Laboratory Report 2
Detecting Signals in the Presence of Noise & Astronomical Imaging

Kirsten Howley [1]
Group 1
September 25, 2001

1. BACKGROUND & MOTIVATION

In the detection and interpretation of data from faint sources, and sources with bright backgrounds, it is essential to be able to differentiate between the signal of the desired source and any accompanying noise. The unwanted noise might be a result of thermal fluctuations, constant background noise, and/or Poisson noise. The goal of this experiment is to develop techniques that will enable us to extract our desired signal from the rest of the chatter and measure its strength and accuracy relative to the noise.

In addition, this lab provides an introduction to the operation of the CCD (Charged Coupled Device) camera. This exposure (no pun intended) involves taking images, manipulating them, and designing a web page using HTML and CCD camera photos. An understanding of these processes will prove invaluable in future projects.

2. EXPERIMENT & METHOD

The first part of the experiment involved detecting pulses from a flashing light emitting diode (LED) located inside a photomultiplier tube. The pulse widths of the flashing LED were controlled by a Hewlett Packard 8112A Pulse Generator, which was hooked up to an interface box from the Linux PC, and depicted on a Kikusui COS5021 Oscilloscope. The photomultiplier tube also contained a constant LED that could be controlled manually from outside the box. Our photometer was composed of a H5773 photomultiplier module, photon statistics demo unit, pulse amplifier, discriminator, squawker, counter and CIO-CTR05 (counter card). A server program on the IBM PC (located in the node *pulsar.ugastro*) read out the counters.

In the second part of the experiment, images from the CCD camera were analyzed. A Photometerics Ltd. Peltier (thermoelectrically) cooled CCD camera was used and a NuBus card generated the sequences and read out the data (located in the node *quasar.ugastro*). It was necessary to cool the CCD chip (to -45°C) to prevent electrons from being liberated thermally. The LC200 cooling unit, which was used, circulated a water Ethylene Glycol mix.

---

[1] E-mail: kirsten@ugastro.berkeley.edu

3. LED DATA COLLECTION & ANALYSIS

The procedure for data collection involved manually setting the pulse generator to the desired pulse width and running a program in Unix to communicate to the PMT to collect counts. The pulse generator was set to emit flashes every other sample. The constant LED was turned down completely, and a two sets of 1000 samples were taken at a sample rate of 100 Hz, one with a flashing LED pulse width of 5 μs and the other with a flashing LED pulse width of 1 μs. Plotting the data demonstrates the difficulty of detecting the flashing LED at these very small pulse widths. It is nearly impossible to tell the LED is flashing at a 1 μs pulse width.

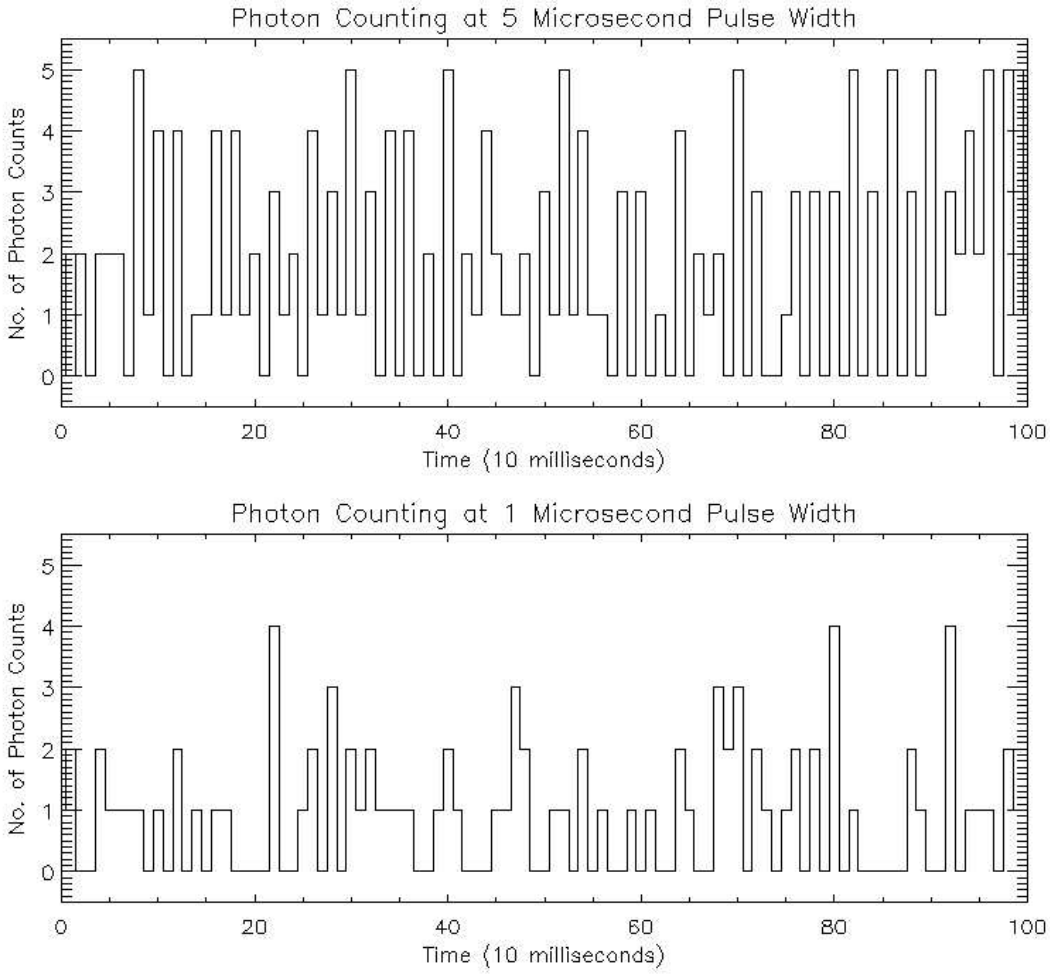

Fig. 1.— Photon counting for flashing LED pulse, thermal noise, and Poisson noise. Each sample period was 10 milliseconds (100 Hz) with a flashing LED (every other sample) of 5 μs and 1 μs pulse widths, respectively; a total of 100 samples were taken for each plot. These two plots demonstrate the difficulty of detecting pulses from the LED at very small pulse widths. At a pulse width of 1 μs it is impossible to tell the LED is flashing.

