Optical Lab, Astronomy 120
Lab 2, Astronomical Spectroscopy: Detecting light with a CCD

Sameen Yunus
sfsyunus@berkeley.edu
Anthony Khodanian & Pavan Muddukrishna

10/10/2017

Abstract

In this lab, we used a USB-2000 spectrometer to take spectra of various light sources such as fluorescent room lights, a neon lamp, an incandescent lamp and sunlight. The spectrometer collects data on a linear CCD (charge-coupled device) array of 2048 pixels. One of the goals of this lab was to find the wavelength calibration of the device by using well-known wavelengths of common light sources. We looked at the residual errors from a linear least squares fit and a second order polynomial fit to find that the mapping is better approximated by the higher order fit. We studied two sources of noise: the intrinsic noise in the detector that comes from bias offset and dark current, and the fluctuations in number of photons detected per pixel on the CCD. We expected the noise in number of photon arrivals to follow Poisson statistics and were able to calculate the gain and read noise from a linear fit to the mean and variance of the intensities.

Introduction

Spectroscopy is a useful astronomy tool that allows us to measure chemical compositions of sources. All chemicals absorb and emit electromagnetic radiation at particular allowed wavelengths. This happens when an electron is excited to a higher energy state and then emits a photon at the corresponding wavelength when it drops down to a lower energy level. We can compare known emission spectra of elements to relative fluxes of stellar absorption spectra to determine what their elemental compositions and abundances are. The USB-2000 spectrometer is sensitive to visible light wavelengths between 370nm and 700nm [1]. The spectrometer has a built-in pixel to wavelength calibration; however, there is a "total estimated uncertainty of within 10% for most Ocean Optics calibration light sources." [2] This motivates us to use the well-known peaks of Ne I and Hg I to calibrate the pixel to wavelength mapping ourselves and calculate the errors in our calibration.

[1] USB-2000 manual, page 1, https://drive.google.com/open?id=0B40Ynk22SiBpbXRnRlJRTHFtdzQ.
[2] Ocean Optics FAQs, https://oceanoptics.com/faq/accurate-calibration-light-sources-ocean-optics/

Data and Apparatus

The instrument used in this lab is an Ocean Optics USB-2000 spectrometer that takes input from an optical fiber. The light from the fiber passes through a filter that removes non-optical wavelengths. The device uses a concave mirror to direct a collimated beam through a diffraction grating that passes light of different wavelengths at different angles. Next, this light is focused by a second concave mirror, then passed through a cylindrical lens, and finally detected at a Sony ILX511 linear CCD array where the pixels respond to the incident wavelengths and intensities.
The CCD array uses the photoelectric effect to induce a current that is recorded in Analog-Digital Units (ADU). The current that is recorded depends on the integration time, which is the time that voltage is allowed to accumulate on a given pixel before the computer reads it. The raw data is then the ADU counts collected at each pixel in the 2048-pixel array. The saturation level is the maximum allowed voltage at each pixel which is set by a 12-bit number that the computer can store: $2^{12} - 1 = 4095$ ADU.

