0)}{kT}\right) \quad (7.22)$$

D. Collision Rate: If we just care about the distribution of velocity in one direction (say, v_x), integrals over the other two directions (v_y, v_z) results in:

$$f_x = n_0 \left[\frac{m}{2\pi kT}\right]^{1/2} \exp\left(-\frac{\frac{1}{2}mv_x^2 - e(V - V_0)}{kT}\right) \quad (7.23)$$

We simplify further by considering the case where the potential is a constant (i.e., $V = V_0$). In order to calculate the number of electrons that hit an area $\mathcal{A}$ during the coming interval Δt, focus on a subset of electrons: those that happen to have a particular velocity v_x. (We

will then add up all possible v_x, from v_{min} to ∞. For this immediate problem $v_{min} = 0$, i.e., any electron moving to the right can hit the wall; however, we'll need the more general case in a few pages.) An electron must be sufficiently close to the wall to hit it: within $v_x \Delta t$. The number of electron hits will be equal to the number of such electrons in the shaded volume in Figure 7.2. To find the total number of hits, integrate over all possible v_x:

$$\text{number of hits} \quad = \quad \int_{v_{min}}^{\infty} dv_x \, f_x \, \mathcal{A} v_x \Delta t \tag{7.24}$$

$$= \quad \mathcal{A} \Delta t \, n_0 \left[\frac{m}{2\pi kT} \right]^{1/2} \int_{v_{min}}^{\infty} \exp\left(-\frac{\frac{1}{2}mv_x^2}{kT} \right) v_x \, dv_x \tag{7.25}$$

$$= \quad \mathcal{A} \Delta t \, n_0 \left[\frac{m}{2\pi kT} \right]^{1/2} \frac{kT}{m} \int_{mv_{min}^2/2kT}^{\infty} e^{-y} dy \tag{7.26}$$

$$= \quad \mathcal{A} \Delta t \, n_0 \left[\frac{kT}{2\pi m} \right]^{1/2} \exp\left(-\frac{\frac{1}{2}mv_{min}^2}{kT} \right) \tag{7.27}$$

$$\longrightarrow \quad \mathcal{A} \Delta t \, n_0 \left[\frac{kT}{2\pi m} \right]^{1/2} \quad \text{for} \quad v_{min} \to 0 \tag{7.28}$$

Thus the particle current density, J_e, (units: hits per second per m^2) is

$$J_e = \frac{\text{hits}}{\mathcal{A} \Delta t} = \frac{1}{4} \, n_e \left[\frac{8kT}{\pi m} \right]^{1/2} = \frac{1}{4} \, n_e \, \langle v \rangle \tag{7.29}$$

B. Debye Length (λ_D): To introduce the Debye length, we consider a very simplified case: the potential variation due to varying electron density in a uniform (and nearly canceling, i.e., $n_i = n_{e0}$) ion density. By Poisson's equation:

$$\epsilon_0 \, \nabla^2 V = -e \left(n_i - n_{e0} e^{eV/kT} \right) = -e n_{e0} \left(1 - e^{eV/kT} \right) \approx e n_{e0} \left(\frac{eV}{kT} \right) \tag{7.30}$$

where we have assumed $eV/kT \ll 1$ so we can Taylor expand. If we consider variation in just one direction (x), we have:

$$\frac{d^2 V}{dx^2} = \left(\frac{e^2 n_{e0}}{\epsilon_0 kT} \right) V = \frac{1}{\lambda_D^2} V \tag{7.31}$$

with exponentially growing/decaying solutions of the form: $V \propto \exp\left(\pm x/\lambda_D \right)$. We learn from this that deviations from charge neutrality take place on a length scale of λ_D, and are self-reinforcing so that very large changes in the potential can be accomplished in a distance of just a few λ_D.

In the cathode glow, high speed electrons (accelerated in the large electric field in the neighboring cathode dark space) collision-ionize the gas. Many of the newly created ions are attracted to the cathode. Accelerated by cathode dark space electric field, the ions crash into the cathode. The collision results in so-called secondary electron emission[15] from the cathode. These secondary electrons are then repelled from the cathode and cause further ionization in the cathode glow region. The observed voltage drop near the cathode (the "cathode fall"), is exactly that required so that, on average, a secondary electron is able to

---

[15]Secondary electron emission is the key to photomultiplier tubes used to detect single photons.
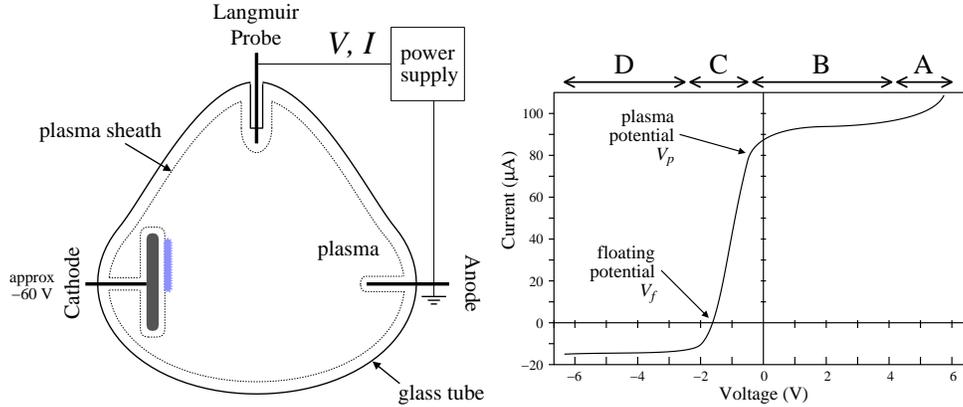
Figure 7.3: At its simplest, a Langmuir probe is just a wire stuck into the plasma. Of course, electrical contact with the plasma is limited by the plasma sheath. An external power supply allows the probe's voltage to be adjusted and the resulting current measured. The characteristics of the plasma can be determined by careful analysis of the resulting IV relationship.

reproduce itself through the above mechanism. Clearly the cathode fall depends both on the gas (how easy it is to ionize) and the cathode material (how easy is it to eject secondary electrons). Luckily we will not need to understand in detail the processes maintaining the discharge near the cathode. A properly operating Langmuir probe does not utilize secondary electron emission, that is we limit voltage drop near the probe to much less than the cathode fall.

Langmuir Probe Theory

As stated above, if a wire is stuck into a plasma, instead of connecting to the plasma potential (V_p) it instead charges to a negative potential (the floating potential: $V_f < V_p$) so as to retard the electron current enough so that it matches the ion current. An equal flow of electrons and ions—zero net current— is achieved and charging stops. If the wire is held at potentials above or below V_f, a net current will flow into the plasma (positive I: plasma electrons attracted to the probe), or out from the plasma (negative I: plasma ions attracted to the probe). Figure 7.3 displays a Langmuir probe IV curve dividing it up into four regions (A–D):

A. When the Langmuir probe is well above the plasma potential it begins to collect some of the discharge current, essentially replacing the anode.

B. When the probe is at the plasma potential (left side of region B), there is no plasma sheath, and the surface of the probe collects ions and electrons that hit it. The electron current is much larger than the ion current, so at V_p the current is approximately:

$$I_p = e\mathcal{A}\frac{1}{4}n_e\left[\frac{8kT_e}{\pi m}\right]^{1/2} \tag{7.32}$$

where $\mathcal{A}$ is the area of the probe. For $V > V_p$, the sheath forms, effectively expanding

slightly the collecting area. Thus the probe current increases slightly and then levels off in this region.

C. When the Langmuir probe's potential is below V_p, it begins to repel electrons. Only electrons with sufficient kinetic energy can hit the probe. The minimum approach velocity v_{min} allowing a probe hit can be determined using conservation of energy:

$$\frac{1}{2} mv_{min}^2 - eV_p = -eV \tag{7.33}$$

$$\frac{1}{2} mv_{min}^2 = e(V_p - V) \tag{7.34}$$

where the probe is at potential V. Equation 7.27 then reports that the resulting electron current is

$$I = e\mathcal{A}\, n_e \left[\frac{kT_e}{2\pi m}\right]^{1/2} \exp\left(-\frac{e(V_p - V)}{kT_e}\right) \tag{7.35}$$

For $V \ll V_p$, very few of the electrons have the required velocity. At $V = V_f$ the electron current has been suppressed so much that it equals the ion current. (The ion current is always present in region C, but typically for $V > V_f$ it is "negligible" in comparison to the electron current.)

D. In this region the probe is surrounded by a well developed sheath repelling all electrons. Ions that random-walk past the sheath boundary will be collected by the probe. As the sheath area is little affected by the probe voltage, the collected ion current is approximately constant. (At very negative voltages, $V \sim -60$ V, secondary electron emission following ion hits leads to large currents, and a glow discharge.) The equation for this ion current is a bit surprising (that is not analogous to the seeming equivalent electron current in region B):

$$I_i \approx -\frac{1}{2}e\mathcal{A}\, n_i\, u_B \tag{7.36}$$

where u_B is the Bohm[16] velocity—a surprising combination of electron temperature and ion mass:

$$u_B = \sqrt{kT_e/M_i} \tag{7.37}$$

Note that Equation 7.15 would have suggested a similar result but with the thermal velocity rather than the Bohm velocity.

Region D: Ion Current

As shown in Figure 7.4, consider bulk plasma ($n_i = n_e = n_0$, located at $x = 0$ with a plasma potential we take to be zero $V_p = 0$) near a planar Langmuir probe (located at $x = b$ biased with a negative voltage). In the bulk plasma, the ions are approaching the probe with a

---

[16]David Bohm (1917–1992) Born: Wiles-Barre, PA, B.S. (1939) PSU, joined Communist Party (1942), Ph.D. (1943) Berkeley — Oppenheimer's last student before he became director at Los Alamos. Cited for contempt by McCarthy's House Un-American Activities Committee, Bohm was arrested in 1950. Although acquitted at trial, he was nevertheless blacklisted, and ended up in London.
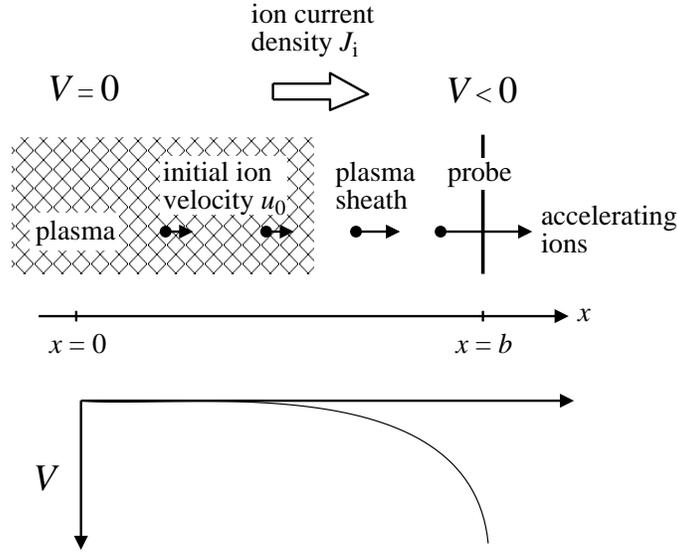
Figure 7.4: When the Langmuir probe is negatively biased relative to the plasma, it attracts ions from the plasma (and repels electrons). Of course, the the accelerating electric field (from voltage difference) is largely confined to the edge of the plasma, the plasma sheath. The resulting ion current density J_i should be continuous: the same for every x.

velocity u_0. The ions are accelerated towards the negatively biased probe; we can determine their velocity, $u(x)$ at any position by applying conservation of energy:

$$\frac{1}{2}M_i u^2 + eV(x) = \frac{1}{2}M_i u_0^2 \tag{7.38}$$

$$u(x) = u_0 \sqrt{1 - \frac{2eV(x)}{M_i u_0^2}} \tag{7.39}$$

The moving ions constitute a steady electric current density:

$$J_i = en(x)u(x) = en_0 u_0 \tag{7.40}$$

(i.e., J_i doesn't depend on position), so $n(x)$ must decrease as the ions speed toward the probe.

$$n(x) = n_0 u_0 / u(x) = \frac{n_0}{\sqrt{1 - \frac{2eV(x)}{M_i u_0^2}}} \tag{7.41}$$

The varying charge density affects the electric potential which in turn affects the electron density through the Boltzmann equation (Equation 7.22). Poisson's equation reads:

$$\frac{d^2V}{dx^2} = -\frac{\rho(x)}{\epsilon_0} = \frac{e}{\epsilon_0} n_0 \left[\exp\left(\frac{eV(x)}{kT_e}\right) - \frac{1}{\sqrt{1 - \frac{2eV(x)}{M_i u_0^2}}} \right] \tag{7.42}$$

The first step in solving most any differential equation, is to convert it to dimensionless form. We adopt dimensionless versions of voltage, position, and velocity:

$$\tilde{V} = \frac{eV}{kT_e} \tag{7.43}$$

$$\tilde{x} = \frac{x}{\lambda_D} = x\left[\frac{n_0 e^2}{\epsilon_0 kT_e}\right]^{1/2} \tag{7.44}$$

$$\tilde{u}_0 = \frac{u_0}{u_B} = \frac{u_0}{\sqrt{kT_e/M_i}} \tag{7.45}$$

If we multiply Equation 7.42, by e/kT_e, we find:

$$\frac{d^2\tilde{V}}{dx^2} = \frac{1}{\lambda_D^2}\left[\exp(\tilde{V}) - \frac{1}{\sqrt{1 - \frac{2\tilde{V}}{\tilde{u}_0^2}}}\right] \tag{7.46}$$

or

$$\frac{d^2\tilde{V}}{d\tilde{x}^2} = \exp(\tilde{V}) - \frac{1}{\sqrt{1 - \frac{2\tilde{V}}{\tilde{u}_0^2}}} \tag{7.47}$$

First let's simplify notation, by dropping all those tildes.

$$\frac{d^2V}{dx^2} = \exp(V) - \frac{1}{\sqrt{1 - \frac{2V}{u_0^2}}} \tag{7.48}$$

The physical solution should show the neutral plasma picking up a positive charge density as we approach the probe (as electrons are repelled and ions attracted to the probe). The r.h.s. of Equation 7.48 is proportional to the charge density $n_e - n_i$ which is nearly zero at $x = 0$ (where $V = 0$) and monotonically declines as x approaches the probe (where $V < 0$). If we Taylor expand the r.h.s., we find:

$$\begin{aligned}
n_e - n_i &\propto \exp(V) - \frac{1}{\sqrt{1 - \frac{2V}{u_0^2}}} \\
&\approx \left(1 + V + \frac{1}{2}V^2 + \cdots\right) - \left(1 + \frac{1}{u_0^2}V + \frac{3}{2u_0^4}V^2 + \cdots\right) \\
&= \left(1 - \frac{1}{u_0^2}\right)V + \frac{1}{2}\left(1 - \frac{3}{u_0^4}\right)V^2 + \cdots \tag{7.49}
\end{aligned}$$

Thus $u_0 \geq 1$ is required for $n_i > n_e$ when $V < 0$. Maximizing the extent of charge neutrality requires $u_0 = 1$. As shown in Figure 7.5, the choice $u_0 = 1$ works throughout the $V < 0$ range (i.e., beyond the range of convergence of the Taylor's series).

