Internship Report: Study of the Linearity of the CCDs of the Vera C. Rubin Observatory

Jerónimo Calderón*
RECA Internship 2022, funded by the **LSST Corporation**
Astronomy B.Sc. Program, Universidad de Antioquia, Colombia
Advisors: **Craig Lage**†& **Andrés Alejandro Plazas Malagón**‡

Sep/30/2022

1 Introduction.

The Vera C. Rubin Observatory[1], scheduled to start operations on 2023, hosts the Simonyi Survey Telescope. With a primary mirror of 8.4 m in diameter, the telescope will conduct the **Legacy Survey of Space and Time** by observing the southern night sky for 10 years. The survey will focus on monitoring the transient phenomena of the deep sky and the Solar System, probing dark energy and dark matter, mapping the Milky Way, and more.

The Rubin Observatory is located in Cerro Pachón, about 100 km from La Serena, Chile, and it is now in the final stages of construction and setup. One of the most important parts of the telescope is its camera: the **LSSTCam**, a mosaic of 189 CCDs with a 3.2 gigapixel resolution designed and built by the SLAC National Accelerator Laboratory at Stanford University.

In order to process the massive amount of data that the LSST will generate (~ 20 TB per night) the Rubin Observatory staff have developed the **LSST Science Pipelines** (unofficialy called 'The Stack')[2]. In this internship we worked using the Stack to process several PTC data from the LSSTCam's CCDs. Our main objective for this project was to learn how to handle PTC data and try several possible setup parameters for the linearization algorithm developed in the Stack.

This document is a report of the work we developed during the internship with funding from the **LSSTC Enabling Science Award 2021-51**, and presents our main results. In sections 2.1 and 2.2 we show our exploration of the first sets of PTC data for detector 22, and inspect how it deviates from polynomial and exponential fits. In section 2.3 we present a tutorial on how to connect to the Rubin Science Platform and access the notebooks and other project files. Then in section 3 we present the approach taken to work with the linearizer algorithm and the results obtained by varying the parameters involved in the different fits that can be used.

This work was developed as one of the research projects for the **RECA Internship Program 2022**[3] funded by the LSST Corporation and under the advice and guidance of **Dr. Craig S. Lage** and **Dr. Andrés Plazas-Malagón**.

*jeronimo.calderong@udea.edu.co
†cslage@ucdavis.edu
‡aplazas@astro.princeton.edu
[1] About the Vera C. Rubin Observatory: `https://www.lsst.org/about`
[2] About the Stack: `https://pipelines.lsst.io/`
[3] RECA Internships: `https://www.astroreca.org/en/internship`