Data Analysis Methods

Bias offset and dark current

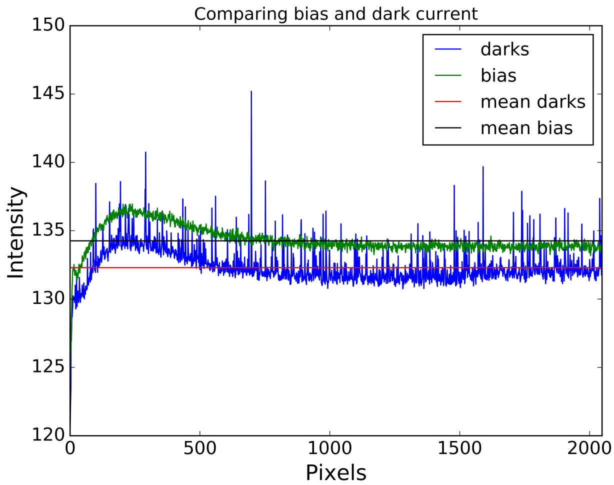


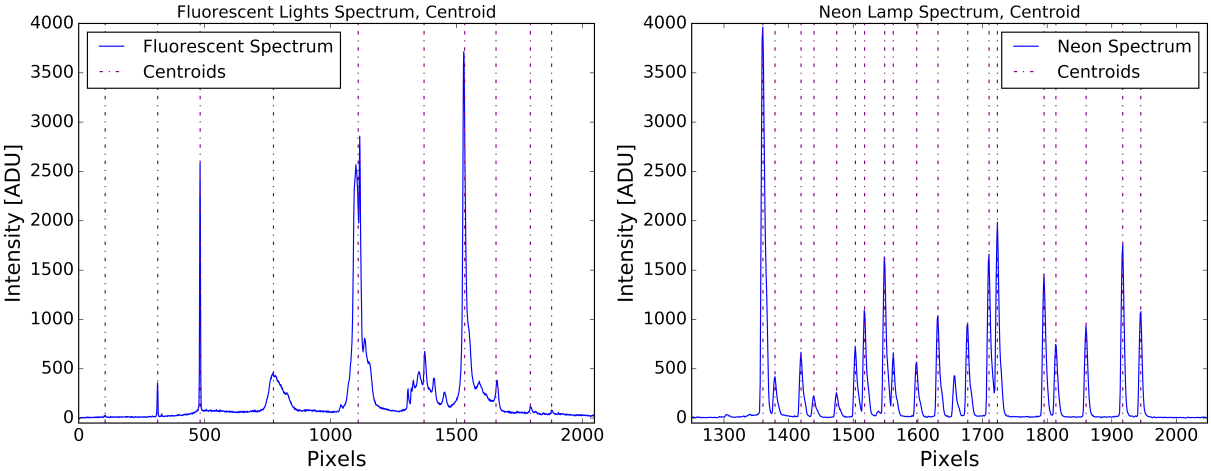
Figure 1: ***Green spectrum**: bias offset measured at an integration time,* t = 3 ms*; this is introduced by the device to avoid negative fluctuations in the ADU counts since the computer cannot output negative values.* ***Black line**: mean value of the bias offset:* 134.256 ADU. ***Blue spectrum**: dark current recorded for* t = 100 ms. ***Red line**: mean value of dark current:* 132.293 ADU.

The bias offset is a minimum positive voltage at each pixel that is introduced by the device because the detector cannot output negative values; this is set by the computer's 12-bit limit that can record values between 0 and 4095. In order to find this offset, we kept the cap on and set the integration time to the minimum value of 3 ms where the dark current is negligible because the CCD array doesn't have enough time to accumulate any charges. We took 100 spectra of the bias offset and then averaged the value of each pixel over those 100 files. This spectrum is then plotted in Figure 1 in the green line. My initial approach was to simply calculate the mean value across all the pixels, which was $\mu_b = 134.256$ ADU; however, when we calculate the standard deviation across all the pixels, we find $\sigma_b = 3.138$ ADU we find that the response varies across each pixel. Therefore it isn't sufficient to simply subtract the mean from all our spectra, we must make the correction for each pixel. However, we must also take into consideration the dark current that accumulates on the CCD array for our correction.
Dark current is the non-uniform accumulation of charges across the pixels due to the

residual electric current flowing in the photo-diode when there are no photons incident. Since all of our spectra are for short integration times of 100 ms, it is sufficient to find the average current accumulated across each pixel (this is shown in Figure 1 in the blue spectrum). We can see there are a few pixels that generate more dark current than the others; however, because we average over 100 spectra, we can simply subtract these values as they are. Seeing as how the response varies with the pixels, it is not sufficient to simply subtract the mean, $\mu_d = 132.293$ ADU from our data. We find the standard deviation $\sigma_d = 3.216$ ADU which is almost the same as σ_b. This tells us that the dark current is roughly constant with time; therefore, for our dark correction, we can simply subtract the average dark current accumulated at each pixel from our spectra. Note[3] that the dark current also includes the bias offset; correcting for the dark current is sufficient.

Centroids and centroid error

In order to find the pixel to wavelength mapping, we had to compare the pixel locations[4] of our spectra to well-known wavelengths of Hg I and Ne I. The best way to find the accurate pixel location was to use the centroids (Figure 2a and 2b) of the peaks because it returns the weighted mean of the peak. First, I found the pixel location corresponding to the maximum pixel value of each peak by comparing the counts of subsequent pixels. If the pixel value was higher than both the preceding and following pixels while also being higher than a base threshold (this was 200 ADU which removes the local maxima from noise) then it was selected as the maximum. Note: there is an inbuilt python function called *argrelextrema()* that finds local maxima or minima. I compared my method to this function and the peak values agreed.