Note that we can numerically solve this differential equation using *Mathematica*:

```
NDSolve[{v''[x]==Exp[v[x]]-1/Sqrt[1-2 v[x]/u0^2],v[0]==0,v'[0]==0},v,{x,0,20}]
```

but, it's not quite that simple. The unique solution with these boundary conditions, is $V = 0$ everywhere. There must be some small electric field ($V'(0) < 0 \iff E > 0$) at $x = 0$ due to probe. The smaller the choice of E, the more distant the probe. See Figure 7.6.
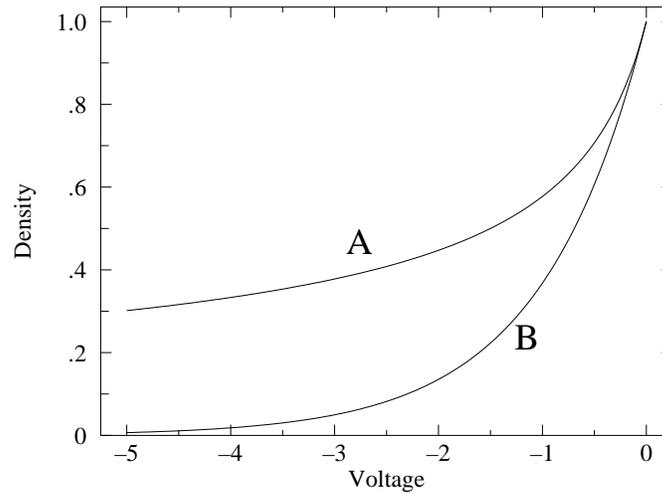
Figure 7.5: A plot of the ion density (A) $n_i \propto \left[1 - 2V(x)/u_0^2\right]^{-1/2}$ and the electron density (B) $n_e \propto e^{V(x)}$ for $V \leq 0$ in the vicinity of the probe. (With this axis choice the bulk plasma, at $V = 0$, is to the right and the probe, at $V < 0$, is on the left.) The case $u_0 = 1$ is displayed. Notice that $n_i > n_e$ throughout the plasma sheath, as is required. If $u_0 < 1$, there would be a region near $V = 0$ where $n_e > n_i$. If $u_0 > 1$, the densities in region near $V = 0$ would not be tangent. There would be a first-order mismatch between n_e and n_i: $n_i > n_e$ for $V < 0$ and $n_i < n_e$ for $V > 0$.
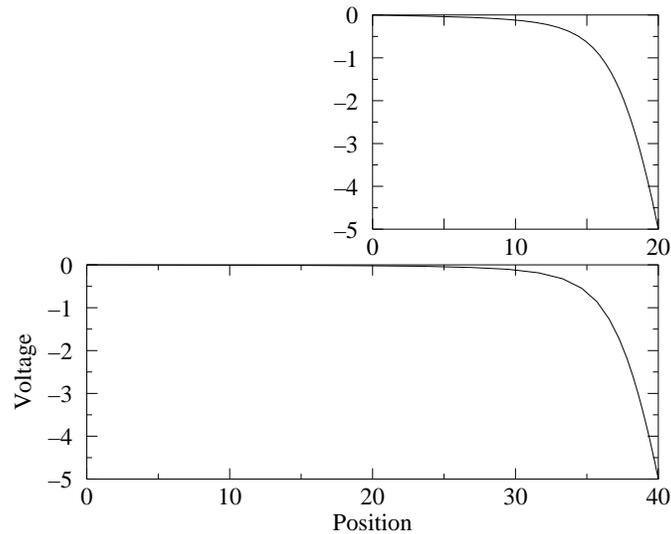
Figure 7.6: *Mathematica* solutions of our differential equation (7.48) with $V'(0) = -.0068$ (top) and $V'(0) = -.00067$ (bottom). Smaller E simply places the probe at a greater distance without much changing the voltage in the vicinity of the probe (i.e., in the sheath).

We have concluded, $u_0 = 1$, or returning to dimensioned variables: $u_0 = u_B$. However, we must now determine how the ions arrived at this velocity. It must be that in the much larger region ("pre-sheath", $x \ll 0$), under conditions of near charge neutrality ($n_e \approx n_i$), the ions traveled "downhill" a potential difference of:

$$V(-\infty) - V(0) = \frac{\text{K.E.}}{e} = \frac{\frac{1}{2}M_i u_B^2}{e} = \frac{kT_e}{2e} \tag{7.50}$$

Thus, at $x = 0$, the plasma density has been depleted compared to that at $x = -\infty$:

$$n_i = n_e = n_0 = n_\infty e^{\Delta V / kT_e} = n_\infty e^{-1/2} \tag{7.51}$$

The corresponding current density is:

$$J = e n_0 u_0 = e n_\infty e^{-1/2} u_B \approx .61 \; e n_\infty \sqrt{\frac{kT_e}{M_i}} \tag{7.52}$$

Clearly this is an approximate argument. Other arguments can give lower (e.g., .40) constants. We adopt as our final equation (if only approximate: $\pm 20\%$)

$$J_i \approx \frac{1}{2} \; e n_\infty \sqrt{\frac{kT_e}{M_i}} \tag{7.53}$$

Region C

Plasma Potential, V_p

We have said that for $V < V_p$, the electron current to the probe is exponentially growing essentially because of the exponential form of the Maxwell-Boltzmann distribution. For $V > V_p$, the electron current to the probe continues to grow, but only because of expanding collecting area due to an expanding plasma sheath. The boundary between two cases is defined by the point of maximum slope. The point where the slope is a maximum, of course, has the second derivative zero—an inflection point. Thus V_p is defined by $I''(V_p) = 0$. (This is called the Druyvesteyn criteria.)

In calculus you've learned how to apply a precise definition of derivative to find the derivatives of various functions, but how can you determine the second derivative from a set of data?

We begin with a qualitative treatment. If you have a set of equally spaced data points: $x_i = ih$ where $i \in \mathbb{Z}$ (i is an integer, h might have been called Δx), then $(f(x_{i+1}) - f(x_i))/h$ (the slope of a line through the points $(x_i, f(x_i))$ & $(x_{i+1}, f(x_{i+1}))$) ought to be something like $f'(x_i + h/2)$ (the derivative half way between x_i & x_{i+1}). Similarly, $f'(x_i - h/2) \approx (f(x_i) - f(x_{i-1}))/h$. Thus:

$$f'' \approx \frac{f'(x_i + h/2) - f'(x_i - h/2)}{h} = \frac{f(x_{i+1}) + f(x_{i-1}) - 2f(x_i)}{h^2} \tag{7.54}$$

where we imagine f'' is evaluated half way between $x_i + h/2$ & $x_i - h/2$, that is, at x_i.

More formally, theoretically f has a Taylor expansion:

$$f(x_i + y) = f(x_i) + f'(x_i)y + \frac{1}{2} \; f''(x_i)y^2 + \frac{1}{6} \; f'''(x_i)y^3 + \frac{1}{24} \; f''''(x_i)y^4 + \cdots \tag{7.55}$$
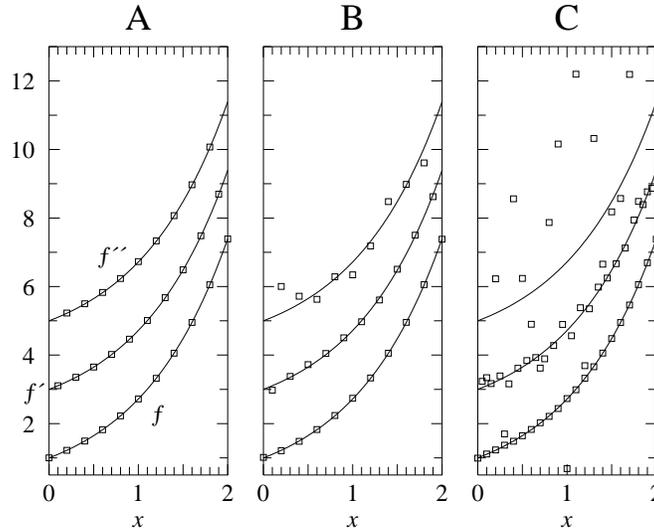
Figure 7.7: Numerical differentiation applied to the function $f(x) = e^x$. f' and f'' have been vertically offset by 2 and 4 respectively so that they do not coincide with f. (A) Here with $h = .2$, the numerically calculated f' and f'' (based on the shown data points) accurately track the actual derivatives (shown as curves). (B) With random noise of magnitude $\pm.02$ added to the y values, the data points seem to accurately follow the function, but numerically calculated f' and f'' increasingly diverge. See that numerical differentiation exacerbates noise. (C) Doubling the number of data points (with the same level of random noise as in B), makes the situation much worse. While the data points seem to follow f adequately, f' shows large deviations and f'' does not at all track the actual derivative.

So:

$$f(x_i + h) = f(x_{i+1}) = f(x_i) + f'(x_i)h + \frac{1}{2}\,f''(x_i)h^2 + \frac{1}{6}\,f'''(x_i)h^3 + \frac{1}{24}\,f''''(x_i)h^4 + \cdots$$

$$f(x_i - h) = f(x_{i-1}) = f(x_i) - f'(x_i)h + \frac{1}{2}\,f''(x_i)h^2 - \frac{1}{6}\,f'''(x_i)h^3 + \frac{1}{24}\,f''''(x_i)h^4 + \cdots$$

$$f(x_i + h) + f(x_i - h) = 2f(x_i) + f''(x_i)h^2 + \frac{1}{12}\,f''''(x_i)h^4 + \cdots \tag{7.56}$$

with the result:

$$\frac{f(x_i + h) + f(x_i - h) - 2f(x_i)}{h^2} = f''(x_i) + \frac{1}{12}\,f''''(x_i)h^2 + \cdots \tag{7.57}$$

Thus if $f''''(x_i)h^2/f''(x_i) \ll 1$ (which should follow for sufficiently small h), Equation 7.54 should provide a good approximation for $f''(x_i)$. While ever smaller h looks good mathematically, Figure 7.7 shows that too small h plus noise, is a problem.

Root Finding—Floating Potential, V_f & Plasma Potential, V_p

The floating potential is defined by $I(V_f) = 0$. Of course, it is unlikely that any collected data point exactly has $I = 0$. Instead a sequence of points with $I < 0$ will be followed

by points with $I > 0$, with V_f lying somewhere between two measured points. We seek to interpolate to find the best estimate for V_f.

We begin with the generic case, where we seek a root, i.e., the x value such that $y(x) = 0$. We start with two data points (x_1, y_1) and (x_2, y_2) with $y_1 < 0$ and $y_2 > 0$. The line connecting these two points has equation:

$$y = \frac{\Delta y}{\Delta x}\,(x - x_1) + y_1 = \frac{y_2 - y_1}{x_2 - x_1}\,(x - x_1) + y_1 \tag{7.58}$$

We seek the x value for which the corresponding y value is zero:

$$\frac{\Delta y}{\Delta x}\,(x - x_1) + y_1 \;=\; y = 0 \tag{7.59}$$

$$x - x_1 \;=\; -y_1\,\frac{\Delta x}{\Delta y} \tag{7.60}$$

$$x \;=\; x_1 - y_1\,\frac{\Delta x}{\Delta y} \tag{7.61}$$

In the case of the floating potential $(I(V_f) = 0)$, our (x_i, y_i) are a sequence of voltages with measured currents: $x_i = V_i$, $y_i = I(V_i)$, and we can apply the above formula to find the voltage, V_f where the current is zero.

The same generic result can be used to estimate the plasma potential, where $I''(V_p) = 0$. Here $x_i = V_i$, $y_i = I''(V_i)$. Just as in the floating voltage case, it is unlikely that any collected data point exactly has $I'' = 0$. Instead a sequence of points with $I'' > 0$ will be followed by points with $I'' < 0$, with V_p lying somewhere between two measured points. You will apply the generic interpolate formula to find the best estimate for V_p.