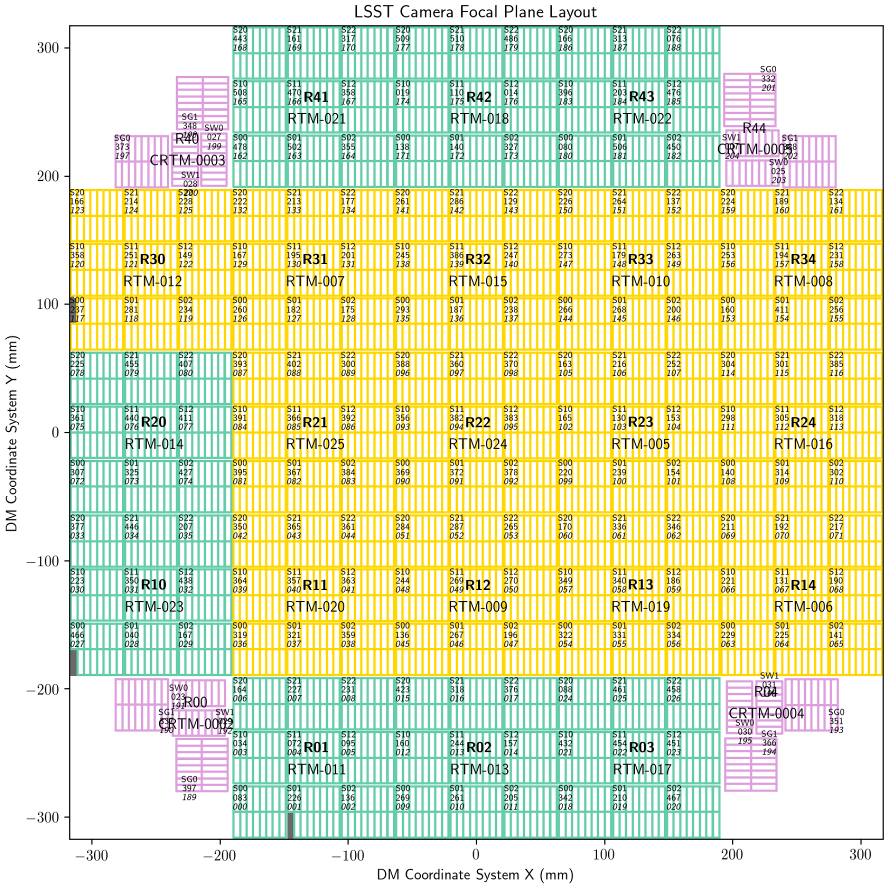


Figure 1: *Focal Plane map of all the CCDs on the Rubin Observatory LSST Camera. Yellow and green areas show CCDs manufactured by two different vendors. Our dataset for task 1 comes from a CCD on the third raft at the bottom right corner.*

2 Getting familiar with LSST Data through NCSA and the RSP workflow.

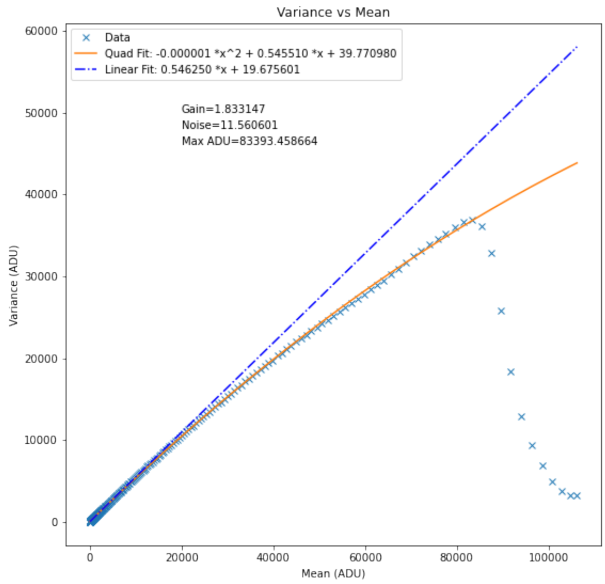
To get familiar with the **Photon Transfer Curve (PTC)** data, Dr. Lage provided some data sets with some tasks to start exploring the workflow and to help us get to know the RSP better. Here we show the statements of the tasks and the results obtained from our exploration.

2.1 Task 1: Exploring PTC Data.

Statement: "This is Photon Transfer Curve (PTC) data for one amplifier from the LSST Camera. The attached "FP_layout_DM.png" (1) is a map of the focal plane. The R refers to which of the 21 "rafts" (modules with 9 CCDS) it came from, and the S refers to which of the 9 CCDs it is. The C refers to which of the 16 amps it is. See if you can find which CCD this data came from. Note the yellow and green colors since the CCDs are from two different vendors. The file "PTC_R03_S11_C06_Det22.txt" is a text file with the measured mean and variance numbers from one amplifier, and the file "PTC_R03_S11_C06_Det22.png" is a plot of these. See if you can write a Jupyter notebook to read the .txt file and make a similar plot. Try to understand where the gain, noise, and A00 numbers come from and the difference between the red and green lines."

Plots and fits:

Figures 2 and 3 show the plots that resulted from solving task 1.