(a) *Fluorescent room lights spectrum; we are interested only in the Hg I lines.*

(b) *Neon lamp spectrum and centroids; we are interested only in the prominent Ne I lines.*

Figure 2: *These spectra are dark subtracted. We later select the relevant centroids by hand for the pixel to wavelength mapping.*

[3] Our darks and bias were recorded on different days; this could explain why the bias has a higher average than darks.

[4] In this section, I use pixel location to mean the pixel number between 0 and 2047 and pixel value to mean the ADU counts for a given pixel location.

I then found the centroid values by taking a threshold around each peak (this was 10 pixels for fluorescent spectrum and 5 pixels for neon spectrum) and using the centroid Equation 01 for that region. Here $\langle x \rangle$ is the position of the centroid, x_i is the pixel location of each point used to calculate the centroid and I_i is the corresponding pixel value. This gives us the centroid values which we can compare to well-known wavelengths of Hg I and Ne I.

$$\langle x \rangle = \sum_i x_i I_i \Big/ \sum_i I_i \tag{01}$$

Centroid Error - In order to find the error in each of our calculated centroids, we can use Equation 02 which follows from error propagation:

$$\sigma^2_{\langle x \rangle} = \left(\langle x^2 \rangle - \langle x \rangle^2 \right) \Big/ \sum_i I_i \tag{02}$$

Where $\sigma^2_{\langle x \rangle}$ is the error in each centroid position, the numerator is the variance of the pixel locations and I_i is the pixel value. We calculated these errors for each centroid position for the fluorescent and neon spectra and then found the mean of these errors: $\bar{\sigma}^2_{\langle x \rangle, Ne} = 0.00204$ pixels and $\bar{\sigma}^2_{\langle x \rangle, Fl} = 0.0149$ pixels. Note the mean error in the fluorescent spectrum centroids is higher; this is likely because the centroids are spaced farther apart and we take the centroid over a broader area[5].

Pixel to wavelength calibration

Once we had our centroids, we used the numbers provided by the National Institute of Standards and Technology[6] to hand pick the ones that correspond to well-known wavelengths. Table 1 below summarizes the values of centroids and wavelengths.

Hg I λ [nm]	Hg I centroid	Hg I λ [nm]	Hg I centroid	Ne I λ [nm]	Ne I centroid
585.2487	1360.334	633.4427	1677.004	365.0153	104.267
594.4834	1419.351	638.2991	1710.1194	404.6563	313.934
614.3062	1548.528	640.2248	1723.062	435.8328	482.314
616.3593	1561.916	650.6528	1795.375	546.0735	1110.053
621.7281	1598.085	659.8952	1860.315		
626.6495	1631.113	667.8276	1916.893		

Table 1: *Known wavelengths of Hg I and Ne I corresponding to the centroids we calculated. We used twelve known data points for Hg I and four for Ne I.*

Then, by plotting the values in Table 1 with the known wavelengths as the independent variable and the calculated centroids as the dependent variable, we got Figure 3. We then fit a straight line and a second order polynomial to our data; this is shown by the red and

[5] I verified this by changing my code to find the centroid errors over different ranges between 5 and 30 pixels for the fluorescent spectrum and found that the error is higher for a broader range

[6] Hg I wavelengths, https://physics.nist.gov/PhysRefData/Handbook/Tables/mercurytable2.htm. Ne I wavelengths, https://physics.nist.gov/PhysRefData/Handbook/Tables/neontable2.htm

purple lines in Figure 3. I calculated the linear fit in two ways; I first used *np.polyfit* with degree one which returns a linear fit. The second method was the application of Equations 03 to our data wherein x is the wavelengths and y is the centroids. This tells us that the slope is the ratio of the covariance of x and y to the variance of x. This result is derived from error propagation[7]. Both methods returned the same slope and intercept.

$$m = \left(\overline{xy} - \bar{x}\bar{y}\right)/\left(\overline{x^2} - \bar{x}^2\right) \qquad c = \bar{y} - m\bar{x} \tag{03}$$