Electron Temperature, T_e

If we combine Equation 7.36 for the ion current with Equation 7.35 for the electron current, we have an estimate for the total current through out Region C:

$$I \;=\; -\frac{1}{2}\,e\mathcal{A}\,n_i\,u_B + \frac{1}{4}e\mathcal{A}\,n_e\left[\frac{8kT_e}{\pi m}\right]^{1/2}\exp\left(-\frac{e(V_p - V)}{kT_e}\right) \tag{7.62}$$

$$=\; \frac{1}{2}e\mathcal{A}\,nu_B\left\{-1 + \left[\frac{2M_i}{\pi m}\right]^{1/2}\exp\left(-\frac{e(V_p - V)}{kT_e}\right)\right\} \tag{7.63}$$

$$=\; k_1 + k_2\exp\left((V - V_p)/k_3\right) \tag{7.64}$$

Use of this equation is based on a long list of assumptions (e.g., constant collection area $\mathcal{A}$, Maxwell-Boltzmann electron speed distribution, $n_i = n_e = n\ldots$). These assumptions are not exactly true, so we do not expect this equation to be exactly satisfied: we are seeking a simplified model of reality not reality itself. By fitting this model equation to measured IV data, we can estimate the parameters k_1, k_2, k_3. Since $k_3 = kT_e/e$ we can use it to estimate the electron temperature. Similarly, k_1 represents a measure of the ion current from which (once T_e is known) n_i can be calculated using Equation 7.36.

In order to start a non-linear fit as in Equation 7.64, we need initial estimates for the parameters k_i. Measurement of V_f (where $I = 0$) and Equation 7.63, provide an estimate

for T_e:

$$0 = I \;=\; \frac{1}{2}\, e\mathcal{A}\, n\, u_B \left\{ -1 + \left[\frac{2M_i}{\pi m}\right]^{1/2} \exp\left(\frac{e(V_f - V_p)}{kT_e}\right) \right\} \tag{7.65}$$

$$1 \;=\; \left[\frac{2M_i}{\pi m}\right]^{1/2} \exp\left(\frac{e(V_f - V_p)}{kT_e}\right) \tag{7.66}$$

$$0 \;=\; \frac{1}{2}\,\ln\left[\frac{2M_i}{\pi m}\right] - \left(\frac{e(V_p - V_f)}{kT_e}\right) \tag{7.67}$$

$$\frac{e}{kT_e}\,(V_p - V_f) \;=\; \frac{1}{2}\,\ln\left[\frac{2M_i}{\pi m}\right] \tag{7.68}$$

$$\frac{kT_e}{e} \;=\; \frac{2(V_p - V_f)}{\ln\left[\frac{2M_i}{\pi m}\right]} \tag{7.69}$$

Notice that the unit for $k_3 = kT_e/e$ is volts, so $ek_3 = kT_e$ in units of J or, most simply, $k_3 = kT_e$ in units eV[17] (since $eV = e \times \text{Volt}$). It is common practice in plasma physics to report "the temperature" [meaning kT] in eV.

The ion current k_1 can be estimated from the "saturated current" in Region D, i.e., the nearly constant current a volt or so below V_f. (For future convenience, we name this "saturated current" in Region D, I_i.)

See that for $V = V_p$, $I = k_1 + k_2$. Since the ion current is "negligible" for most of region C, we can estimate k_2 from the measured current near V_p. (For future convenience, we name the current actually measured at the data point just below V_p, I_p.)

Plasma Number Density, n

Given kT_e, we have several of measuring n:

  A. measured I_i and Equation 7.36, and

  B. measured I_p and Equation 7.32,

Additionally we could use the fit values of I_i or I_p (k1 or k2, in Equation 7.64). These methods will give answers that differ by a factor of 5 or more! When different ways of measuring the same thing give different results, "systematic error" is the name of the problem. The source of this problem is our imperfect model of current flow in Region C (all those inaccurate assumptions). In particular, both (A) and (B) are hindered by the assumption of a Maxwell-Boltzmann speed distribution. (In fact measurements in Region C are commonly used to *measure*[18] that speed distribution.) Often fit parameters are preferred to individual data points (essentially because the fit averages over several data points), but that is not the case here. Thus the method considered most accurate is (A), although it

---

[17]Recall: 1 eV = 1.6022×10^{-19} J is the energy an electron gains in going through a potential difference of 1 V.

[18]For example, Langmuir probes have been used to measure the electron speed distribution in the plasma that gives rise to aurora (northern lights). The results show a non-Maxwellian speed distribution: lots of high-speed "suprathermal" electrons.

Table 7.1: R.C.A. 0A4-G Gas Triode Specifications

probe length	0.34 cm
probe diameter	0.08 cm
tube volume (approx.)	40 cm^3
peak cathode current	100 mA
DC cathode current	25 mA
starter anode drop (approx.)	60 V
anode drop (approx.)	70 V
minimum anode to cathode breakdown voltage	
(starter anode potential 0 volts)	225 V

could be improved[19] to account for the slowly varying—rather than constant—ion current in Region D. (That is, how precisely is <u>the</u> ion saturation current determined?) However, even with that ambiguity resolved, Equation 7.36 itself has expected variations of the order of 20%. Mostly physicists just live with that level of accuracy, as improved measurement methods (like microwave phase shifts due to the index of refraction of the plasma) are often not worth the effort.

Apparatus: 0A4-G Gas Triode Tube

Figure 7.8 displays the anatomy of a 0A4-G gas triode[20]. As shown in Figure 7.9, a discharge through the argon gas is controlled by a Keithley 2400 current source. Various cathode currents ($I_c = -5, -10, -20, -40$ mA) will produce various plasma densities. During tube operation, you should see the cathode glow expand as larger discharge currents are produced. Note also that the cathode voltage ($V_c \sim -60$ V) varies just a bit, over this factor of 8 increase in I_c. A glow discharge does not act like a resistor! A Keithley 2420 is used to sweep the probe voltage and simultaneously measure the probe current. Figure 7.10 shows representative results. You should note that the cathode glow is not perfectly stable. It can jump in position for no obvious reason. If a jump occurs during a probe sweep, the resulting data will look noisy (I'' randomly jumping in sign rather than smoothly going from $I'' > 0$ to $I'' < 0$) and cannot be used. Figure 7.11 shows fits to the data between V_f and V_p. The resulting reduced χ^2 were of order 10^5: the measurement errors are much smaller than the deviations between the reality and the model. While the model is "wrong", it nevertheless supplies a reasonable representation of the data. A `fudge` of the errors allows parameter error estimates to be extracted from the covariance matrix, but it's hard to give meaning to the resulting error.

---

[19]This issue is addressed by Chen in report LTP-111 `Chen206R.pdf` listed as a web reference at the end of this document. Models of the Universe can usually be improved at a cost of greater complexity. Choosing an appropriate level of complexity is something of an art. Here we are using the simplest possible model.

[20]This tube is also called a cold cathode control tube. In its usual applications, what is here called the anode is called the starter anode and what is here called the Langmuir probe is called the anode.
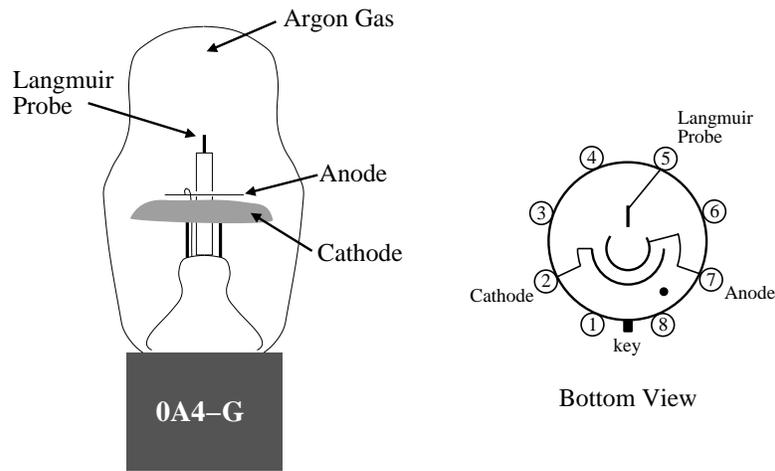
Figure 7.8: The 0A4-G triode has a large cold cathode and two "anodes" surrounded by low pressure argon gas. A glow discharge in the argon gas may be maintained by an approximate -60 V drop between the cathode (pin 2) and the "starter anode" (pin 7). The pin 5 anode may then be used as a Langmuir probe in the resulting plasma. The figure shows an R.C.A. 0A4G; A Sylvania 0A4G has the same components arranged differently.
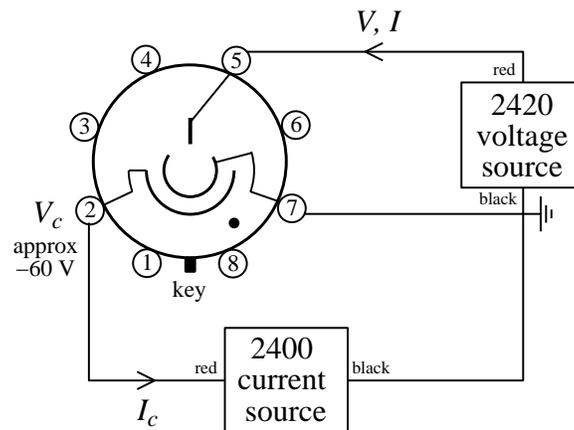
Figure 7.9: When the cathode is held at a voltage of $V_c \approx -60$ V relative to the anode, the argon gas in the tube partially ionizes and a discharge is set up between the anode and cathode. The discharge current is controlled by a Keithley 2400 in current-source mode, e.g., $I_c = -20$ mA. The voltage V on the Langmuir probe is swept by the Keithley 2420, and the current I is simultaneously measured. The resulting IV curve allows us to determine the characteristics of the plasma. Note that the pinout shows the tube as viewed from the bottom.
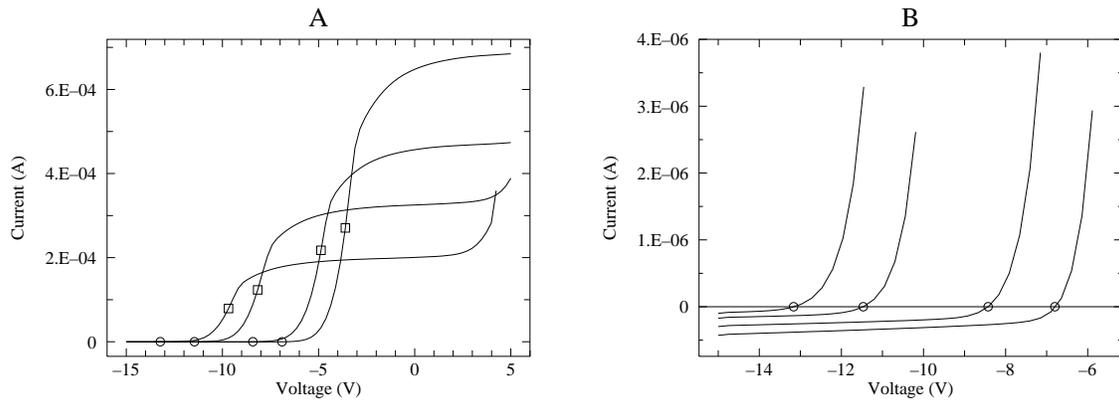
Figure 7.10: The *IV* curves for the Langmuir probe in a 0A4G tube for $I_c = -40, -20, -10, -5$ mA. The plasma potential V_p is marked with a square; The floating potential V_f is marked with a circle. Since the ion current is so much smaller than the electron current, blowing up the y scale by a factor of 200 is required to see it (see plot B). In this lab we are primarily concerned with Region C: between V_f and V_p.
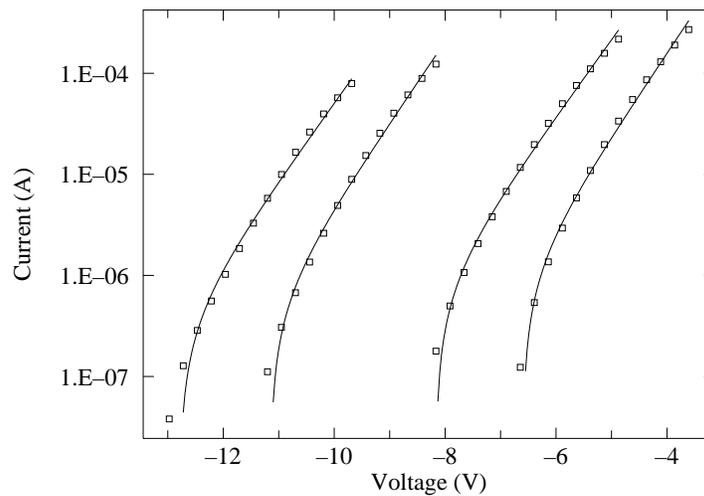
Figure 7.11: In Region C, between V_f and V_p, $I > 0$ so we can display it on a log scale. Assuming a constant ion current (k_1), constant sheath area, and a Maxwell-Boltzmann distribution of electron speed, we can fit: $I(V) = k_1 + k_2 * \exp((V - V_p)/k_3)$. The horrendous reduced χ^2 shows that these assumptions are not exact, nevertheless the fit does a reasonable job of representing the data (say, $\pm 10\%$)

Computer Data Collection

As part of this experiment you will write a computer program to control the experiment. *Plagiarism Warning*: like all lab work, this program is to be your own work! Programs strikingly similar to previous programs will alarm the grader. I understand that programming may be new (and difficult) experience for you. Please consult with me as you write the program, and test the program (with tube disconnected!) before attempting a final data-collecting run.