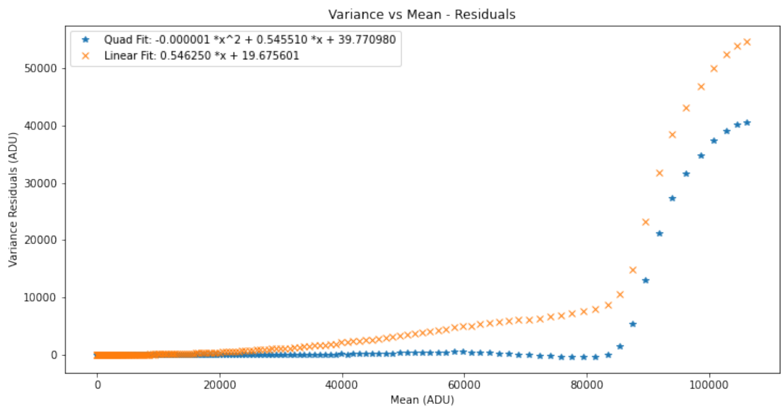


Figure 2: *Top: Quadratic and linear fit made with np.polyfit. Quadratic fit made using data points up to maximum variance. Linear fit made with points with* $\mu < 10000$ *ADU. Bottom: Residuals for both fits.*

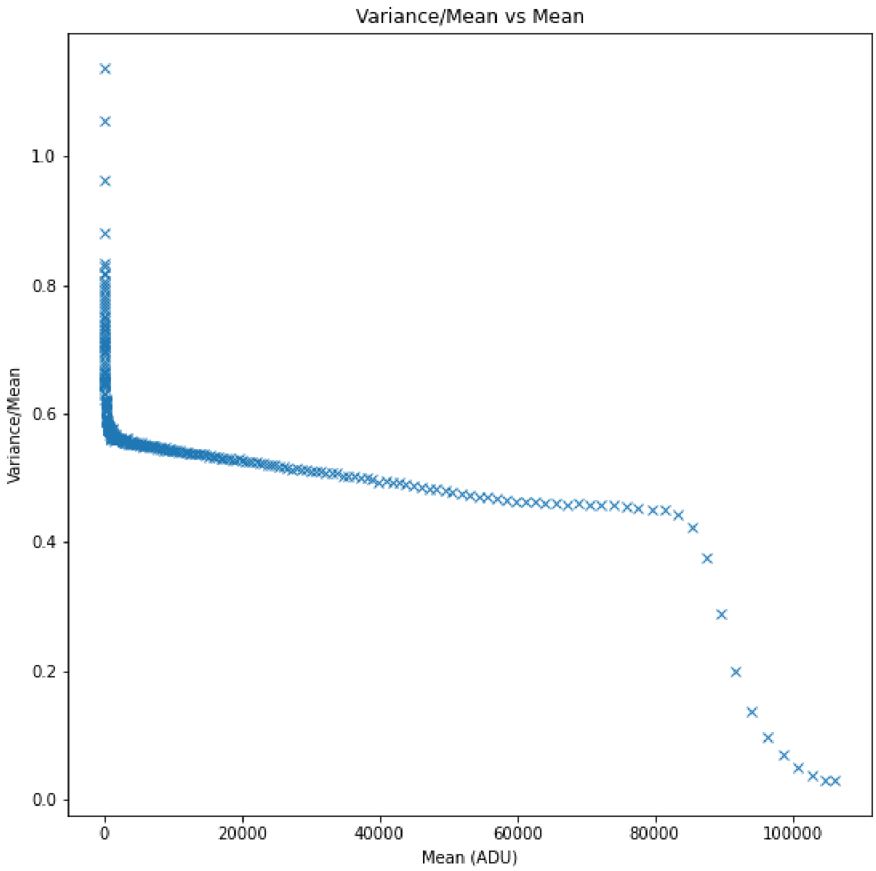


Figure 3: *Variance (calibrated by the mean) vs mean for the first data set.*

By plotting the PTC (variance vs mean, fig. 2) and its residuals from the fit model (see equation 16 from Astier et al. (2019)) wee see how the data presents higher absolute deviation more severely in the high flux regime. But if we plot the PTC normalized by the mean (fig. 3) we see how the low flux regime also presents high variances compared to the mean value obtained there. This is expected due to the effects of the reading noise of the detector which add randomness in the lectures for very low flux.