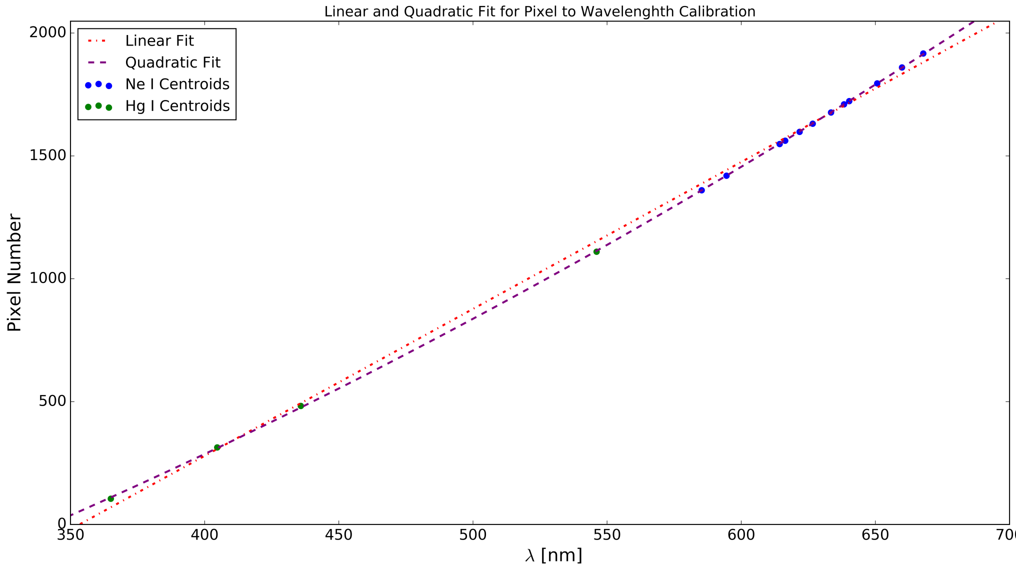


Figure 3: *Here we plot the calculated centroids against known values of wavelengths. The blue points are Ne I centroids and the green points are Hg I centroids. The line in red is the linear fit using the least squares method and the purple line is a quadratic fit using a numpy function. Note the quadratic fit is a better approximation to the data.*

We can also find the errors in our calculations of the slope and intercept using Equations 04 where σ_m^2 and σ_c^2 are the errors in the slope and intercept respectively, σ^2 is the variance, N is the number of samples, and x is the independent variable. Table 2 summarizes the linear and quadratic coefficients[8] and the errors in the linear fit.

$$\sigma_m^2 = \frac{\sigma^2}{N\sum x_i^2 - \sum x_i{}^2} \qquad \sigma_c^2 = \frac{\sigma^2 \sum x_i^2}{N\sum x_i^2 - \sum x_i{}^2} \tag{04}$$

Linear coefficients	Errors	Quadratic coefficients
m=5.9811567 [pixels/nm]	σ_m^2=2.589 × 10^{-6} [pixels^2/nm^2]	a_0=-1237 [pixels]
c=-2113.515 [pixels]	σ_c^2=5.592 [pixels^2]	a_1=2.4509 [pixels/nm
		a_2=3.395 × 10^{-3} [pixels/nm^2]

Table 2: *Fit parameters for linear and quadratic fit*

[7] page 4, https://drive.google.com/file/d/0B40Ynk22SiBpSU1DN2dPN3pzNXc/view

[8] linear: $y = mx + c$, quadratic: $y = a_2x^2 + a_1x + a_0$

In order to find the quadratic fit, I applied a numpy method, *np.polyfit(wavelengths, centroids, 2, full=True*) where it calculated a second order fit to wavelengths and centroids and also returns the sum of the residual errors. We find the average error for the quadratic fit is $\mu_{err} \approx 0.06$ and the average error for the linear fit is $\mu_{err} \approx 12$. Having found the linear and quadratic fits, we then had to find their accuracy. In order to do this, we found the residuals of both fits; this is the distance of each point from the line of best fit using the known wavelengths. These data are plotted in Figure 4;