In the following I'm assuming the probe voltage V is stored in array `v(i)`, probe current I is stored in array `a(i)`, and the second derivative of prove current I'' is stored in array `app(i)`. Your program will control all aspects of data collection. In particular it will:

0. Declare and define all variables.

1. Open (i.e., create integer nicknames—i.e., `iunit`—for) the enets `gpib1` and `gpib2`.

2. Initialize the source-meters which must be told the maximum voltage and current to occur during the experiment. For the 2420, you can limit V, I to 25. V and .005 A; For the 2400, limit V, I to 100. V and 0.1 A;

3. Display the status of all devices before starting data collection.

4. Open files:

    (a) `VI.dump.dat` (intended for all V, I, I'' of probe, with comments (!) for cathode I_c, V_c)

    (b) `VI.fit.dat` (intended for Region C: $V, I, \delta I, I''$ of probe, with comments for cathode I_c, V_c, calculated V_f, V_p, estimated T_e, measured I_i, I_p, and the number of data points. The data points in this file are for fitting Equation 7.64:

    $$f(\mathtt{x}) = \mathtt{k1} + \mathtt{k2} * \exp((\mathtt{x} - \mathtt{k4})/\mathtt{k3}) \qquad (7.70)$$

    Note that `k4`=V_p is a constant, not an adjustable parameter.)

5. Tell the 2400 source-meter to source a cathode current, $I_c = -20$ mA

6. Let the system sleep for 60 seconds to approach thermal equilibrium.

7. Repeat the below data collection process <u>six</u> times. Since you will need just three repeats for each I_c, this will probably produce more data than is needed. However some data sets may be noisy because of unstable (moving, flickering) cathode glow. Noisy data will have multiple sign changes in calculated I''; Discard this data. If one run of this program fails to produce enough good data (three repeats), simply rename the data files (to preserve the data produced in the initial run), and re-run the program.

    Do the following for four different I_c: 5, 10, 20 40 mA (e.g., `acath=-.005*2**j` for `j=0,3`). Thus there will be 4×6 voltage sweeps.

    (a) Tell the 2400 source-meter to source a cathode current, I_c

    (b) Let the system sleep for 10 seconds to approach thermal equilibrium.

(c) Tell the 2420 to perform a linear probe voltage sweep from V_{min} to V_{max}, including N data points (say, $N = 100$). In Figure 7.10, you can see that the choice made was $V_{min} = -15.$ and $V_{max} = +5.$, but your choices will vary and must be determined by trial and error (see below). The aim is to find a range that includes from a few volts below Region C to a few volts above Region C for every cathode current I_c.

(d) Turn off the probe voltage.

(e) Repeat (a) thus obtaining up-to-date values for I_c, V_c

(f) Write a comment ('!') line to the file `VI.dump.dat` containing I_c, V_c from the 2400.

(g) Write a line to the file `VI.dump.dat` containing the first probe (V, I) data point: `v(1), a(1)`

(h) Do for `i=2,N-1` the following:

    i. Calculate I'' and store the value in the ith spot of an array i.e., `app(i)=a(i+1)+a(i-1)-2*a(i)`.

    ii. Write to the file `VI.dump.dat` the probe data: V, I, I'', i.e., `v(i), a(i), app(i)`.

(i) Write a line to the file: `VI.dump.dat` containing the last (`i=N`) data point V, I.

(j) Find the data point just before V_f with code[21] like:

```
        do i=2,N-1
                if(a(i)>0.and.a(i+1)>0)goto 100
        enddo
        STOP 'a<0'
100     ivf=i-1
        if(ivf.lt.15) STOP
```

The final line halts the program if `ivf`< 15, as we need Region D data to find the ion current, I_i. If the program `STOP`s, V_{min} will need to be reduced to capture this data, and the program re-run.

(k) Using Equation 7.61 and the data points at `ivf` and `ivf+1`, find V_f. Note: in the general case Equation 7.61, we were seeking x such that $y(x) = 0$; Here we are seeking V_f which is defined as the voltage such that $I(V_f) = 0$, so, for example, $x_1 \rightarrow$`v(ivf)`, $y_1 \rightarrow$`a(ivf)`, and $x \rightarrow V_f$.

(l) Find the data point just before V_p with code[22] like:

```
        do i=ivf+1,N-2
                if(app(i)<0.and.app(i+1)<0)goto 200
        enddo
        STOP 'app>0'
200     ivp=i-1
```

Note that the program is halted if the plasma potential has not been found before we run out of data. In that case V_{max} will need to be increased to capture this data, and the program re-run.

---

[21] The check for two successive data points gone positive is done to avoid mistaking one point of noise for V_f.

[22] The check for two successive data points with $I'' < 0$ is done to avoid mistaking one point of noise for V_p

(m) Using Equation 7.61 and the data points at `ivp` and `ivp+1`, find V_p. Note V_p is defined as the voltage such that $I''(V_p) = 0$, so, for example, $x_1 \rightarrow$`v(ivp)`, $y_1 \rightarrow$`app(ivp)` , and $x \rightarrow V_p$.

(n) Write comment lines to the file: `VI.fit.dat` recording:

    i. I_c, V_c, V_f, V_p (also `print` this to the screen, so you can monitor data collection)

    ii. `'!  set k1=',a(ivf-15),' k2=',a(ivp),' k3=',0.18615*`$(V_p - V_f)$

    The aim here is to record basic plasma parameters (which are also needed as an initial guess in the fit to Equation 7.70) in a format that allows easy copy and paste into `plot` and `fit`. `k1` is the measured ion current I_i; `k2` is the measured current at V_p (denoted I_p); `k3` is an estimate for kT_e/e, the electron temperature in eV.

    Notes: The estimate for `k3` is based on Equation 7.69 where $0.18615 = 2/\ln(2M_i/\pi m)$. We are assuming that `ivf-15` is far enough below V_f (i.e., in Region D) that $I_i = $ `a(ivf − 15)`. You should check (after the fact) that this is the case, i.e., $V_f - $`v(ivf − 15)`$ \gg $ `k3`. If this condition is not met, simply use a larger offset than `15` and re-run the program.

    iii. `'!  set k4=',  `V_p`, ' npoint=', ivp-ivf`

    `k4` is the (not adjusted) plasma potential, V_p, whereas `k1-k3` are varied from the above initial guesses to get the best fit to Equation 7.70 in Region C.

(o) For `i=ivf+1,ivp` write the probe data $V, I, \delta I, I''$ to the file `VI.fit.dat` with one $V, I, \delta I, I''$ data 'point' per line.

8. Turn off the output of the 2400.

9. Close all files.

Data Analysis

If all has gone well you have three good data sets for four different I_c, something like 1000 data points. We could spend the next semester analyzing this data! Instead I suggest below a simplified analysis scheme.

Start by making a composite plot similar to Figure 7.10A showing one IV curve for each of the four I_c. For these plots, choose V_{min} & V_{max} so that all behaviors (Regions A–D) are displayed. Make a similar plot showing three IV curves for one of the four I_c. The aim here is to test for reproducibility; because of time limitations we'll just check the reproducibility of this one selected I_c. (I'll call this selected I_c data set I_c^*.) These six data sets (one IV curve is on both plots) will be analyzed in greater detail. In order to properly `fit` the data, you will want to have more than 15 data points in the region between V_f and V_p (this may require reducing the range V_{min} through V_{max} for the probe voltage sweep so it focuses just on Region C: say from about 4 V negative of the lowest V_f to 2 V positive of the highest V_p and then re-taking the data). Additionally you should find that all six data sets have approximately the same number of data points. (Find the `npoint=` in the file `VI..fit.dat`.

Electron Temperature in eV: `k3`
Fit Equation 7.70 to the six data sets. Note that the file `VI.fit.dat` itself contains the required initial guesses for adjusted parameters `k1-k3` and the fixed value for the plasma

potential `k4` in a format that should allow easy copy & paste into `fit`. When using `fit` adjust only `k1-k3` (it would be 'degenerate' to have fit also change V_p=`k4`). Expect to see large reduced χ^2 which signals a too-simple model. *No definite meaning can be attached to error estimates from such poor fits*, nevertheless some sort of nonsense needs to be reported. On page 16 we explored options for dealing with such 'unusual' fits. Option #5: 'In dire circumstances' you can use `fudge` command in `fit` to change your errors so that a reduced χ^2 near one will be obtained. A re-fit with these enlarged errors then gives a new covariance matrix from which an estimate of parameter errors can be determined. Parameter values within the range allowed by these errors would produce as good (or bad) a fit to the data as the 'best' values. As the name suggests `fudge` is not exactly legitimate[23], nevertheless it is what Linfit has been silently doing all these years. Option #4: 'Bootstrap' the data: repeatedly fit subsets of the data and see how the resulting fit parameters vary with the multiple fits. This is essentially a way to repeat the experiment without taking new data. The `fit` command `boots` will report the results of `nboot` (default: 25) re-fits to subsets of your data along with the standard deviation of the fit parameters. Using either option, estimate the error in the fit value of `k3`: use the square root of the appropriate diagonal element of the covariance matrix of a `fudge`d fit or the reported standard deviation of `k3` from a `boot`strap.

For I_c^*, compare the two errors for k_3: σ (standard deviation of 3 fit values of k_3) and that from the fudged covariance matrix. Are they in the same ballpark (say, within a factor of two[24])? Calculate the electron temperature of the plasma both in eV and K. Comment on the relationship (if any) between T_e and I_c.

Plasma Number Density: n

We are interested in two sorts of 'error' in the plasma density n: reproducibility and an estimate of 'systematic' error. In the first case we're asking: "given the same I_c is applied, how much do conditions in the plasma vary?" In the second case we interested in what the actual value of n is. We can estimate this systematic error by measuring n by two different methods:[25] n calculated using I_i (Method A: let's call this n_i) and n calculated using I_p (Method B: let's call this n_p). Let's be clear here: n_i, n_p, n_e, n are all $\underline{\text{supposed}}$ to be the same thing: the number of electrons (or ions) per m^3 in the plasma. So

$$\frac{n_p}{n_i} = \frac{-I_p/I_i}{\sqrt{2M_i/\pi m}} \tag{7.71}$$

should be one; but it won't be: systematic error is present!

In your lab notebook record these results in tables similar that shown in Table 7.2[26]. (For this example, I selected $I_c^* = 20$ mA.) Note particularly to record exactly the proper number of significant figures in these tables! You should also copy & paste each full `fit` report into a long concatenated file and include the resulting file in your notebook. As previously noted, the reduced χ^2 for these fits is likely to be "horrendous". Pick out your highest reduced χ^2 fit, and plot the fitted function along with the data points on semi-log paper. (The results should look similar to Figure 7.11, but with just one data set.) Do the same for the best fit.

---

[23]for further discussion see: `http://www.physics.csbsju.edu/stats/WAPP2_fit.html`

[24]An estimate for the uncertainty in a calculated σ is: $\delta\sigma = \sigma/\sqrt{2(n-1)}$, so here with $n = 3$ points, 50% errors in σ are expected.

[25]See page 164; we are discussing here only methods A and B; methods using the fit parameters `k1` and `k2` would be additional options.

[26]The spreadsheet `gnumeric` may be of use.

Table 7.2: Simplified data table for reporting Langmuir Probe results.

k_3 (V)			I_c (mA)	k_3 (V)	fit error	σ
			40			
			20			
			10			
			5			

I_i (A)			I_c (mA)	median I_i	σ	% error in I_i	n_i (m^{-3})	% error in n_i
			40					
			20					
			10					
			5					

I_p (A)			I_c (mA)	median I_p	σ	% error in I_p	n_p/n_i
			40				
			20				
			10				
			5				

Calculate[27] n for each I_c based on the median I_i. Comment on the relationship (if any) between n and I_c. Make a power-law fit and log-log plot of the data[28]

Derive Equation 7.71. Calculate the n_p/n_i and see that n has systematic error, i.e., n calculated from I_p will be several times larger than the value of n calculated from I_i. This proves that there are problems with our simple theory.

Miscellaneous Calculations: 'Lawson Product' $n\tau$, λ_D, f_p
In 1957 J. D. Lawson determined that power generation from thermonuclear fusion required temperature, plasma density n and the plasma confinement time τ to meet certain criteria: a temperature of at least 10^4 eV with the product: $n\tau > 10^{21}$ m$^{-3} \cdot$ s. We can estimate τ because we know that our plasma is moving toward any surface at the Bohm velocity u_B. Given that typically the plasma is within about 1 cm of a wall, find how long it remains confined, and calculate the "Lawson product" $n\tau$ for the I_c^* plasma. (It is possible to directly measure the plasma confinement time by using a scope to time the decay of the plasma when the glow discharge is suddenly turned off.)

Calculate the Debye length (λ_D, Eq. 7.14) and the plasma frequency (f_p, Eq. 7.1) for I_c^*. Compare λ_D to the diameter of the Langmuir probe. Will the sheath substantially expand the collecting area $\mathcal{A}$? According to Koller (p. 140, reporting the results of Compton & Langmuir[29]) the mean free path of an electron in a 1 torr argon gas is 0.45 mm. Compare

---

[27]http://www.physics.csbsju.edu/cgi-bin/twk/plasma.html can do this in one click.
[28]I'd use WAPP$^+$: http://www.physics.csbsju.edu/stats/WAPP2.html.
[29]Rev. Mod. Phys. **2** (1930) 208

your λ_D to this electron mean free path.