The difficulty in detecting the signal comes from noise generated by thermal fluctuations and Poisson statistics. The noise adds counts to our desired signal; since our signal is very faint (< 10 counts per sample), it becomes clouded by the accompanying noise.

To make detecting the signal even more difficult, the constant LED was added to the existing noise and signal. Since this was done manually, the squawker was used to determine when the constant LED was comparable to the flashing LED.Initially, after sampling the data, locating the source signal was done by viewing the results to determine when the LED was flashing and when it was not (the LED was set to flash every other sample). However, using this method alone would make it impossible to tell when the LED was flashing for very small pulse widths[2]. Fortunately, due to the fact that the LED was flashing synchronously, it was possible to measure the number of counts from each pulse by using a "Mark" command at the Unix command line. This automatically added 2^{14} counts to each sample that the LED was instructed to flash. It was then a simple task to write a program in IDL that recognized the high counts and extracted the corresponding data. This was extremely useful, later in the lab, for recognizing flashing signals generating counts that were only a tiny fraction of the noise.

With the constant LED added, 1000 samples of data were taken using the "Mark" command at a sample rate of 100 Hz and a flashing pulse width of 1 ms. The first second of sampling is shown in Figure 2. It is clear at this high pulse width that the LED is flashing. However, even though it is obvious that the count rate is alternating up and down due to the flashing LED, the signal belonging solely to the flash is not quite as clear. To obtain values corresponding only to the LED flash, the noise must be subtracted. An account of the background noise, LED flash & background noise, and our desired LED flash is depicted in Figure 3.

4. COMPARISON TO THEORY

4.1. Propagation of Errors

It is important to understand the effect uncertainties in our measurements have on the error of the result calculated from these measurements. It is simple to calculate the standard deviation of each set of our data. From Lab 1:

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

$$variance = s^2$$

The question arises, what is the application when multiple sets of data have been combined. If both sets of data are independent of one another (that is there is no covariance), and the measurements were either added or subtracted, then the variances add. A proof of this is shown:

[2]It was nearly impossible to notice the signal in Figure 1.

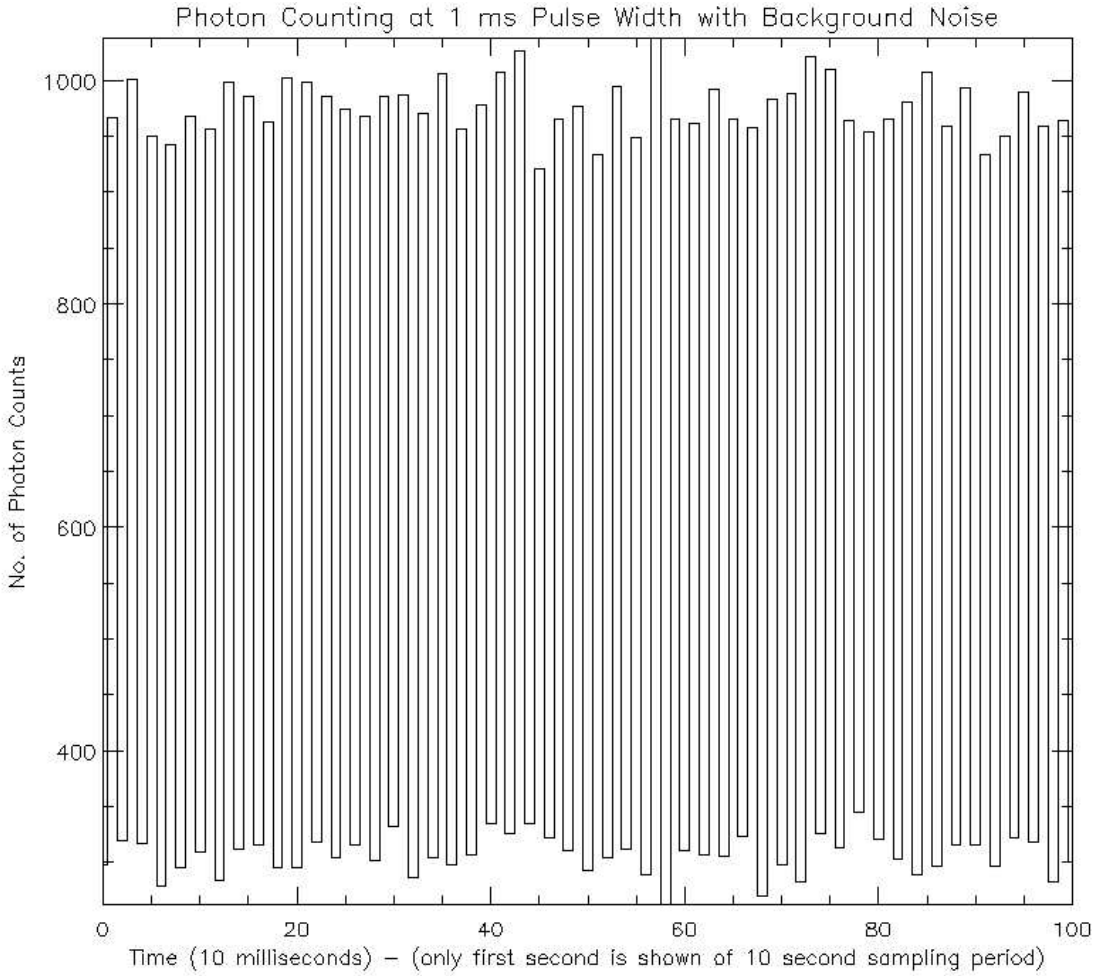

Fig. 2.— Photon counting for flashing LED pulse, constant LED, thermal noise, and Poisson noise. Each sample period was 10 milliseconds (100 Hz) with a flashing LED (every other sample) of 1 ms pulse width. A total of 1000 samples were taken; the first 1 second of sampling is shown (the first 100 samples).