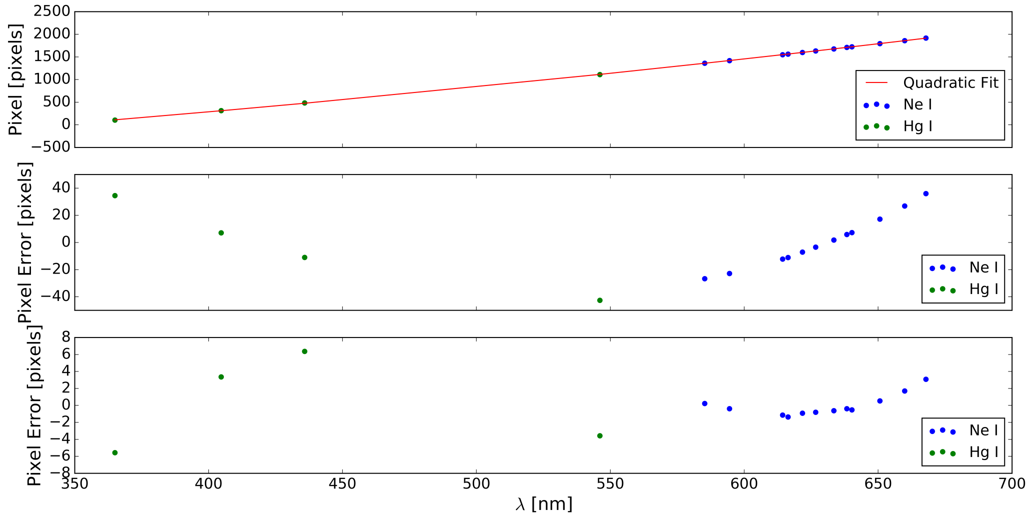


Figure 4: ***Top***: *Plot of wavelength versus pixels showing only the quadratic fit.* ***Middle***: *Linear residuals; note that they plot a quadratic which would suggest that a second order correction is called for.* ***Bottom***: *Quadratic residuals; note the variation in the error has decreased.*

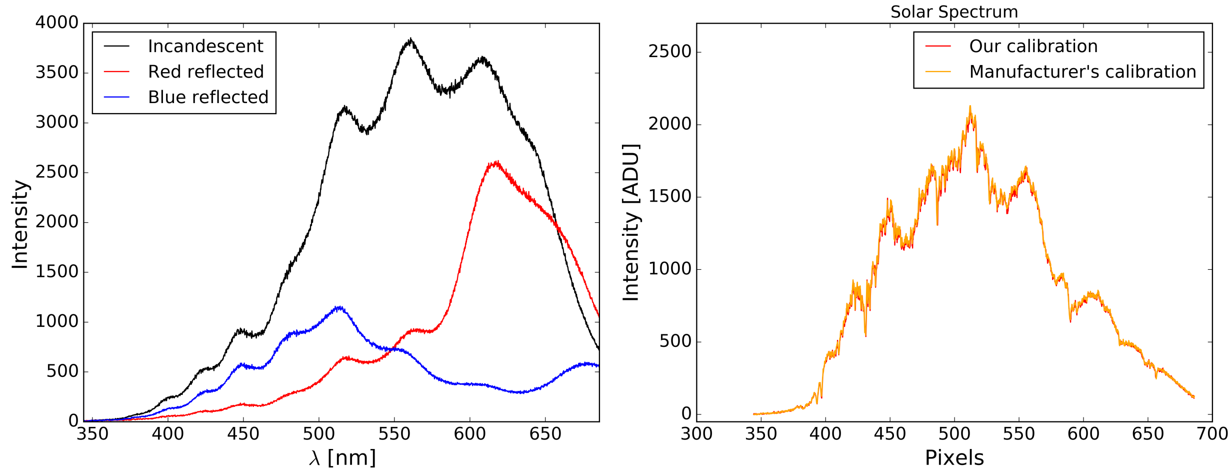
We can compare the standard deviation of the residuals to confirm that a higher order fit is necessary; $\sigma_{lin} = 3.5322$ nm while $\sigma_{quad} = 0.2451$ nm. The quadratic fit residuals have a deviation that is an order of magnitude lower than the linear fit. It is also worth noting that if we were to use the manufacturer's coefficients[9], given in the USB-2000 manual to find their residuals and calculate the standard deviation, we find $\sigma_{man} = 0.2673$ nm $\approx \sigma_{quad}$. Therefore, we are motivated to use a quadratic pixel to wavelength mapping: $\lambda = \frac{-a_1 + \sqrt{a_1^2 - 4(a_2)(a_0 - \text{pixel})}}{2a_2}$ where a_i are the coefficients in Table 2.