Report Checklist

1. Write an introductory paragraph describing the basic physics behind this experiment. For example, why does the Langmuir probe current increase exponentially with probe voltage? Why is it that probe currents allow the calculation of plasma density? (This manual has many pages on these topics; your job is condense this into a few sentences and no equations.)

2. Computer program: Print out a copy of your program and tape it into your lab notebook.

3. Plots:

   (a) Similar to Figure 7.10A showing one IV curve for each of the four I_c.

   (b) Similar to Figure 7.10A but showing three IV curves for I_c^*.

   (c) Two plots similar to Figure 7.11, showing the Region C fit to the data. (These plots are to display the best and worst reduced χ^2; Record the reduced χ^2 on each plot.)

   (d) Power law fit and log-log plot of four (I_c, n) data points.

4. Tabulated results from six `fit`s for k_3 (T_e) at four different cathode currents including `fudge`d or `boots`trapped results for δk_3 and the standard deviation of three k_3 for I_c^*. You should copy & paste each full `fit` report into a long concatenated file. Print that file and tape it into your notebook. Record the identifying letter on your tube.

5. Tabulated results for measured I_i and I_p at four different cathode currents including reproducibility errors estimated from the standard deviation.

6. Calculations (self-document spreadsheet or show sample calculations):

   (a) kT_e in units if eV for four I_c with estimates for errors.

   (b) T_e in units of K for four I_c.

   (c) n calculated from median I_i and `k3` for four I_c (with and estimate of reproducibility error)

   (d) n_p/n_i calculated from median I_i and I_p (which provides an estimate of systematic error)

   (e) λ_D

   (f) f_p

   (g) Lawson product at I_c^*

7. Derivation of n_p/n_i (Eq. 7.71).

8. Answers to the questions posed in the Data Analysis section:

   (a) Are they in the same ballpark?

   (b) Comment on the relationship (if any) between T_e and I_c.

    (c) Comment on the relationship (if any) between n and I_c.

    (d) Will the sheath substantially expand the collecting area $\mathcal{A}$?

    (e) Compare your λ_D to this electron mean free path.

9. Discussion of errors:

    (a) Two methods were used to find "errors" in the electron temperature, T_e: (A) fudged/bootstrap error from a fit and (B) σ from lack of reproducibility. What is the meaning and significance of fudged/bootstrap error? What is the meaning and significance of reproducibility error? How would you respond to the question: "What is the error in T_e?" (Note a few sentences are required, not a number!)

    (b) Two methods were used to find the plasma number density, n: methods based on I_i and I_p. What is the meaning and significance of the fact that two different ways of measuring n produced different results. How would you respond to the question: "What is n?" (Note a few sentences are required, not a number!)

    (c) Consider any one of the basic plasma parameters (n, T_e, V_f, V_p) measured in this lab. Report any evidence that there is systematic error in the parameter. Report your best guess for the total error (systematic and random) in the parameter. Report how this error could be reduced.

Comment: Uncertainty

Area: Systemic Error

Both methods of calculating plasma density (n_i and n_p) used the probe area $\mathcal{A}$, which entered as an overall factor. The probe area was calculated based on the probe geometry data listed in Table 7.1 which was supplied by reference #9 with no uncertainties. Clearly if the actual probe geometry differs from that in Table 7.1 there will be a systematic error in both methods of calculating n. If I guess the uncertainty in length and diameter measurements based on the number of supplied sigfigs, I find a 6% uncertainty in $\mathcal{A}$. In 2008 a Sylvania 0A4G tube was destroyed, and I took the opportunity to measure the probe. The results[30] were quite different from those reported in Table 7.1: $\mathcal{A}$ was about 20% larger for this Sylvania tube then for the RCA tube of reference #9. (The Sylvania/GE 0A4G also looks different from the RCA 0A4G.)

There is an additional significant problem: because the plasma sheath extends several λ_D beyond the physical probe, there will be particle collection beyond the surface area of the probe—the effective area is larger than the geometric surface area; we have made a Spherical Cow by the simple assumption of $\mathcal{A} = \pi d\ell$. A glance at Fig. 7.6 on page 160, shows that the plasma sheath extends about $10 \times \lambda_D$ for $\Delta V \sim 5 \times kT_e/e$, which amounts to a large correction to probe diameter. Since the plasma sheath for ions[31] in Region D has no reason to be identical to the plasma sheath for electrons at the plasma potential, the effective $\mathcal{A}$ in the two methods is probably not the same, and hence $\mathcal{A}$ does not really cancel out in deriving Eq. 7.71.

---

[30] diameter= 0.025", length= .2", the surrounding metal can is about 2 cm×3 cm.

[31] For example, the mean free path for an electron is much larger than the mean free path for an ion (approximately $4\sqrt{2}\times$).

While we have lots of systematic and spherical cow error in our measurement of T_e and n, it should also be emphasized that we do have 'the right end of the stick'.

Interpreting Results

Inspection of Figure 7.11 shows the unmistakable signs of "large reduced χ^2": The fitted curve misses many error bars. The miss might be called "small", but the error bar is smaller still (in fact too small to show in this figure). The plasma density calculated from I_i disagrees with that calculated from I_p. What should be recorded as our uncertainty in T_e and n? The problem is the result of using simplified theory. What can we conclude from the fits using inexact[32] theory?

First the exponential *IV* relationship is clearly reflecting the Boltzmann factor at work. To the extent that there is <u>an</u> electron temperature, our k_3 estimate must be fairly accurate.

On the other hand our n values disagree by a factor of three, and there are reasons to suggest (e.g., uncertain $\mathcal{A}$) the systematic error may be even larger. The disagreement between n_i and n_p could be improved by better theory. (The assumed Maxwell-Boltzmann speed distribution, collisions, varying collecting areas due to sheath expansion, . . . can be corrected —see particularly References 3 & 5.) However, for many practical purposes one is interested in reproducible control rather measurement. One might be told that: "silicon wafer etching is to proceed when the <u>indicated</u> n reaches 10^{14} m^{-3}" with no concern for what n actually is. Usually the reproducibility error found in the lab is only a few percent, which is often good enough for industrial control.

Of course, physics is most interested in reality, and what we have found is systematic error: two different ways of measuring n disagree. While one might argue that calculation of n based on I_i is more robust than that based on I_p, fundamentally what is needed is additional methods of determining n to resolve the problem (see Reference 8). This is also beyond the aims of this lab. The best physics you can do based on this data is to report our estimated errors with a clear warning that the systematic errors may be larger than the reported uncertainty.

References

1. Maxwell & Benedict *Theory of Gaseous Conduction and Electronics* 1941 Chapter X

2. Koller *Physics of Electron Tubes* 1937 Chapter X

3. ed. Huddlestone & Leonard *Plasma Diagnostic Techniques* 1965 Chapter 4 by F. F. Chen Dewey 537.16028 H87

4. Chen *Introduction to Plasma Physics and Controlled Fusion* 1984 p. 295

5. ed. Auciello & Flamm *Plasma Diagnostics* 1989 Volume 1: Discharge Parameters and Chemistry, Chapter 3 by N. Hershkowitz (ILL Madison)

6. Hutchinson *Principles of Plasma Diagnostics* 2001 Chapter 3 (ILL Bemidji)

---

[32]Note that <u>all</u> theory is inexact.

7. `http://www.physics.uiowa.edu/~fskiff/Physics_132/Lab%20Manual/B4Langmuir.PDF`

8. `http://www.ee.ucla.edu/~ffchen/Publs/Chen206R.pdf`

9. I. Alexeff, J. T. Pytlinski, and N. L. Oleson Am. J. Phys. **45** (1977) 860–866, also see Am. J. Phys. **45** (1977) 1196–1199, Am. J. Phys. **46** (1978) 1276–1278, Am. J. Phys. **52** (1984) 753–754

Appendix A: `fit`

`fit` is a non-linear least-squares fitting program that allows you to fit (at most) nine parameters of a function to data contained in a file. The function must be expressed as a single line of intrinsic Fortran functions, real numbers, operations (`+`, `-`, `*`, `/`, `**` or `^`) and adjustable parameters `K1-K9`. Both `plot` and `fit` use the prompt "`*`" to request commands for action from the user.

A typical `fit` session begins by using the editor to create a data file. When creating the data file, follow this convention: reserve the first line of the file for an identifying sentence, and follow that line with columns of data separated by tabs or spaces. Typically you will put x values in the first column, y values in the second column, and y-errors (if available) in the third column. `fit` assumes this file is called `fit.dat`, but other filenames may be SET.

Once you have created the file, start the program by typing `fit` to the *linux* `$` prompt:

```
$ fit
*
```

(At any time you may exit `fit` by typing `quit`. **Note:** the common windoze command *Ctrl* C <u>kills</u> programs rather than copies; simply highlight by sweeping with mouse and then hit the center mouse button to paste at the desired location.) SET `fit`'s FILE variable equal to the *filename* of the file you just created by typing:

```
* SET FILE=filename
```

If you have no y-errors (an unlikely situation!), inform the program by typing

```
* SET YECOL=0
```

(Don't forget to hit the *Return* or *Enter* key after each line.)

If any other columns are out of place, reSET `fit`'s other column location variables: XCOL and YCOL. Note that `fit` will use no more than 512 data points. When `fit` is properly informed about your file type:

```
* READ
```

and `fit` will read your data into its internal data storage area.

The next step is to tell `fit` the function you want it to fit to your data. As stated above, you must be able to express this function in one line using the intrinsic Fortran functions (e.g., `log`, `log10`, `asin`, ...), constant real numbers (e.g., `2`, `6.03e23`), operations (`+`, `-`, `*`, `/`, `**` or `^`), and (possibly) varying parameters (`K1-K9`). Examples:

```
* set f(x)=k1*exp(-((x-k2)/k3)**2)

* set f(x)=k1*(1-exp(k2*x))

* set f(x)=k1/sqrt((x-k2)^2+k3^2)
```

The function should be typed *without* using spaces. Note that the program automatically takes the absolute value of negative numbers raised to a power (e.g., $(-2)^3 = 8$), but the `SQRT` of a negative number is zero. Use repeated multiplications to simulate integer powers (e.g., $x^3 = $ `x*x*x`), if the proper sign is required. Note that the default `f(x)` is the polynomial `k1` $+$ `k2` $*$ `x` $+$ `k3` $*$ `x`$^2 + \cdots +$ `k9` $*$ `x`^{8}, defined with repeated multiplication.

Now it's time to `SET` your initial guess for the `K` parameters that occur in your function, with a command like:

```
* SET K1=3.14159 K2=4
```

Unlike linear least squares programs like *WAPP*$^+$you must provide a moderately accurate initial guess for each adjustable parameter. Time spent in careful consideration of parameter values will be repaid many-fold. Bad initial parameter guesses will result in program crashes, long computer `fit` times, and inaccurate computer results. `fit` will step into the first chi-square (χ^2) valley (χ^2 minimum) it finds, and then stay there; so you must start `fit` close to the right valley. It is advisable to try several different starting guesses for your parameters to find the path that leads to the lowest value of χ^2.

Once you have a good initial guess for your `K` parameters, you may ask the computer to iterate to a minimum χ^2. Of course, you must also tell the computer what `K` parameters it should vary to find that minimum:

```
* fit
Enter list of Ks to vary, e.g.  K1-K3,K5:
```

After you tell the computer what `ks` to vary, the machine will start its search. How does `fit` know when to stop? It will, of course, never reach the exact valley bottom (i.e., χ^2 minimum). If `fit` "**finished**" there is some evidence of convergence: either χ^2 changed hardly at all during its last downward step ($\Delta\chi^2$=`DELCHI`, default=.1), or the parameters' first `KSIG` digits (default=4) didn't change during its last downward step. Note: unchanging digits for one step, should not be interpreted as unchanging digits for lots more steps, nor should it be interpreted as an error. If `fit` "**quit**" then it's sick of trying (`SET` by `NITER`, default=100), so you should assume more steps are needed to reach the valley bottom. This suggests something is wrong...your initial guesses, the function, the data.

Once `fit` has quit or finished, it will print out summary results. Longer versions of the results (including the covariance matrix needed to figure errors[1]) may be printed, typed on the screen or placed in a file, at the users option. Always retain a copy of successful fit results! Sometimes the found χ^2 is so "bad" that you will be tempted to `FUDGE`, i.e., adjust your y-errors until you get a reduced-χ^2 of one. If you `FUDGE` be sure to write down why you think this cheating was justified.

A good way to estimate errors is to repeat the experiment lots of times. However very often the required time is not available. `fit` provides a possibility legitimate way to simulate actual repeats: it repeatedly fits using random subsets of your data. This process is called bootstrapping and it is invoked by the command `BOOTST`.

The command

    * PRINT

will display your data (x, y, y-error, and $f(x)$) on the screen. `fit` will mark (?, !, :) points that are 3, 2, or 1 y-errors away from $f(x)$. Use this feature to see which data points are poorly fit. Alternatively, you may use the program `plot` to display your data with fitted curve.

You may find it helpful to do some parameter stepping yourself before you hand the final fit over to `fit`. Put in a parameter guess, enter

    * CHI

to find the resulting χ^2, modify your guess and repeat the process.

`fit`ting functions to data is an important and frustrating part of your education in physics. Good luck!