2.2 Task 2: Exploring Linearity with Diode Data.

Statement: "Here are the two files, one for each vendor. Each one has the IDs of the exposure pair, the mean, the variance, the exposure time, and the monitor diode. First, try plotting the monitor diode reading against the exposure time to see how linear this is. Second, try plotting the mean against the monitor diode to see how linear the detector response is. Since it will be quite linear, then try fitting the mean vs monitor diode with a straight line (linear fit) and plot the departure from the linear fit."
Plots and fits:
Figures 4 and 5 show the plots that resulted from solving task 2.

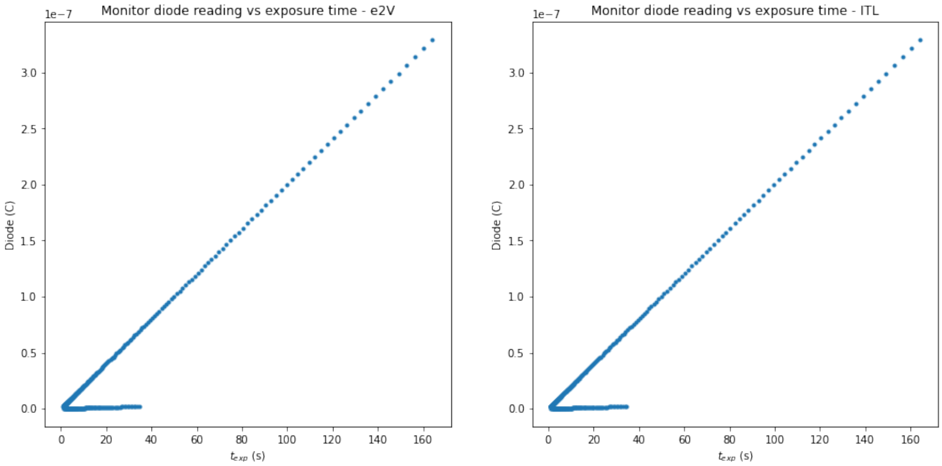


Figure 4: *Diode readings vs exposure time for two detectors, one E2V and one ITL. The expected linearity is noted.*