Common spectra

We can test our pixel to wavelength mapping by looking at some spectra of common sources. In Figure 5a we look at three spectra of incandescent light; the black spectrum is with the optical fiber directly pointed at the lamp while the red and blue spectra are the incandescent lamp reflecting off red and blue surfaces (textbooks) respectively. This plot shows us some fine scale variations in the pixel response that are common to all three

[9] $a_2 = -1.491 \times 10^{-5}, a_1 = 0.1974, a_0 = 344.3$

spectra; we study these in the next section. It is also worth noting that the array is more sensitive to redder wavelengths; this is shown in the higher counts for the red reflected spectrum compared to the blue reflected spectrum.
Figure 5b shows two spectra of the sun. The orange line is the manufacturer's calibration that was recorded using the wavelength setting of the SpectralSuite software. There is an underlying red spectrum that is a plot with the same instrument setup, but using our derived pixel to wavelength mapping instead. We can see some overlapping in the spectra, and the emission lines match up well. The solar spectrum is also a fascinating spectrum to study because we can recognize the features of a G-type star in its shape and spectral features.



(a) *Spectra of an incandescent desk lamp, and the same light reflected off a blue and red surface at the same brightness.*

(b) *Spectra of the sun; the orange line is the wavelength setting on the SpectralSuite software while the red line is our calibration.*

Figure 5: *Dark subtracted spectra of some interesting sources. Notice the small scale variations in pixels, we study these in the following section.*

Read noise and gain

The counts recorded by the computer in ADU are directly proportional to the voltage on each pixel with the constant bias offset discussed in a previous section. The number recorded by the computer is then:

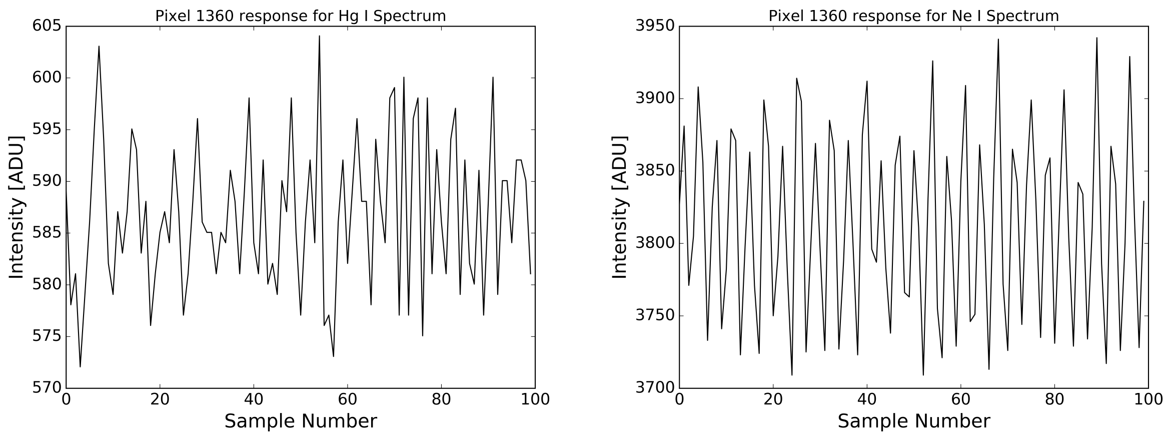
$$\mathrm{ADU} = \mathrm{gNe/C} + \mathrm{ADU}_0 \tag{05}$$

where N is the number of electrons excited in the CCD, the quantity ge/C is simply known as the gain factor G, and ADU_0 is the constant bias offset. Read noise is the noise generated by the device as the charge present on the pixels transfers to the ADC; since this noise is added to each pixel as it is read out, it is applied uniformly across the CCD. Gain is the conversion factor between electrons and the counts measured by the computer in electrons/ADU. Equation 05 implies that the ADU counts are a function of N and ADU_0. We can apply the law of error propagation and our understanding of Poisson statistics applied to the photoelectric effect from Lab 1 to get the following:

$$\sigma^2_{ADU} = G(ADU - ADU_0) + \sigma_0^2 \tag{06}$$

where we expect a linear relationship[10] between σ^2_{ADU} and $ADU - ADU_0$; the x value is the mean of the data and the y value is the variance[11]. We can then find the gain factor G, and the read noise σ_0^2 from the slope and intercept of this equation. σ^2_{ADU} is the variance in ADU counts calculated for each pixel using the average of 100 spectra; for this analysis we use the average of both fluorescent light and a neon lamp.