Report problems and/or suggestions to Tom Kirkman.

---

[1]If the reduced χ^2 suggests a reasonable fit, the parameter uncertainties may be calculated from the square root of the corresponding diagonal element of the covariance matrix.

Special Commands

Start—in a terminal type: `fit`
`$`*linux* command — do any *linux* command from inside `fit`
`@`*filename* — execute indirectly a file of `fit` commands
Examples:

      * `$ls *.dat`

      * `@test.fit`

`fit` Commands

(unambiguous abbreviations and lowercase allowed)

Syntax:

      * `COMMAND`

      * `COMMAND VARIABLENAME =` value`, VARIABLENAME =` value `...`

Examples:

      * `SET K1=0, K2=1, K3=-1, F(X)=K1+K2*X**2+K3*X**4`

      * `re yec=0, fi=expt.dat`

`BOOTST`   performs a bootstrap with `NBOOT` iterations

`CHI`   calculates χ^2 using `F(X)`, Ks, and data.

`FIT`   fits `F(X)` to your data (requests list of Ks to vary)

`FUDGE`   destroys existing y-errors and renorms them so $\chi^2 = $ `NPOINT`

`HELP`   displays these messages

`PRINT`   displays data stored in program and `F(X)` (marks large-deviation points)

`QUIT`   exit the `fit` program

`READ`   reads data from `FILE`: see `XCOL, YCOL, YECOL, ROW, IBEGIN, NPOINT, FILE`

`SET`   allows you to set variables values without doing any action

`SHOW`   shows variable values

`WRITE`   writes x, y, y-error, and $f(x)$ to `FILE`

`fit` **Variables**

A   constant for use in `F(X)` (0)

B   same
     (Note: `A` and `B` are used in `WINDOW(X)`: `WINDOW(X)=1` iff $a \le x \le b$, else zero)

C   same

CHISQ most recent value of χ^2 calculated

DELCHI `fit` stops if a step results in a χ^2 change less than `DELCHI` (.1)

F(X)   function use: `A`, `ABS`, `ACOS`, `ASIN`, `ATAN`, `B`, `C`, `COS`, `COSH`, `EXP`, `INORM`,
     `K`, `K1-K9`, `LOG`, `LOG10`, `NORM`, `PI`, `SIN`, `SINH`, `SQRT`, `TAN`, `TANH`, `WINDOW`,
     `X`, any real number, operations `+`, `-`, `*`, `/`, `**`, or `^`
     (`K1+X*(K2+X*(K3+X*(K4+X*(K5+X*(K6+X*(K7+X*(K8+X*K9)))))))`)

FILE   `READ`/`WRITE`s data with this filename (`fit.dat`)

IBEGIN first array element used during `CHI`, `FIT`, `READ`, `PRINT`, `WRITE` (1)

K1-K9 possibly adjustable variables in `F(X)`

KSIG   `fit` stops after a step if all Ks retain `KSIG` unchanged digits (4)

NBOOT number of bootstrap iterations performed (25)

NITER   `fit` stops if more than `NITER` test steps are needed (100)

NPOINT number of points used during `CHI`, `FIT`, `READ`, `PRINT`, `WRITE` (512)

ROW   first row of data in `FILE` (2)

SEED   if negative, re-initializes random generator's seed (-1)

XCOL   column of x data in `FILE` — used during `READ` (1)

YCOL   column of y data in `FILE` — used during `READ` (2)

YECOL   column of y-error data in `FILE` — used during `READ` (3)
     Note: unread y-errors default to 1

Appendix B: `plot`

`plot` is a general purpose, advanced plotting program. `plot` can be frustrating to use because it assumes you know what you want to do and actually mean what you say—so if you don't know what you're doing and don't know what you say means to the machine, you can draw a mess fast. Of course, after you know the program you should be able to draw a beautiful graph as fast as you drew that mess.

The most unusual feature of `plot` is the way you give values to the parameters needed to draw a graph, e.g., the size of the graph, maximum and minimum values, etc. Instead of asking you if you want to change each of those 39 parameters, `plot` allows you to set them in any order, at any time you want. For example, if you want to change the minimum value of x to 0, you type

```
* SET XMIN=0
```

(Don't forget to hit the *Return* or *Enter* key after each line.)

So the first thing you need to know is the name of the parameters. You will find a complete list at the end of this appendix but let me mention three important sets of parameters now. The first set determines the way the axes look: `XMIN, XMAX, XSCALE` and similarly for y. `XMIN` and `XMAX` should be self explanatory. `XSCALE` determines the nature of the scale:

`XSCALE=1` is a linear (normal) scale;

`XSCALE=2` is a logarithmic scale (distance in plot is proportional to $\log x$);

`XSCALE=3` is an inverse scale (distance in plot is proportional to $\frac{1}{x}$); and

`XSCALE=4` is a probability scale (distance in plot is related to $\mathrm{erf}^{-1}(x)$).

(Note: you will rarely actually set `XMIN` and `XMAX` as there are autoscale commands, but you will often be setting the type of scale.)

The second set determines how the program finds the proper data to plot. Basically to get numbers (x, y, and errors) into the program you need to make a file with the editor. The first line of this file should be a one-line description of the data in the file (i.e., `free fall experiment - trial #2`). The following lines should consist of data—an x value, some spaces (or tabs or commas), the corresponding y value, some spaces, and the corresponding absolute error in y. Thus when looked at as a whole, the first column will consist of x values, the second column the corresponding y values, etc. If you called this file `plot.dat` you have made a file exactly in the default manner. The program will be able to `READ` this

file without telling it anything. If you put the x values in column 2, you will have to inform the program by typing

```
* SET XCOL=2
```

That is, `XCOL` is a variable that tells the program which column contains the x values. Similarly, `YCOL,` `YECOL` (y-error) and `XECOL` (x-error) are variable names. If you have no errors, tell the program by typing

```
* SET YECOL=0
```

If you didn't call your file `plot.dat`, you must tell the program by typing

```
* SET FILE=filename
```

Note the above `SET`ing does not actually change the data inside the program. The data inside the program is only changed when you:

```
* READ
```

Finally the third set determines how functions are plotted.

```
* SET F(X)=SIN(PI*X)
* SET PMIN=-1 PMAX=1
* FCURVE
```

`FCURVE` will plot the function in the domain `PMIN` to `PMAX` (the "p" is for parameter— remember `xmin` and `xmax` have another meaning). Note that the border parameters (`ymin`, `ymax`, etc.) must be appropriately[1] set before the function curve is added to the plot. `F(X)` can be set using the usual computer syntax[2] for algebra. There are eleven built-in functions that correspond to the usual $WAPP^+$ functions with the usual parameters `A`, `B`, and `C`. These can be plotted by `SET`ting the function number `NFUNCT` to a value between 1 and 11. `NFUNCT=0` (the default) plots the function `F(X)`.

You use `SET` to tell the program on how you want your data displayed. To actually display data you need action commands. The sequence:

```
* READ
* SCALE
* BORDER
* DPOINT
```

---

[1]`FSCALE` will find appropriate `ymin`, `ymax`, etc., based on the <u>function</u> `f(x)` and domain `PMIN` to `PMAX`, but more commonly the border parameters — including `PMIN` and `PMAX` — are set to fit the <u>data</u> with `SCALE`

[2]Note that you must use `*` for multiplication not just juxtaposition.

tells the program to read the data (from the previously set `FILE`); `SCALE` tells the program to find appropriate `XMIN`, etc. given the data contained in the program; `BORDER` tells the program to display the graph outline you have described; `DPOINT` tells the program to plot the data points. Type `FCURVE` to plot your function. Another useful command is `HELP`—it types out a listing of the commands and the variables.

Actually all this `SET`ting and commanding can be done together as shown below: (note one command per line, but you may define several variables)

```
* READ YECOL=0, FILE=freefall1.dat
* SCALE
* BORDER TITLE='Free Fall', XLABEL='Time (s)', YLABEL='Z (m)'
* DPOINT
* FCURVE PMIN=0, PMAX=.65, NFUNCT=10, A=0, B=0, C=4.9
```

Two final comments: The equal signs and commas in the above example make the reading of the line easier—but they make the typing of the line harder. You may substitute spaces for them to make typing easier. Similarly, one gets tired of typing in the full command in capitals—lowercase and any unambiguous abbreviations may also be used. Thus instead of the above you may type:

```
* re yec 0 fi freefall1.dat
* sc
* bo tit 'FREE FALL' xl 'TIME (s)' yl 'Z (m)'
* dp
* fc pmin 0 pmax .65 nfu 10 a 0 b 0 c 4.9
```

Getting Hardcopy Plots

Now that you have a plot displayed on the terminal, you should want a copy of the screen to put in your lab notebook. The command `PCOPY` will make a copy of the displayed plot to the file specified by `PFILE`. If `PFILE="  "` (as it is by default), the output goes to the printer.

Special Commands

Start—in a terminal type: `plot`
$*linux* command — do any *linux* command from inside `plot`
@*filename* — execute indirectly a file of `plot` commands
Examples:

```
* $ls *.dat
* @fet.plt
```

Report problems/suggestions to Tom Kirkman.

`plot` **Commands**

(unambiguous abbreviations and lowercase allowed)

Syntax:

> * COMMAND
> * COMMAND VARIABLENAME = value, VARIABLENAME = value ...

Examples:

> * BORDER XMIN=0, XMAX=1, YMIN=-1, YMAX=1, TITLE='now is the time'
> * bo xmi 0 xma 1 ymi -1 yma 1 ti 'now is "the time"'

BORDER draws a border for a graph

CLEAR clears the screen

CROSSH digitizing mode—using the mouse move the 'crosshair' to the desired point. Hold the Shift key and click; the x and y value at crosshair is reported.

DCURVE connects with a LINE NPOINT data points, starting with IBEGIN.

DPOINT plots NPOINT individual data points starting with IBEGIN; point symbol determined by POINT, error bars by XECOL and YECOL

FCURVE plots the function NFUNCT for p from PMIN to PMAX in NSTEP steps. $p = x$ for NFUNCTs 0–11, NFUNCT=0 displays F(X), link in funct(x,y,p,nfunct) for NFUNCT > 11

FSCALE finds appropriate mins and maxs for x and y axes for FCURVE

HELP displays these messages

LABEL writes LABEL to the right of most recent CROSSHair

PCOPY copies plot to printer (lpr -Plp) or to PFILE

PRINTD displays data stored in program

QUIT exit the `plot` program

READ reads data from FILE— see XCOL, YCOL, XECOL, YECOL, ROW, IBEGIN, NPOINT

SCALE finds appropriate min and max for x and y axes from NPOINT data pairs starting at IBEGIN (automatic 5% excess unless scale=0). Room for error bars included if x & y ECOL not zero

SET allows you to set variables values without doing any action

SHOW shows variable values

`plot` **Variables**

(Default settings are indicated in parentheses.)

A   used to give values to function parameters (see `NFUNCT`) (0)

B   same
   (Note: `A` and `B` are used in `WINDOW(X)`: `WINDOW(X)=1` iff $a \leq x \leq b$, else zero)

C   same

F(X) function: use `A`, `ABS`, `ACOS`, `ASIN`, `ATAN`, `B`, `C`, `COS`, `COSH`, `EXP`, `INORM`, `K`, `K1-K9`, `LOG`, `LOG10`, `NORM`, `P0-P9`, `PI`, `SIN`, `SINH`, `SQRT`, `TAN`, `TANH`, `WINDOW`, `X`, any real number, operations `+`, `-`, `*`, `/`, `**`, or `^` (default: 8th degree polynomial)

FILE `READ` seeks data from a file with this name (plot.dat)

FONT character size— 1:small, 2:standard, 3:big, 4:huge; 3&4 plotter only (2)

IBEGIN first array element used during `READ`, `DPOINT`, `DCURVE`, `SCALE`, `PRINT` (1)

LABEL string to be placed in plot using `CROSSH` and `LABEL` (LABEL)

LANG graphics language (1)

LINE pattern— 1:solid, 2:dotted, 3:dashed, 4:long-dash, 5:DASH-dot, 6:DASH-dash, 7:DASH-dash-dash (plotter only) (1)

NFUNCT 0:`F(X)`, 1:$y = a + bx$, 2:$y = ax^b$, 3:$y = a\exp(bx)$, 4:$y = a + b\ln(x)$, 5:$y = a + b/x$, 6:$1/y = a + bx$, 7:$1/y = a + b/x$, 8:$y = a\exp(b/x)$, 9:$1/y = a + b\ln(x)$, 10:$y = a + bx + cx^2$, 11:$y = ax^b$ >11:link funct(x,y,p,nfunct) (0)

NPOINT number of points used during `READ`, `PRINT`, `DPOINT`, `DCURVE`, `SCALE` (512)

NSTEP number of subdivisions of p range — used during `FCURVE` (100)

PATSIZ spacing of `LINE` pattern in fraction of screen width (.04)

PEN   plotter pen number (1)

POINT type of data symbol 0:dot, 1:circle, 2:square, 3:diamond, 4:triangle up, 5:triangle down; `POINT`<0 for filled symbol (2)

POISIZ size of symbol in fraction of screen width (.004)

PMAX maximum value of p — used during `FCURVE` (p equals x for `NFUNCT`=0–11) (1)

PMIN minimum value of p — used during `FCURVE` (0)

ROW   first row of data in `FILE`— used during `READ` (2)

TERM 1:visual 550, 2:selanar, 3:else (2)

TICSIZ size of axes ticks in fraction of screen (.015)

`TITLE`	title of graph—put on during `BORDER` (GRAPH)
`XCOL`	column of x data in `FILE`— used during `READ` (1)
`XECOL`	column of x-error data in `FILE`— used during `READ, SCALE, DPOINT`; `XECOL`=0 suppresses x-error bars (0)
`XLABEL`	label on x axis— put on during `BORDER` (X)
`XMAX`	largest x value in plot (1)
`XMIN`	smallest x value in plot (0)
`XOFF`	offset of the start of the x axis— unit: fraction of screen (.18)
`XSCALE`	type of x scale— 1:linear, 2:logarithm, 3:inverse, 4:probability (1)
`XSIZE`	size of x axis— unit: fraction of screen (`XSIZE` + `XOFF` < 1.) (.8)
`YCOL`	column of y data in `FILE`— used during `READ` (2)
`YECOL`	column of y-error data in `FILE`— used during `READ, SCALE, DPOINT`; `YECOL`=0 suppresses y-error bars (3)
`YLABEL`	label on y axis— put on during `BORDER` (Y)
`YMAX`	largest y value in plot (1)
`YMIN`	smallest y value in the plot (-1)
`YOFF`	offset of the start of the y axis— unit: fraction of screen (.1)
`YSCALE`	type of scale— 1:linear, 2:logarithm, 3:inverse, 4:probability (1)
`YSIZE`	size of y axis—unit: fraction of screen (`YSIZE` + `YOFF` < .75) (.6)

If there were some machine, somewhere, that could spit out universes with randomly chosen values for their fundamental constants, then for every universe like ours it would produce 10^{229} duds.