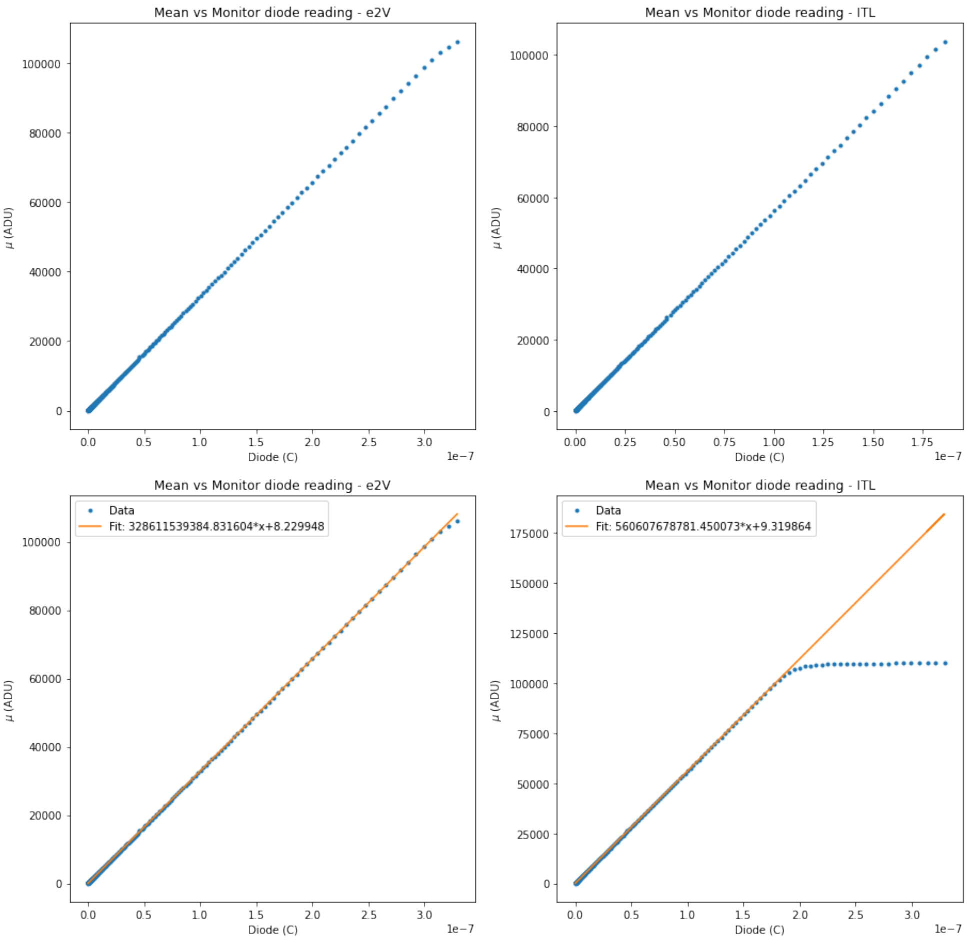


Figure 5: *Mean intensity values vs diode readings for both detectors without and with linear fits until linearity is lost.*

In this task we can see how **diodes are flux calibrators** to check the response of the detectors, since the monitor diodes measure flux by getting charged under a well-known linear behavior. In figures 4 and 5 we see how the linear regime ends when the detectors achieve fluxes of around 100000 ADUs, getting to the full well capacity of the pixels in the detector.

2.3 Task 3: Connection to RSP

Cisco AnyConnect VPN Service: NCSA requires connection through the Cisco AnyConnect VPN. Installation can be achieved easily by downloading the installer from the webpage: https://its.gmu.edu/knowledge-base/how-to-install-cisco-anyconnect-on-linux/ (Linux) or https://its.gmu.edu/knowledge-base/how-to-install-cisco-anyconnect-on-a-windows-computer/ (Windows).

After installation one must connect to the NCSA server as shown in figure 6. Make sure you are connecting to "NCSA SSLVPN" with the group "ncsa-vpn-default", with your assigned NCSA username and password. For second password you can type "push" to get sent a push notification through the Duo App or type in one of your Duo Passcodes (set a Duo account linked to your NCSA account if you have not done it already: https://wiki.ncsa.illinois.edu/display/cybersec/Duo+at+NCSA).

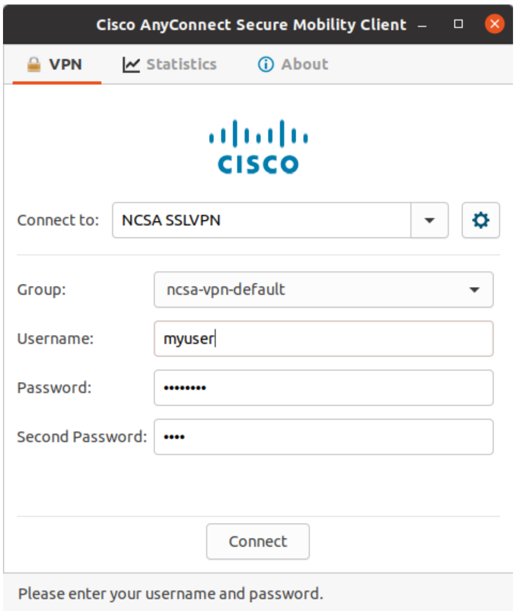


Figure 6: *Cisco Anyconnect GUI.*

After aproving connection via the Duo push notification (or Duo passcode) the status of the VPN should appear as "Connected".

RSP: To access the Rubin Science Platform follow the link: https://lsst-lsp-stable.ncsa.illinois.edu/, you should get the page shown in figure 7, there we only have access to the "Notebooks" section, which will let us work with Notebooks, Terminals and files that we have in our user folder and the main directories.