For a given set of x & y values (where $u = x$ or y):

$$z = x + y$$

$$\mu = \langle u \rangle = \int_{-\infty}^{\infty} uP(u)\,du \tag{1}$$

$$\sigma^2 = \langle (u - \langle u \rangle)^2 \rangle = \int_{-\infty}^{\infty} (u - \langle u \rangle)^2 P(u)\,du \tag{2}$$

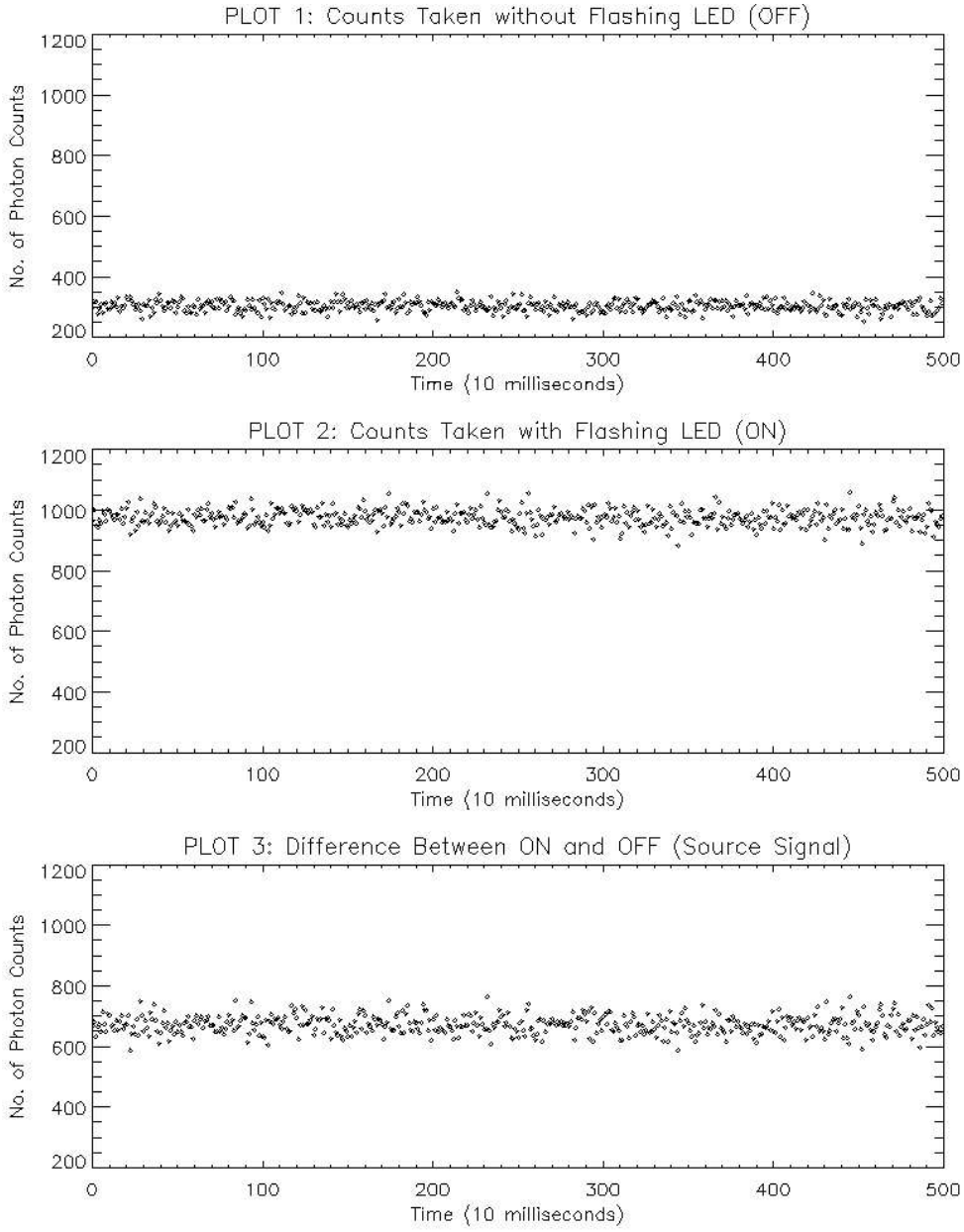

Fig. 3.— Data from Figure 2. Plot 1 shows the number of photon counts from every other sample when the flashing LED was off. Plot 2 shows the number of photon counts when the flashing LED was on. Plot 3 shows the strength of the flashing LED alone; it is the difference between Plot 1 and Plot 2.

.