It is important to verify that our data is not dominated by external factors such as pixel illumination and dark current. We therefore plot the time series for pixel number 1360 that is used in our analysis in Figure 6a and Figure 6b. These plots show the counts measured at this pixel across all 100 spectra. This returns an interesting result: we note that the neon time series has a distinct periodicity and upon careful examination we can see that it also appears to be modulated. There are two periods to the time series, one is the 60Hz wall noise that occurs with a period of 1/120 Hz = 8.3ms, this occurs in both time series where each point sample is recorded after 100ms. There's an additional period corresponding to the flickering of the neon lamp due to gas being ionized in the tube so that the pixel value oscillates between some range of values. This phenomenon also occurs in the fluorescent lights but it has a longer period so it doesn't show up in our time series.



(a) *Pixel 1360 time series using spectral data for fluorescent light.*

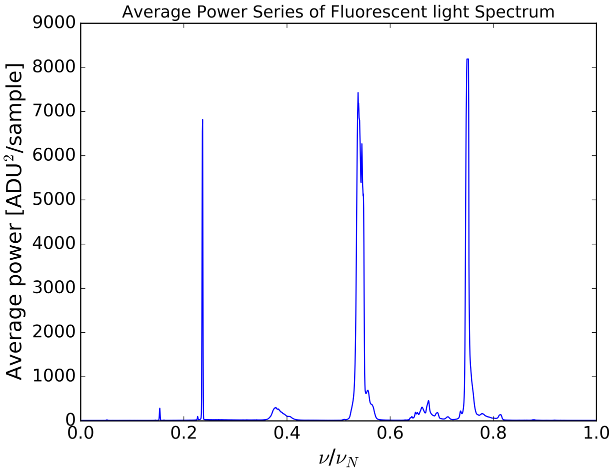
(b) *Pixel 1360 time series using spectral data for neon lamp.*

Figure 6: *Dark subtracted ADU response for pixel 1360 for two different spectral types, averaged over 100 samples. Note the periodicity in the neon time series.*

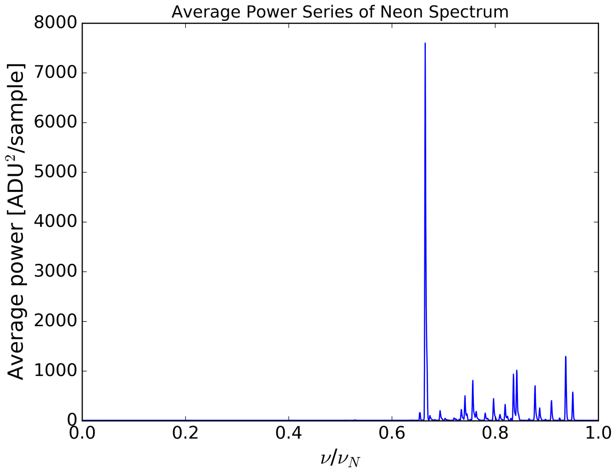
In order to confirm that our variance is not affected by external factors, we can plot the power spectrum of the fluorescent lights and neon lamp. As expected, the spectra in Figures 7a and 7b are flat except for the peaks at known elemental frequencies. We plot the spectra in units of ν/ν_f where ν_f is the Nyquist frequency; or half the sampling rate which sets the highest waveform frequency.

[10] The derivation of Equation 06 from Equation 05 is found on page 2 of the CCD Noise Handout, https://drive.google.com/file/d/0B40Ynk22SiBpcEVwU25IcHF0R1E/view

[11] $\mu = \sigma^2$ for Poisson statistics



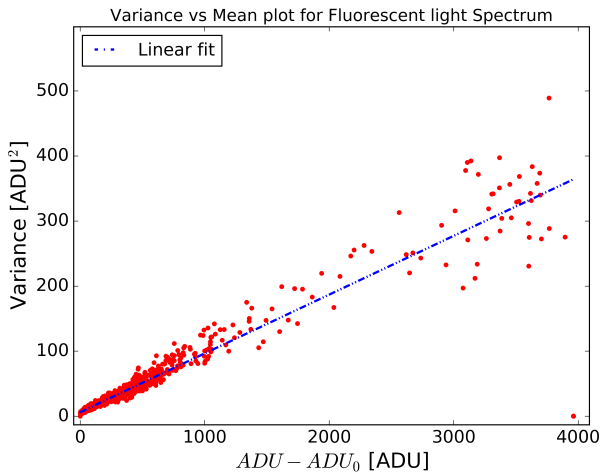
(a) *Fluorescent lights Fourier power spectrum.*