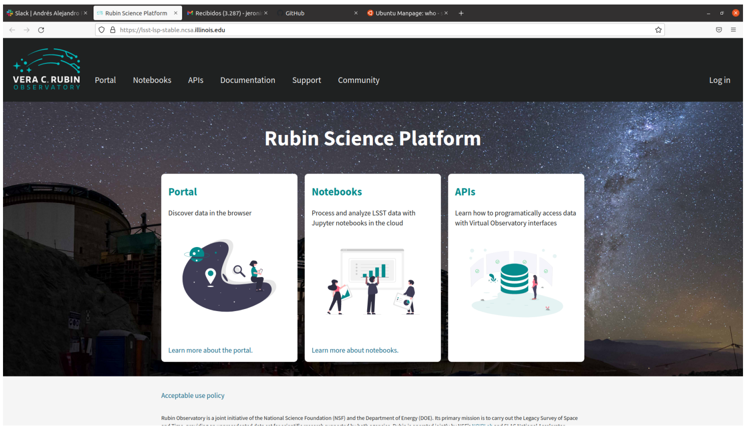


Figure 7: *Rubin Science Platform login page. Here we only have access to the "Notebooks" section.*

Via SSH:

1. Login to NCSA through one of the servers:
   `ssh myuser@141.142.181.16`
   `lsst-login01.ncsa.illinois.edu`
   `lsst-login02.ncsa.illinois.edu`
   `lsst-login03.ncsa.illinois.edu`

2. Login as dev:
   `ssh lsst-devl01.ncsa.illinois.edu`

3. Load LSST: `. /software/lsstsw/stack3/loadLSST.bash`

4. Setup: `setup lsst_distrib`

5. Create yaml file for authentication, this is done with a text editor like nano or vi in the folder: `/home/USER/.lsst/db-auth.yaml`
   ```
   - url: "postgresql://lsst-pg-prod1.ncsa.illinois.edu:5432/lsstdb1"
   username: "myuser"
   password: "mysecretpasswd"
   ```
   Check that the words "username" and "password" are aligned with the "u" in "url". Keep the dash in the beggining.

6. Data example:
   ```
   pipetask run -j 32 -d "detector IN (0) AND instrument='LATISS' AND exposure IN
   (2022050500251..2022050500290) AND exposure.observation_type='flat'" -b /repo/main -c
   isr:doFlat=False -i LATISS/raw/all,LATISS/calib -o u/jcalderong/test.1 -p
   $CP_PIPE_DIR/pipelines/Latiss/cpPtc.yaml --register-dataset-types
   ```
   This runs for a couple of minutes and creates the file 'test.1'.

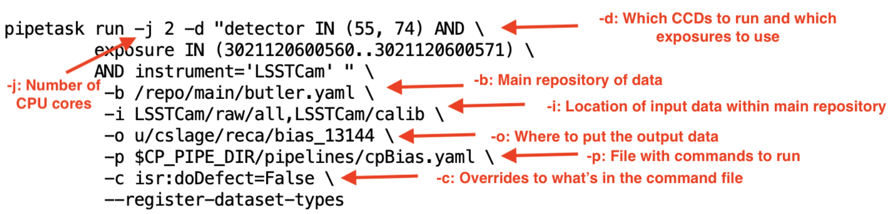


Figure 8: *Explanation of the pipetask commands.*

3 Running the Stack's Linearizer.

We worked with the Stack's linearization algorithm as shown in the notebook attached in the link `https://github.com/jerocalderong/LinearityRubinObservatoryCCDs`. Here we see the obtained results.

In figure 9 we show a comparison for detector 74 **without and with linearization** for the default settings (10 uniformly distributed spline knots). We see how the deviation from linearity gets significantly reduced, but we want to explore different values that may improve the PTC residual curve.