If we assume x & y are independent then:

$$\langle z\rangle = \int_{-\infty}^{\infty} zP(z)\,dz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)P(x)P(y)\,dx\,dy$$

$$\langle z\rangle = \int_{-\infty}^{\infty} xP(x)\,dx \int_{-\infty}^{\infty} yP(y)\,dy + \int_{-\infty}^{\infty} P(x)\,dx \int_{-\infty}^{\infty} P(y)\,dy$$

From Equations 1 & 2 and the fact that the integral of a probability over all possibilities is just 1, we get:

$$\langle z\rangle = \langle x\rangle + \langle y\rangle$$

Calculating the uncertainty[3]:

$$\sigma_z^2 = \langle (z-\langle z\rangle)^2\rangle = \int_{-\infty}^{\infty}(z-\langle z\rangle)^2 P(z)\,dz$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\langle x\rangle + y - \langle y\rangle)^2 P(x)P(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\langle x\rangle)^2 P(x)P(y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(y-\langle y\rangle)^2 P(x)P(y)\,dx\,dy + 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\langle x\rangle)(y-\langle y\rangle)P(x)P(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty}(x-\langle x\rangle)^2 P(x)\,dx + \int_{-\infty}^{\infty}(y-\langle y\rangle)^2 P(y)\,dy$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2$$

This information can be used to calculate the uncertainty in our signal from the flashing LED. Calculating the variance of the data when the flashing LED was on and when the LED was off and comparing that to the variance of the difference of on and off yielded the following[4]:

$$s_{on}^2 + s_{off}^2 = 804 + 273 = 1077 \qquad vs. \qquad s_{on-off}^2 = 1096$$

Theory and estimate were accurate to one another to 2%! Both the error in the signal and the error in the background contributed to the total error.

4.2. Signal-to-Noise Ratio

In detecting faint signals, it is necessary to be able to determine if you are *really* detecting a signal, or, if instead, you are detecting fluctuations in background, thermal or Poisson noise. The signal can be estimated as the mean of the source signal extracted from the noise. The standard deviation of the signal (how much the mean of the signal fluctuates from the mean value) constitutes

[3]The term in the derivation $2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\langle x\rangle)(y-\langle y\rangle)P(x)P(y)\,dx\,dy$ is the covariance. Since we are assuming x and y are independent, we do not need the covariance for this particular calculation.

[4]Using the same data from Figures 1 & 2

the noise. The more data we accumulate, the closer we come to obtaining μ, the parent population mean, increasing our ability to detect fainter signals. It is also possible to estimate the Signal-to-Noise (SNR) ratio by summing each source signal and dividing by the square root of all counts. The reason this is also a good estimate of the SNR is because hidden in this equation is essentially the mean(signal) and the variance of the mean(noise)! Using Poisson statistics ($variance = mean$) we are able to show this correlation:

$$SNR = signal/noise = mean(on - off)/\sqrt{(vom)}$$
$$SNR = [mean(on - off) * samples]/\sqrt{(vom * samples^2)}$$
$$SNR = total(on - off)/\sqrt{(variance(on - off)/samples)} * (samples)$$
$$SNR = total(on - off)/\sqrt{variance(on - off)} * \sqrt{samples}$$
$$SNR = total(on - off)/\sqrt{variance(on) + variance(off)} * \sqrt{samples}$$
$$SNR = total(on - off)/\sqrt{[mean(on) + mean(off)] * samples}$$
$$SNR_{estimate} = total(on - off)/\sqrt{total(on + off)}$$

Using both methods to calculate SNR yielded the following results[5]:

$$SNR = 453 \qquad SNR_{estimate} = 420$$

A very good estimate indeed!

The experiment was ran for smaller pulse widths of 10 μs, 30 μs, 50 μs, 100 μs, 300 μs, and 500 μs; each time the sample rate was set at 100 Hz, and 1000 samples of data were taken. Figure 4 plots the SNR versus pulse width for each set. Based on the data, a theoretically prediction[6] of the Signal-to-Noise Ratio is plotted. Notice how the signal gets stronger as the pulse width increases. Curves for SNR will vary with the amount of constant LED. For the particular conditions of this experiment, the estimated smallest detectable signal was at about 300 nanoseconds. However, at this pulse width, the error in our measurement is so much greater than the obtained value that it would be difficult to say with great confidence that a signal had really been detected. If the errors in our measurements are independent of one another, then the error in our SNR can be computed by:[7]:

$$\delta q = \sqrt{(\frac{\partial q}{\partial x}\delta x)^2 + ... + (\frac{\partial q}{\partial z}\delta z)^2} \tag{3}$$

[5]Using the same data from Figures 1 & 2

[6]Using the estimate of the SNR derived earlier in this section

[7]Taylor, John R.