Though I haven't sat down and run the numbers on it, to me this seems comparable to the probability of making a Unix computer do something useful by logging into a tty and typing in command lines when you have forgotten all of the little options and keywords.

In the Beginning. . . Was the Command Line by Neal Stephenson (1999)

Appendix C: `fit` & `plot` **Example**

This appendix provides a longer example using `fit` and `plot` to analyze and display data. A good way to start is to watch the online video: `AppendixC_live.ogg` available at the class web site and youtube `http://youtu.be/_fWpFLcTm50`

The example data set is a recapitulation of work done in your last PHYS 200 lab on a.c. circuits and resonance: the current flowing in a LRC circuit is measured as a function of the frequency f of the a.c. voltage source. You may recall[1] that because of frequency dependent reactances, a series LRC circuit has a current, I, that depends on frequency. You can follow along with this example if you grab a copy of the data file. At the *linux* `$` prompt type:

    $ cp /usr/local/physics/help/LRC.dat .

and a copy of the datafile should appear in your current directory (which is called "."). Similarly you can get a copy of the actual commands used in the below example by:

    $ cp /usr/local/physics/help/LRC.com .

Theoretically the current as a function of frequency should be given by:

$$I = \frac{V}{\sqrt{R^2 + \left(\frac{1}{2\pi f C} - 2\pi f L\right)^2}} \tag{C.1}$$

Start the program `fit` by typing to the *linux* `$` prompt:

[1]or see Electrical Measurements Review particularly Fig. 1.1 on page 22
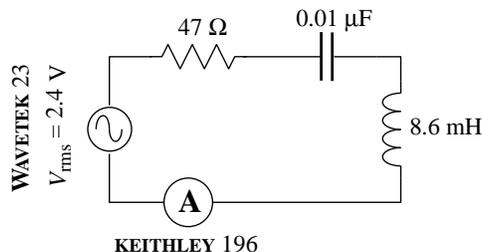
Figure C.1: This series LRC circuit was constructed. A WAVETEK 23 function generator provided a GPIB-controlled sine wave; a KEITHLEY 196 reported the resulting current via GPIB. The program `LRC.f` controlled the data collection.

```
$ fit
*
```

The program responds with its own prompt: `*` . We need to tell `fit` the name of the datafile we want to analyze and then read in that data:

```
* set file=LRC.dat
* read
freq(Hz) I(A) dI(A) by default this line ignored!
READ Done.
```

The first line of the file (`freq(Hz)`...) is simply echoed back as a way of confirming that you are reading the intended file. It should be noted the file `LRC.dat` was written in the default form, with x values in the first column (`XCOL=1`) , y values in the second column (`YCOL=2`), and y-error values in the third column (`YECOL=3`). Next we need to tell `fit` what function should match the data.

```
* set f(x)=k1/sqrt(k2^2+(1/(2*pi*k3*x)-2*pi*k4*x)^2)
```

where $V = k_1$, $R = k_2$, $C = k_3$, $L = k_4$ and $f = x$. For nonlinear fits we must always give starting parameter estimates. In this case that's easy as I measured the component values when I built the circuit:

```
* set k1=2.4 k2=47 k3=1e-8 k4=.0086
```

Next I command a `fit`: please adjust the parameters k_1–k_4 to achieve the smallest possible reduced χ^2:

```
* fit
Enter list of Ks to vary, e.g.   K1-K3,K5 k1-k4

FIT finished with change in chi-square= 1.5258789E-04
6 iterations used
REDUCED chi-squared= 3.790897 chi-squared= 363.9261
K1= 2.262721 K2= 190.7750 K3= 0.9930065E-08
K4= 0.8230284E-02
Display covariance/curvature matrices?  No, Screen, File [N,S,F]
```

The result is pretty much as expected: reduced χ^2 is high, but in the usually acceptable range, the parameters are close to my original guess, with the exception of $k_2 = R$. On reflection even R is OK as the inductor itself has resistance (77.8 Ω) and according to Fig. 1.1 on page 22, the function generator also has an internal 50 Ω resistor. Thus: $47 + 77.8 + 50 = 167.8$ Ω should have been expected. Currently `fit` is asking if I want to look at the covariance matrix, but I'll delay replying <u>N</u>o, <u>S</u>creen, or <u>F</u>ile until I look at the fitted curve.

I start another terminal to display, using the program `plot`, what the fitted curve looks like: at the *linux* `$` prompt I type:

```
$ plot
```

A new blue terminal (labeled plot.exe) pops up with its own * prompt. Exactly as with `fit` I must tell the program the name of the datafile and then read in the data:

```
* set file LRC.dat

* read

freq(Hz) I(A) dI(A) by default this line ignored!
READ Done.
```

The x-data (frequency) ranges from 10000 to 30000; y-data (current) ranges from .002 to .012. I could set by hand those ranges (e.g., `XMIN=10000`), but its much easier to have the program make what it thinks are reasonable choices using the command:

```
* scale
```

I could now draw a border and display the datapoints, but first I tell `plot` the proper axes labels:

```
* set xlabel='Frequency (Hz)' ylabel="RMS Current (A)" title=Resonance

* border

* dpoint
```

`Plot` is to use exactly the same function as `fit`:

```
* set f(x)=k1/sqrt(k2^2+(1/(2*pi*k3*x)-2*pi*k4*x)^2)
```

and I can copy&paste[2] from the `fit` screen the values `fit` found for those parameters:

```
* set K1= 2.262721 K2= 190.7750 K3= 0.9930065E-08

* set K4= 0.8230284E-02
```

Then I can have `plot` display that function curve and print a hardcopy:

```
* fcurve

* pcopy
```

---

[2]Warning repeated: *Ctrl* C kills not copies; highlight and middle-mouse for copy

Resonance
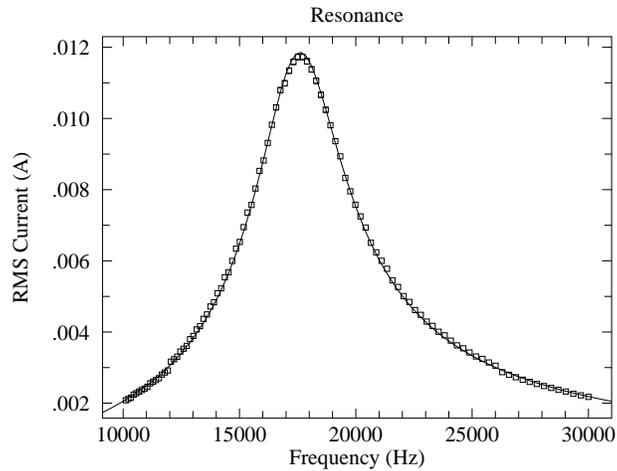
If I go back to the `fit` screen and type `s` in answer to the question about matrices:

```
Display covariance/curvature matrices? No, Screen, File [N,S,F] s

 Results of a fit to the function
 f(x)=K1/SQRT(K2^2+(1/(2*PI*K3*X)-2*PI*K4*X)^2)
 at 15:15 on 13-MAY-2009
 found a REDUCED chi-square of     3.790897      and a chi-squared of     363.9261
 using the following parameter estimates
 K1=   2.262721         K2=    190.7750              K3=  0.9930065E-08
 K4=  0.8230284E-02
 and 100 data points at row    2 in the file:
 LRC.dat


 Below are displayed the covariance matrix = inverse curvature matrix and the
 curvature matrix = .5 * Hessian = .5 * matrix of 2nd partials of chi-square
 see Numerical Recipes 15.5-15.6 for a discussion of these matrices

 Matrix form:          K1 K2 K3 K4
                   K1
                   K2
                   K3
                   K4

 COVARIANCE MATRIX:
  -.55E+01   -.46E+03   0.24E-07   -.20E-01
  -.46E+03   -.39E+05   0.20E-05   -.17E+01
  0.24E-07   0.20E-05   -.11E-15   0.87E-10
  -.20E-01   -.17E+01   0.87E-10   -.72E-04

 CURVATURE MATRIX:
  0.37E+06   -.12E+04   0.34E+14   -.34E+08
  -.12E+04   0.83E+01   -.54E+11   0.63E+05
  0.34E+14   -.54E+11   0.53E+23   0.56E+17
  -.34E+08   0.63E+05   0.56E+17   0.76E+11
```

The error results are a total disaster! The usual rule is that the error in a parameter is given by the square root of the diagonal element of the covariance matrix. Thus (supposedly):

$$\delta k_1 = \sqrt{-.55 \times 10^1} \tag{C.2}$$

$$\delta k_2 = \sqrt{-.39 \times 10^5} \tag{C.3}$$

$$\delta k_3 = \sqrt{-.11 \times 10^{-15}} \tag{C.4}$$

$$\delta k_4 = \sqrt{-.72 \times 10^{-4}} \tag{C.5}$$

Clearly this is some sort of nonsense as the errors cannot be imaginary. Further note that in every case $|\delta k_i| \sim k_i$.

A hint about what's going on is gained by making a different starting guess for the values of the parameters:

```
* set k1=240 k2=.47 k3=1e-7 k4=.00086
```

Fit results show the same reduced χ^2, but with radically different parameters:

```
REDUCED chi-squared=    3.790896       chi-squared=    363.9260
K1=   0.2419768         K2=   20.40159            K3=   0.9285606E-07
K4=   0.8801518E-03
```

There is a curve of (L, R, C, V) values that produces exactly the same resonance curve. Substituting

$$L \rightarrow \alpha L \tag{C.6}$$

$$C \rightarrow C/\alpha \tag{C.7}$$

$$R \rightarrow \alpha R \tag{C.8}$$

$$V \rightarrow \alpha V \tag{C.9}$$

$$\tag{C.10}$$

into Eq. C.1 results in an unchanged function (the α cancels out).

We can rewrite our function with just three parameters:

$$I = \frac{I_0}{\sqrt{1 + Q^2 \left(\frac{f_0}{f} - \frac{f}{f_0} \right)^2}} \tag{C.11}$$

where:

$$I_0 = V/R \tag{C.12}$$

$$Q = \frac{\sqrt{L/C}}{R} \tag{C.13}$$

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{C.14}$$

$$\tag{C.15}$$

This reparametrization of the function is both required and easier, in the sense that initial guesses for the parameters can be immediately obtained from the raw data: $I_0 \sim .012$ A is the current at resonance, $f_0 \sim 17600$ Hz is the resonant frequency, and the quality factor Q is $f_0/\Delta f$, where $\Delta f \sim 19500 - 15800 = 3700$ is the full width at $I = I_0/\sqrt{2} \sim .0085$ A.

```
* set f(x)=k1/sqrt(1+k2^2*(k3/x-x/k3)^2)
* set k1=.012 k2=4.7 k3=17.6e3
* fit
Enter list of Ks to vary, e.g. K1-K3,K5 k1-k3
 FIT finished with change in Ks implying 4 significant figures
   2 iterations used
 REDUCED chi-squared=   3.751816      chi-squared=   363.9261
 K1=  0.1186070E-01     K2=   4.772110          K3=   17605.02
Display covariance/curvature matrices? No, Screen, File [N,S,F] s

 Results of a fit to the function
 f(x)=K1/SQRT(1+K2^2*(K3/X-X/K3)^2)
 at 16:05 on 13-MAY-2009
 found a REDUCED chi-square of    3.751816     and a chi-squared of    363.9261
 using the following parameter estimates
 K1=  0.1186070E-01     K2=   4.772110          K3=   17605.02
 and 100 data points at row   2 in the file:
 LRC.dat


 Below are displayed the covariance matrix = inverse curvature matrix and the
 curvature matrix = .5 * Hessian = .5 * matrix of 2nd partials of chi-square
 see Numerical Recipes 15.5-15.6 for a discussion of these matrices

 Matrix form:          K1 K2 K3
                  K1
                  K2
                  K3

 COVARIANCE MATRIX:
  0.52E-09  0.25E-06  -.68E-06
  0.25E-06  0.14E-03  -.77E-03
  -.68E-06  -.77E-03  0.16E+02

 CURVATURE MATRIX:
  0.13E+11  -.24E+08  -.60E+03
  -.24E+08  0.52E+05  0.15E+01
  -.60E+03  0.15E+01  0.63E-01
```