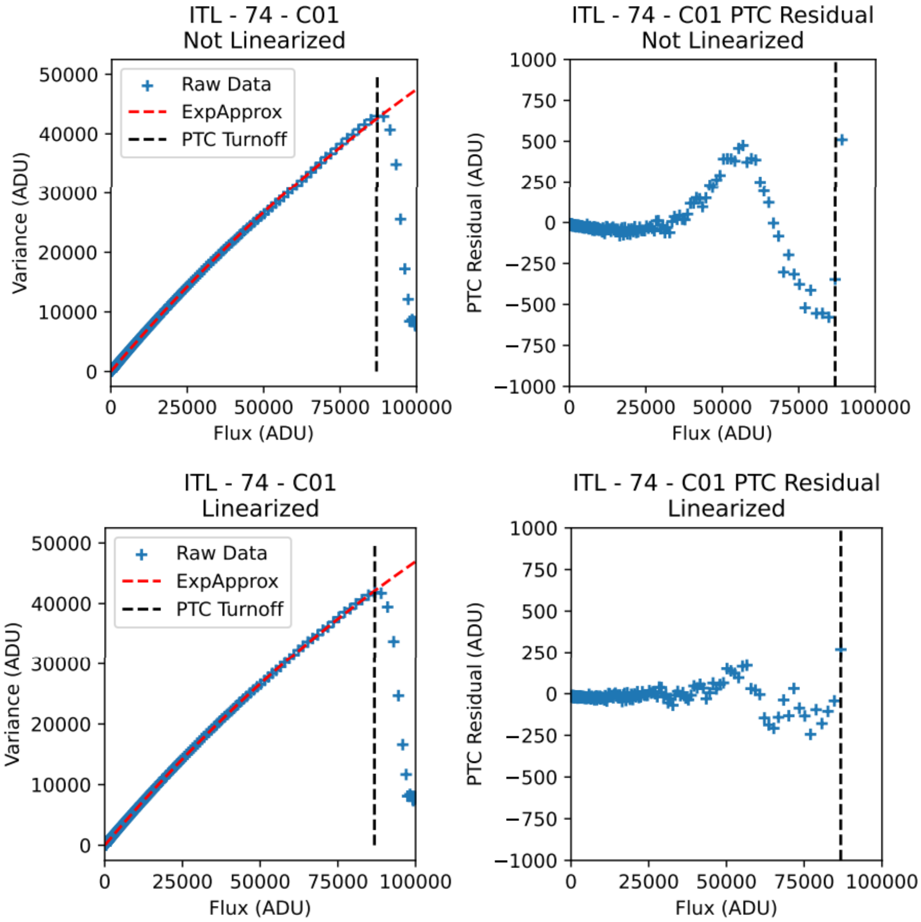


Figure 9: *Top: Non-linearized data. Bottom: Linearized data with 10 Spline knots (default setting).*

Figure 10 shows the **linearization plots obtained with 11 Spline knots** for the same detector. Here we see in the top row the same plots we generated for fig. 9 but also in the bottom row we plot the knot distribution and the fit residual for the two iterations of the fit that the linearizer runs.