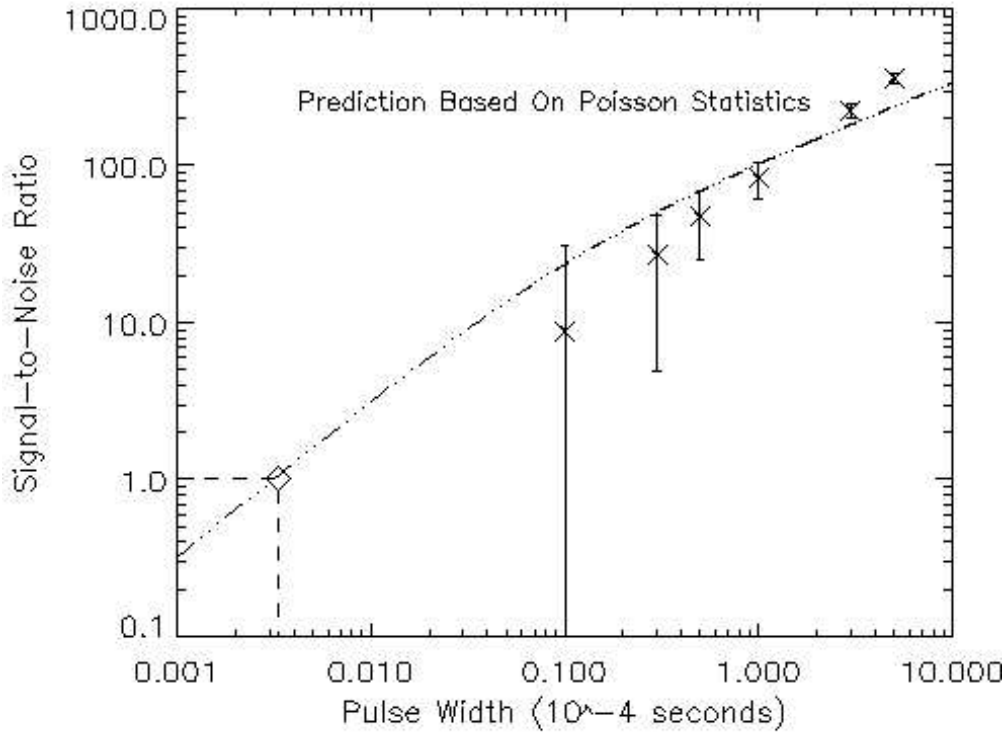

Fig. 4.— Pulse Width vs. Signal to Noise Ratio. For each point, 1000 samples were taken and a rate of 100 Hz for 6 different pulse widths.

Where q is a function of $(x, ..., z)$. It can be shown that for SNR, the uncertainty is[8]:

$$\delta q \approx \sqrt{N} \qquad , where\ N = \#\ of\ samples \tag{4}$$

Ultimately, the weakest signal detectable with the constant LED was a 400 ns pulse width. The signal was one half a percent of the constant LED. It had a mean of 1.5 counts, and a SNR of 2. A plot of the signal is shown in Figure 5. The negative values are a result of Poisson statistics and thermal fluctuations in the background noise; the uncertainties associated with these caused the noise to register higher in some samples than the noise and pulse in other samples. As mentioned, the error in this measurement was so high that it would be difficult to conclude with any certainty the true presence of a signal:

$$\delta q \approx \sqrt{N} \approx \sqrt{\frac{1000}{2}} \approx \pm 22$$

[8]Friedman, Andy & Huss, Lee

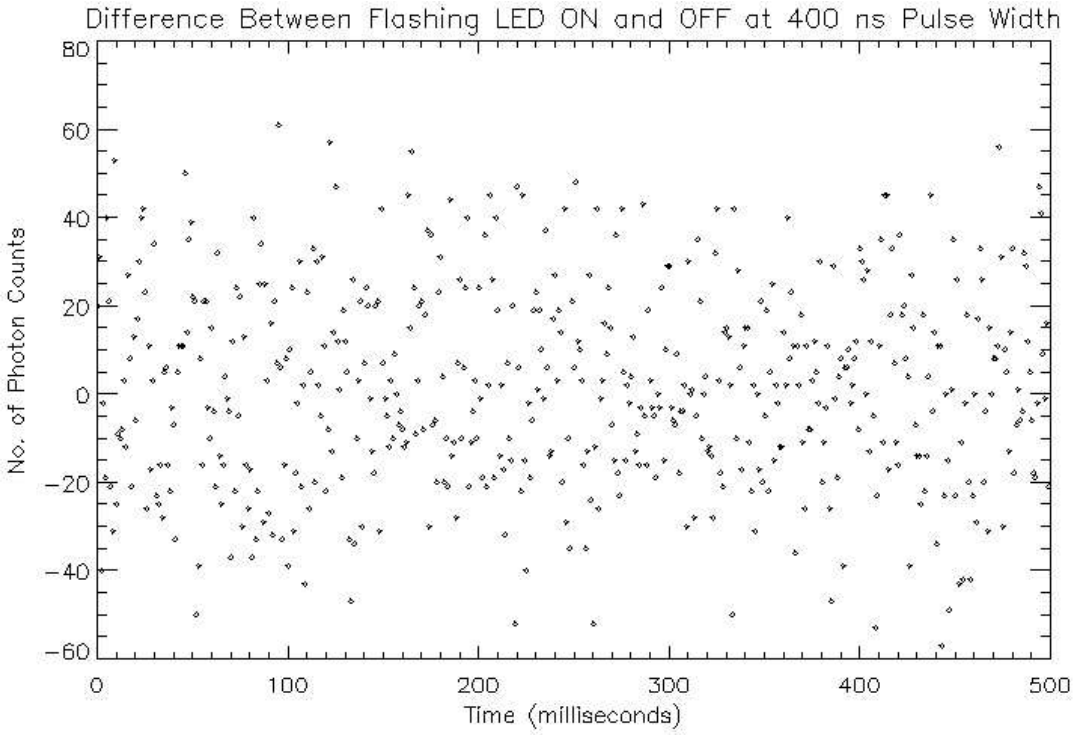

Fig. 5.— Difference between Flashing LED on and off at a 400 ns Pulse Width. A total of 1000 samples were taken at a rate of 100 Hz. This was the weakest signal detectable under sampling conditions. Note that the uncertainty in the signal is so high that it is difficult to conclude with any certainty that a signal had been detected.

5. Astronomical Imaging Data Collection and Analysis

The procedure for collecting images from the CCD camera involved manually focusing the lens and running a program in Unix to communicate to the CCD to take an image. The CCD obtained these images using the photoelectric effect. Electrons were liberated by incoming photons and then captured by nearby pixels with attractive voltages. The pixels then communicate to an analog digital converter which converts the data into an integer array after manipulating the data with some gain. It is then possible to generate an image whose intensity at different points is a function of these integers. The images can then be viewed and manipulated in IDL or XV.

Data was collected synchronously from the CCD at 1 ms, 10 ms, 100 ms, 250 ms, 300 ms, and 350 ms light exposure times. For each collection, two images were taken with the shutter open, and another image was taken with the shutter closed. In an attempt to acquire a fully illuminated image, the CCD was pointed directly at a white piece of paper placed beneath a fluorescent ceiling light.