We now have reasonable results for errors:

$$\delta k_1 = \sqrt{.52 \times 10^{-9}} = 2.3 \times 10^{-5} \text{ A} \tag{C.16}$$

$$\delta k_2 = \sqrt{.14 \times 10^{-3}} = .012 \tag{C.17}$$

$$\delta k_3 = \sqrt{.16 \times 10^2} = 4 \text{ Hz} \tag{C.18}$$

Another way of estimating errors is to bootstrap:

```
    * boots
```

The results:

```
 MEANS
 1.187386E-02     4.77916          17605.1
 STANDARD DEVIATIONS
 2.891049E-05   1.747892E-02     8.34518
```

are similar to those above. With the largish reduced χ^2 one might want to enlarge y-error values until reduced $\chi^2 \approx 1 \ldots$ `fit` calls that a `fudge`. Note that once the y-errors have been modified the only way to return to the unmodified data is to re-read the file.

```
* fudge
* fit
Enter list of Ks to vary, e.g. K1-K3,K5 k1-k3
 FIT finished with change in chi-square=  9.8335266E-02
   1 iterations used
 REDUCED chi-squared=  0.9858394      chi-squared=   95.62643
 K1=  0.1186207E-01     K2=   4.772794          K3=   17605.02
Display covariance/curvature matrices? No, Screen, File [N,S,F] s

 Results of a fit to the function
 f(x)=K1/SQRT(1+K2^2*(K3/X-X/K3)^2)
 at 16:15 on 13-MAY-2009
 found a REDUCED chi-square of   0.9858394     and a chi-squared of    95.62643
 using the following parameter estimates
 K1=  0.1186207E-01     K2=   4.772794          K3=   17605.02
 and 100 data points at row   2 in the file:
 LRC.dat

 Below are displayed the covariance matrix = inverse curvature matrix and the
 curvature matrix = .5 * Hessian = .5 * matrix of 2nd partials of chi-square
 see Numerical Recipes 15.5-15.6 for a discussion of these matrices

 Matrix form:           K1 K2 K3
                    K1
                    K2
                    K3


 COVARIANCE MATRIX:
  0.20E-08  0.94E-06  -.26E-05
  0.94E-06  0.51E-03  -.29E-02
  -.26E-05  -.29E-02  0.60E+02
```

Once the error bars have been expanded (`fudge`) this new `fit` gives a new covariance matrix

from which fudged errors can be determined:

$$\delta k_1 = \sqrt{.20 \times 10^{-8}} = 4.5 \times 10^{-5} \text{ A} \qquad (C.19)$$

$$\delta k_2 = \sqrt{.51 \times 10^{-3}} = .023 \qquad (C.20)$$

$$\delta k_3 = \sqrt{.60 \times 10^2} = 7.7 \text{ Hz} \qquad (C.21)$$

What is the source of the uncomfortably large reduced χ^2? If you look at the plot you'll notice that there are glitches at the two frequencies where $I = .003$ A — that is where the ammeter has automatically switched scales. A scale switch can result in different systematic error and it certainly results in a different voltage burden[3]. If we just fit to the data with $I > .003$ A we can avoid those effects. This data lies between rows 18 and 88 in the file `LRC.dat`; we can selectively have `fit` read this data:

```
* set row=18 npoint=70
* read
* print
```

Note that when you `read` all 17 rows of unwanted stuff in the file is echoed to your screen. The command `print` displays the data that is inside the program, allowing eyeballs to confirm that `fit`'s current data has $I > .003$ A. The `fit` result is:

```
 REDUCED chi-squared=  0.6348079       chi-squared=   42.53213
 K1=  0.1174845E-01     K2=   4.634116          K3=   17598.93
```

One can achieve almost the same thing by telling `fit` to limit its analysis to a subset block of the data it holds. When the data was originally read in, it was stored in adjacent cells labeled 1,2,3,...,`npoint`. We can tell `fit` to analyze any subset block of that data by changing which cell holds the "first" data point (`ibegin`) and the length of the run of wanted data.

```
    * set ibegin=17 npoint=70
```

In this case we should not re-`read` the data, as the intent is to use the data already held in `fit`. (Earlier in this example we changed the y-errors when we commanded `fudge`, so here we really do need to re-read the unmodified data and errors rather than just use a subset of the modified data/errors.)

We can use `plot` to display this new fitted curve with the entire dataset. All we need to do is change the function, set the newly found parameters k_1-k_3 via copy&paste from `fit`, and then redraw the plot:

```
* clear
* set f(x)=k1/sqrt(1+k2^2*(k3/x-x/k3)^2)
* set  K1=0.1174845E-01   K2=4.634116      K3=17598.93
* border
* dpoint
* fcurve
```

[3]see page 25
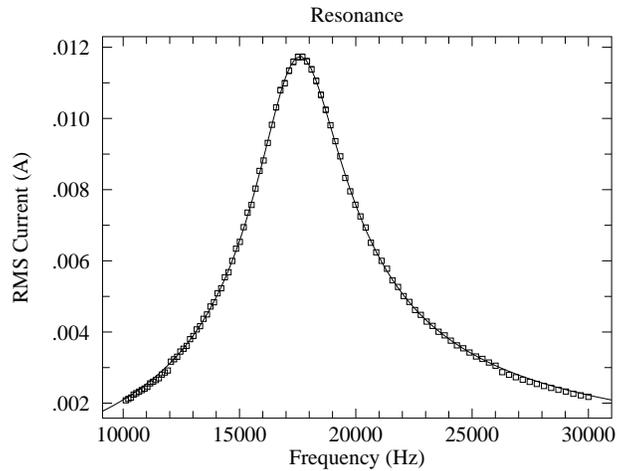
Mathematica

The program *Mathematica* can also find these nonlinear fits; of course it uses a different set of commands and data structures. You can follow along with this example if you grab a copy of the (slightly modified[4]) data file. At the *linux* $ prompt type:

```
$ cp /usr/local/physics/help/LRCm.dat .
```

and a copy of the datafile should appear in your current directory (which is called ".").

There are two version of *Mathematica* a command line version (`math`) and a browser version[5] (`mathematica`). The commands are the same in the two versions; the main difference is an absence of user friendly features—like arrow-key line editing, on-line help, drop-down menus—in the kernel version (`math`). I every much prefer the no-frills `math` version, but it is an acquired taste.

In a terminal, `cd` to the directory that contains the file `LRCm.dat`. Start *Mathematica* by typing to the *linux* $ prompt:

```
$ mathematica
```

or, if you're hard-core:

```
$ math
```

The file `LRC.com` that you previously copied also contains the *Mathematica* commands discussed below. Note that in `mathematica` to execute a command you need to simultaneously hit Shift and Enter, whereas in `math` it's just the usual Enter.

In *Mathematica* data is stored in set notation: `xy`=$\{\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_N, y_N\}\}$ with the errors stored in a different set (which must of course be in the same order):

---

[4]the first row of text has been removed

[5]i.e., with a <u>G</u>raphical <u>U</u>ser <u>I</u>nterface—GUI, pronounced "gooey"—with menus and line editing

`ey=`$\{\delta y_1, \delta y_2, \ldots, \delta y_N\}$ To read the data into *Mathematica* in this format we need to do some rearrangement:

```
xyey=ReadList["LRCm.dat",{{Number,Number},Number}]
xy=Part[xyey,All,1]
ey=Part[xyey,All,2]
```

The following line must be typed in exactly (including case) as shown:

```
nlm = NonlinearModelFit[xy, k1/Sqrt[1+k2^2*(k3/x-x/k3)^2],{{k1,.012},{k2,4.7},
      {k3,17600}},{x},Weights->1/ey^2,VarianceEstimatorFunction->(1 &)]
```

I hope the parts make some sense to you: notice particularly how the initial parameter guesses have been entered (e.g., $\{k1,.012\}$). The result should look at bit like:

$$\texttt{FittedModel}\left[\frac{.0118607}{\sqrt{1 + 22.7729 \ll 1 \gg^2}}\right]$$

The symbol `nlm` now contains all the information about the fit; proper requests can extract that information:

```
In[5]:= nlm["ParameterConfidenceIntervalTable"]
Out[5]=      Estimate    Standard Error    Confidence Interval
                                           0.0118153
        k1   0.0118607   0.000022857       0.011906


                                           4.74902
        k2   4.77209     0.0116262         4.79517


                                           17597.1
        k3   17605.      3.98307           17612.9

In[6]:= nlm["CovarianceMatrix"]
                  -10             -7               -7                   -7
Out[6]= {{5.2244 10    , 2.46067 10   , -7.06763 10  }, {2.46067 10   , 0.000135169,
                      -7
>     -0.00078076}, {-7.06763 10   , -0.00078076, 15.8649}}

In[7]:= nlm["ANOVATable"]
Out[7]=                       DF    SS             MS
                                           6
        Model                 3     1.88937 10     629790.

        Error                 97    363.926        3.75182

                                           6
        Uncorrected Total     100   1.88974 10

        Corrected Total       99    377421.
```

`ParameterConfidenceIntervalTable` gives us the fitted parameters with usual errors. As with `fit`, these estimates are the square root of the diagonal elements of the `CovarianceMatrix`. The quality of the fit is available in the `ANOVATable`: the `Error` row, the `SS` column is the χ^2 and the `MS` column is the reduced χ^2.

Mathematica needs some help to do plots with error bars. Additionally the error bars and the data need to be put together in exactly the right way:

```
Needs["ErrorBarPlots'"]
bar=Map[ErrorBar,ey]
xyEB=Transpose[{xy,bar}]
ErrorListPlot[xyEB]
Show[ErrorListPlot[xyEB], Plot[nlm[x], {x, 10000, 30000}], Frame->True]
```

Appendix D: Linux Commands

Appendix E: Error Formulae

Uncertainty Formulae

In the below equations the quantities $a, b, c, \ldots$ have uncertainties $\delta a, \delta b, \delta c, \ldots$ whereas the quantities $K, k_1, k_2, \ldots$ are "constants" (like π or 2) with zero error or quantities with so small error that they can be treated as error-free (like the mass of a proton: $m_p = (1.67262158 \pm .00000013) \times 10^{-27}$ kg or Avogadro's number $N_A = (6.02214199 \pm .0000047) \times 10^{23}$). This table reports the error in a calculated quantity R (which is assumed to be positive). Note that a quantity like δa is called an *absolute* error; whereas the quantity $\delta a/a$ is called the *relative* error (or, when multiplied by 100, the *percent* error). The odd Pythagorean-theorem-like addition (e.g., $\delta R = \sqrt{\delta a^2 + \delta b^2}$) is called "addition in quadrature". Thus the formula for the error in $R = K\, ab/cd$ could be stated as "the percent error in R is the sum of the percent errors in a, b, c and d added in quadrature".

Equation   Uncertainty

$$R = a + K \qquad\qquad \delta R = \delta a \tag{E.1}$$

$$R = K\, a \qquad\qquad \frac{\delta R}{R} = \frac{\delta a}{a} \quad \text{or} \quad \delta R = |K|\, \delta a \tag{E.2}$$

$$R = \frac{K}{a} \qquad\qquad \frac{\delta R}{R} = \frac{\delta a}{a} \quad \text{or} \quad \delta R = \frac{|K|\, \delta a}{a^2} \tag{E.3}$$

$$R = K\, a^{k_1} \qquad\qquad \frac{\delta R}{R} = \frac{|k_1|\, \delta a}{a} \quad \text{or} \quad \delta R = |k_1 K a^{k_1 - 1}|\, \delta a \tag{E.4}$$

$$R = a \pm b \qquad\qquad \delta R = \sqrt{\delta a^2 + \delta b^2} \tag{E.5}$$

$$R = k_1 a + k_2 b \qquad\qquad \delta R = \sqrt{(k_1 \delta a)^2 + (k_2 \delta b)^2} \tag{E.6}$$

$$R = K\, ab \qquad\qquad \frac{\delta R}{R} = \sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2} \tag{E.7}$$

$$R = K\, \frac{a}{b} \qquad\qquad \frac{\delta R}{R} = \sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2} \tag{E.8}$$

$$R = f(a) \qquad\qquad \delta R = |f'(a)|\, \delta a \tag{E.9}$$

$$R = K\, \frac{ab}{cd} \qquad\qquad \frac{\delta R}{R} = \sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \left(\frac{\delta c}{c}\right)^2 + \left(\frac{\delta d}{d}\right)^2} \tag{E.10}$$

$$R = K\, \frac{a^{k_1} b^{k_2}}{c^{k_3} d^{k_4}} \qquad\qquad \frac{\delta R}{R} = \sqrt{\left(\frac{k_1 \delta a}{a}\right)^2 + \left(\frac{k_2 \delta b}{b}\right)^2 + \left(\frac{k_3 \delta c}{c}\right)^2 + \left(\frac{k_4 \delta d}{d}\right)^2} \tag{E.11}$$

$$R = f(a, b, c, d) \qquad \delta R = \sqrt{\left(\frac{\partial f}{\partial a}\, \delta a\right)^2 + \left(\frac{\partial f}{\partial b}\, \delta b\right)^2 + \left(\frac{\partial f}{\partial c}\, \delta c\right)^2 + \left(\frac{\partial f}{\partial d}\, \delta d\right)^2} \tag{E.12}$$