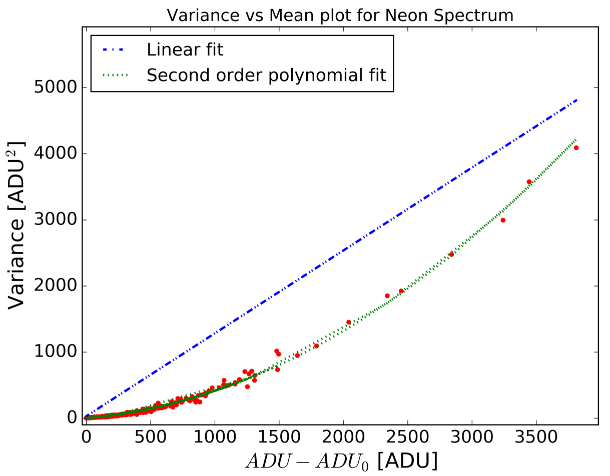
(b) *Neon lamp Fourier power spectrum.*

Figure 7: *Note that the spectra are near a constant value close to zero save the peaks at known wavelengths. This tells us the noise is mostly white noise.*

Having confirmed that any external factors are only due to white noise, we then plot the variance against the mean as in Equation 06. The Figures 8a and 8b show these distributions for both fluorescent light and the neon lamp. Using the method of linear least squares described in "**Pixel to wavelength calibration**"; we use Equations 03 and 04 to find the slope, intercept, and their respective errors. Table 3 summarizes the values of gain, read noise and their errors calculated for the linear fit of both types of spectra.



(a) *Fluorescent room lights variance versus mean. Note the linear fit looks like a good approximation.*



(b) *Neon lamp variance versus mean. The data is better approximated by a second order polynomial fit.*

Figure 8: *Dark subtracted variance versus mean plots for each pixel in the array averaged over 100 samples.*

We find from the plots in Figure 8a and 8b, and the fit parameters in Table 3 that the variance and mean of the fluorescent spectrum obey Poisson statistics since there's a linear

relationship between them: $\mu \propto \sigma^2$ with a scaling factor of gain: G and some constant offset which is the read noise: σ_0^2. If we consider only the fluorescent spectra; we can find the gain: $G = 0.0903$ ADU, and the read noise: $\sigma_0^2 = 6.361$ ADU^2.
However, the neon spectrum μ versus σ^2 plot seems to obey a quadratic relationship. At higher flux levels shown in Figure 8b we see that the data no longer obeys Poisson statistics and the signal is no longer proportional to the flux. This tells us that there is non-linear behavior at some stage of the device, either the CCD, the amplification or the ADC. The non-linearity might also be introduced as a result of the periodicity we observed in the time series of the neon spectra.

Fluorescent	Neon
G=0.0903 [ADU/e] $\sigma_G^2 = 4.13 \times 10^{-6}$ [ADU^2/e^2] $\sigma_0^2 = 6.361$ [ADU^2] $\sigma^2_{\sigma_0^2} = 1.6021$ [ADU^4]	G=1.254 [ADU/e] $\sigma_G^2 = 1.68 \times 10^{-4}$ [ADU^2/e^2] $\sigma_0^2 = 31.822$ [ADU^2] $\sigma^2_{\sigma_0^2} = 5.506$ [ADU^4]

Table 3: *Read noise and gain from linear least squares fit of fluorescent lights and neon lamp. Note the errors in the fluorescent spectrum linear fit are much smaller than the errors in the neon spectrum fit.*

Conclusion

We analyzed the sources of noise in the USB-2000 spectrometer by measuring the bias offset and dark current; we were able to remove these noise sources by averaging over 100 samples and subtracting them from all our spectra. We then calibrated the centroids of peak pixel numbers to the corresponding well-known wavelengths from NIST. We found our calibration of the CCD was accurate with an error of 0.0085 pixels. We then studied the read noise and gain of the device by assuming Poisson statistics and plotting the variance of the the ADU counts versus the mean ADU counts; these values should scale linearly if the data is truly Poisson distributed. This analysis revealed an interesting periodic behavior in the time series of the neon lamp spectrum in addition to evidence of 60Hz wall noise.