The statistics applying to the individual images do not follow Poisson statistics very well. This

is more apparent when the mean and variance are calculated and compared[9]. Figure 6 plots the dark current distribution. Similarly, the count distribution for the illuminated images also fails to closely follow Poisson statistics. This would be expected since the images received by the CCD are fabricated and can take on any intensity distribution desired. Figure 7 shows a comparison of the histogram for a 350 ms illuminated exposure and the corresponding Gaussian distribution.

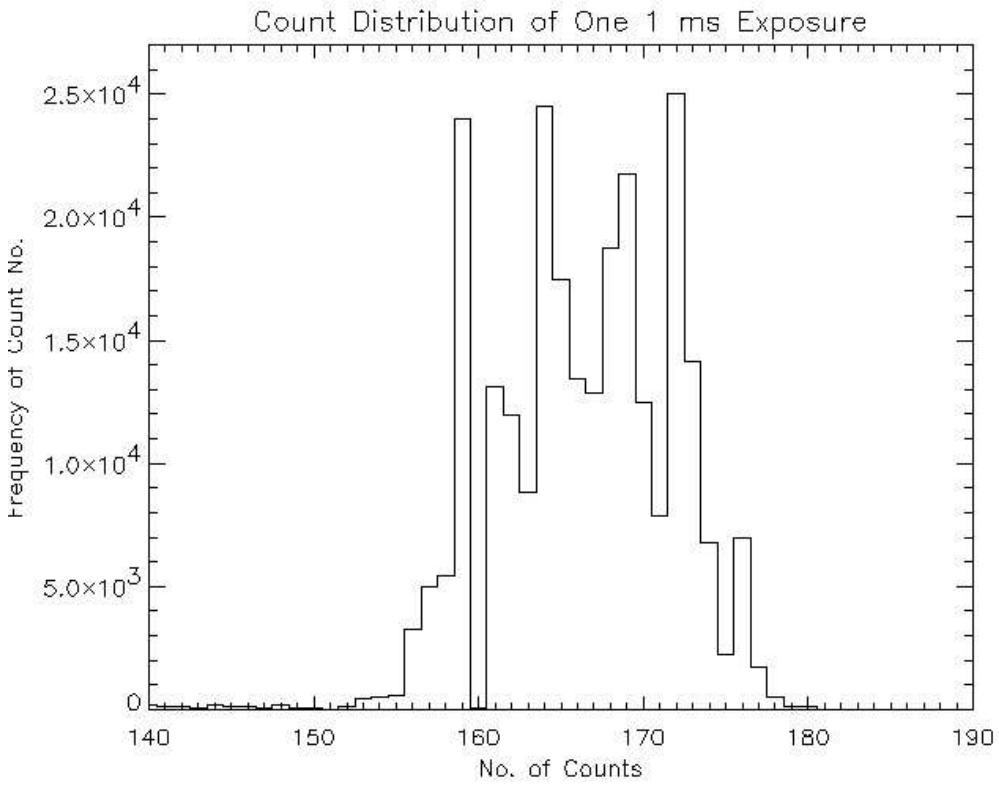

Fig. 6.— Count distribution of dark current taken with CCD camera with a 1 ms exposure time. The distribution has a mean of 166 counts and a variance of 44 counts^2. The distribution deviates from Poisson statistics, notice how it spikes sharply; all the counts are concentrated within a very small range.

Fortunately, there is a way to extract the statistical properties of photons from these images. Since two images taken seconds apart with identical exposure times should be nearly equivalent, subtracting these two images should provide a distribution that mirrors a Gaussian distribution. Figure 8 demonstrates the count distribution of two subtracted illuminated images taken with CCD camera with a 350 ms exposure time. Over plotted is the corresponding Gaussian distribution. Notice how the two match up almost perfectly!

[9]See Figure 6 caption for a comparison.

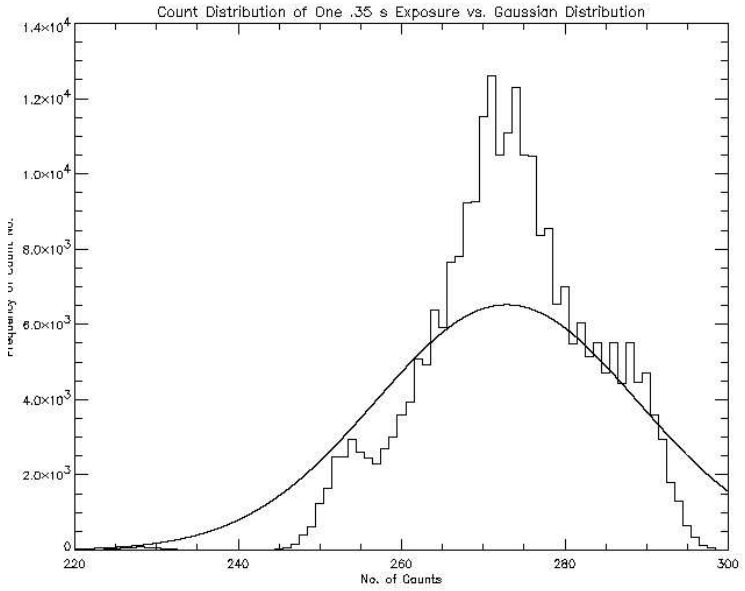

Fig. 7.— Count distribution of illuminated image taken with CCD camera with a 350 ms exposure time. Over plotted is the corresponding Gaussian distribution. Note how the illuminated image deviates from Poisson statistics.

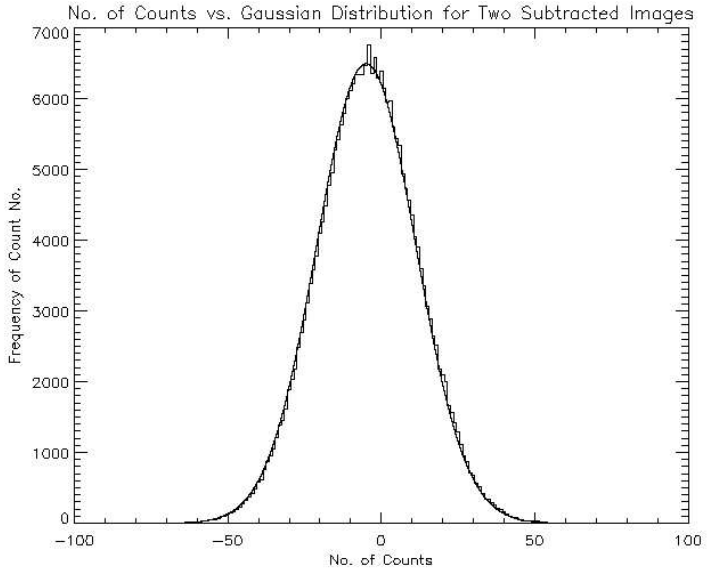

Fig. 8.— Count distribution of two subtracted illuminated images taken with CCD camera with a 350 ms exposure time. Over plotted is the corresponding Gaussian distribution. Notice how the two match up almost perfectly!

6. Comparison to Theory

Each pixel has some amount of readout noise associated with it. The question is how much? Using Poisson statistics we can determine how much readout noise is added to each signal. The count in each pixel is given by:

$$Count/pixel = (arbitray\ offset) + I_\gamma\ t_{exposure}\frac{gain}{capacitance} + I_{dark}\ t_{exposure}\frac{gain}{capacitance}$$

By subtracting images we can correct the signal for any arbitrary offset and dark counts. After correcting for these, we get,

$$signal = I_\gamma\, t_{exposure}\frac{gain}{capacitance} \tag{5}$$

Since each pixel represents a *mean* number of counts, by Poisson statistics, the variance of a count in a pixel is given by this mean. However, in the case of the CCD camera, there is also readout noise and the conversion factor of gain/capacitance. So the variance of a count in a pixel becomes[10]:

$$s^2 = (I_\gamma\ t_{exp} + \sigma^2_{read\,out\,noise})\frac{gain^2}{capaciatance^2}$$

this gives us,

$$s^2 = I_\gamma\ t_{exp}\frac{gain^2}{capaciatance^2} + \sigma^2_{read\,out\,noise}\frac{gain^2}{capaciatance^2}$$

from Equation 5,

$$s^2 = signal\frac{gain}{capaciatance} + \sigma^2_{read\,out\,noise}\frac{gain^2}{capaciatance^2} \tag{6}$$

Notice that this equation has the form of,

$$y_i = mx_i + b \tag{7}$$

Since we have values for x and y, the signal and the variance, we can calculate the best possible values for the slope and intercept. This is done with a process called *Least Squares Fitting*. Using the probability of obtaining any particular value for x_i , we can compute the probability of measuring the corresponding y_i. If we assume a Gaussian distribution:

$$P(y_i) \propto e^{-\frac{1}{2}\left(\frac{y_i - y}{\sigma_i}\right)^2} \propto e^{-\frac{1}{2}\left(\frac{y_i - y(x_i)}{\sigma_i}\right)^2}$$

$$P(y_1, y_2, ..., y_n) = \Pi P(y_i)$$

[10] Assuming $I_\gamma >> I_{dark}$

From Equation 7,

$$\Pi P(y_i) \propto e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - mx_i - b}{\sigma_i}\right)^2}$$

If we define,

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{y_i - mx_i - b}{\sigma_i}\right)^2$$

then we can assume that the most likely values of m and b from Equation 7 will occur when χ^2 is a minimum. If, in fact, Gaussian statistics do apply, then this **is** most likely when χ^2 is a minimum, because that means the probability is at a maximum. Taking the partial and setting it equal to zero to minimize χ^2 gives us,

$$\frac{\partial \chi^2}{\partial m} = 2\sum_{i=1}^{N}\left(\frac{y_i - mx_i}{\sigma_i^2}\right)(-x_i) = 0$$

$$\frac{\partial \chi^2}{\partial m} = 2\sum_{i=1}^{N}\left(\frac{y_i - mx_i}{\sigma_i^2}\right)(-1) = 0$$

expanding this out results in,

$$\sum_{i=1}^{N}\frac{x_i y_i}{\sigma_i^2} - m\sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} - b\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} = 0$$

$$\sum_{i=1}^{N}\frac{y_i}{\sigma_i^2} - m\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} - b\sum_{i=1}^{N}\frac{1}{\sigma_i^2} = 0$$

This is just two equations with two unknowns! Using the data that was collected synchronously from the CCD at 1 ms, 10 ms, 100 ms, 250 ms, 300 ms, and 350 ms exposure times to calculate the sums, a least square fitting line was derived via matrices. The results are displayed in Figure 9 along with the corresponding linear least square fitting equation.

In order to calculate the errors in our resulting slope and intercept, we must once again turn to propagation of errors. Since the values x_i and y_i are a result of different measurements under different circumstances, the spread in their particular values fails to define how accurately we have approximated a linear correlation. However, each measurement of y_i deviates from its counterpart $mx_i + b$ by some standard deviation, s_y. Thus, we can compute this deviation by[11]:

$$s_y = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}(y_i - mx_i - b)^2}$$

[11] Taylor, John R.

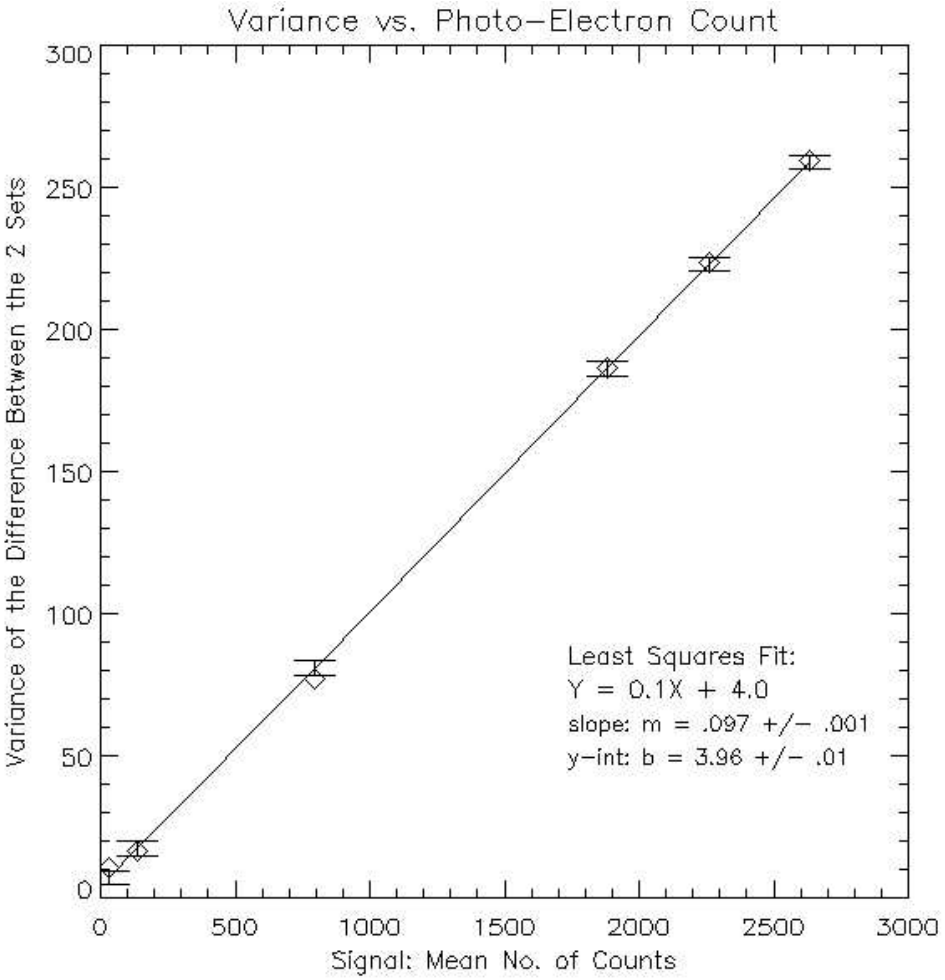

Fig. 9.— Variance vs. Photo-Electron count. Data acquired from CCD camera for various exposure times. The technique of least squares was used to fit a straight line to the points. Each point is the mean of two sets of data each contained 100 samples.

Using Equation 3, the error in the slope and intercept can be computed. The result of this computation is as follows[12]:

$$s_m = s_y \sqrt{\frac{1}{\sum_{i=1}^{N} x^2 - (\sum_{i=1} Nx)^2}}$$

$$s_b = s_y \sqrt{\frac{\sum_{i=1} Nx_2}{\sum_{i=1}^{N} x^2 - (\sum_{i=1}^{N} x)^2}}$$

[12]Taylor, John R.

The error for the data was calculated, and error bars were plotted in Figure 9. Each corresponding data point falls perfectly within range of the least squares fitting line, thus confirming the theoretically line.

This line has a very important interpretation. Equation 6 is the key to understanding these values. The slope corresponds to the $\frac{gain}{capacitance}$, and the intercept corresponds to $\sigma^2_{readoutnoise}\frac{gain^2}{capacitance^2}$. This may seem very vague at first, but, upon inspection, the meanings of these values can be determined. Since the intercept is our value when there is no signal, then it can be interpreted as the variance in the readout noise when there is no signal, or, by Poisson statistics, the mean amount of readout noise per pixel with there is no signal. The slope determines how this amount changes as the mean increases. As can be seen, the variance rises slowly with the mean number of counts. Thus we have discovered the answer to our initial question, *"How much readout noise is associated with each pixel?"*!

6.1. Personal Webpage

Using images obtained from the CCD camera, a personal webpage was designed using HTML code. This webpage can be viewed at www.UGAstro.Berkeley.EDU/~kirsten/.

7. CONCLUSIONS

The ability to understand how effectively our devices are able to detect photons and compute the amplitude of the errors associated with them is essential in analyzing astrophysical data. Poisson statistics and propagation of errors are tools that allow us to conclude, with some given certainty, whether or not a signal we are detecting is, in fact, more than just fluctuations in thermal noise, background noise or Poisson noise. That is, statistical analysis allows us to quantitatively express our confidence in a detected signal.

Equally important, is our ability to predict how a device should respond in a given environment. Since astronomical imaging is essential in astronomy, understanding the machinery and the limitations of its precision and accuracy will allow us to deduce what range of signals we are able to detect.

8. REFERENCES

Taylor, John R. *An Introduction to Error Analysis.* Sausalito: University Science Books, 1997: 75,187-188.

Friedman, Andy & Huss, Lee. *Laboratory 2*

9. ACKNOWLEDGEMENTS

I would like to thank Jim, Andy, Lee, Lindsey, and Eric for explaining a lot of the procedure and theory (as well as the associated problems with them) to me.Without them, I would still be on step #1 of this lab! I would also like to thank Nate for spending 3 hours with me explaining "variance of the mean".