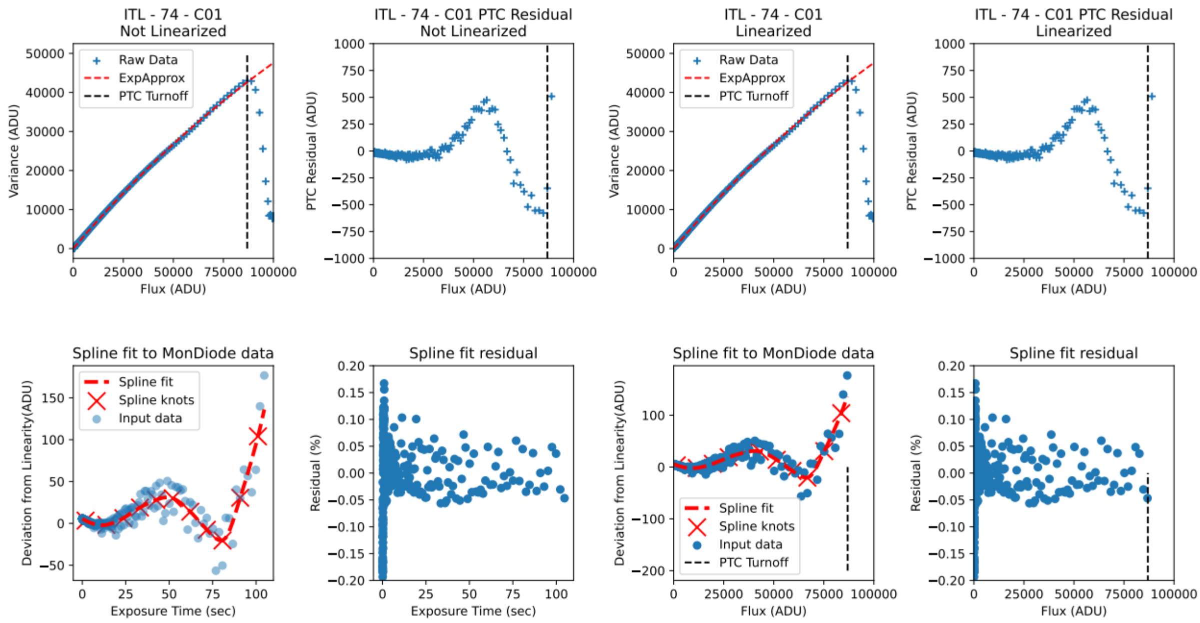


Figure 10: *Linearizer ran using 11 Spline knots. From top to bottom and from left to right we see the plots for: Non-linearized PTC, non-linearized PTC residuals from fit, linearized PTC, linearized PTC residuals from fit, spline knots used with the obtained deviation on the first fit, residual from first spline fit, spline knots used with the obtained deviation on the second fit, and residual from second spline fit.*

In figure 11 we show a **comparison of the linearization with different number of Spline knots**. We notice how the same detector gets different PTC residuals from fitting splines of different number of knots. By inspecting the scattered points we see that 12 knots seems to give the best result, but we analyze this in further detail in section 3.2.

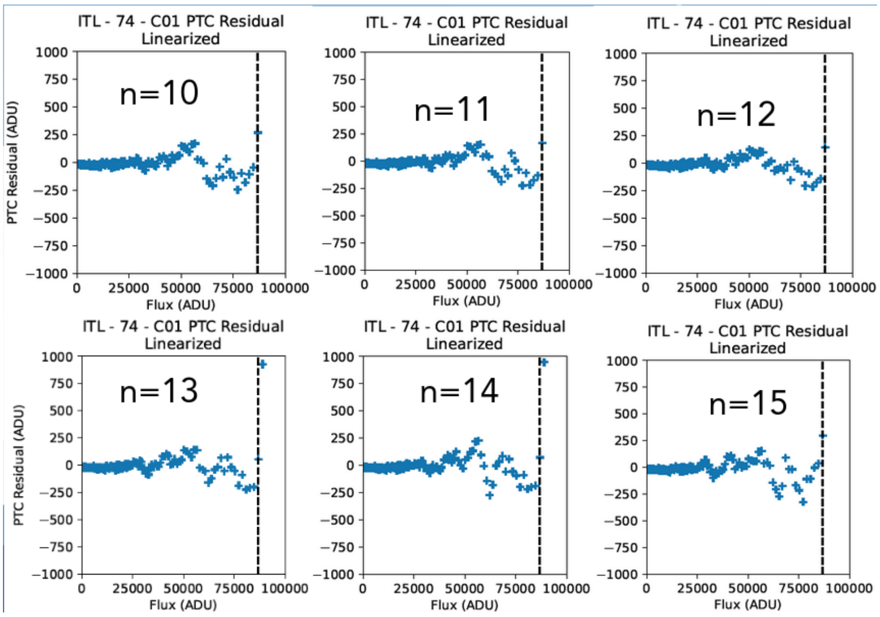


Figure 11: *Deviation from linearity using different amount of knots.*

3.1 Changing the knot distribution

A final test we ran was trying to fit the PTC using a non-uniform distribution of the Spline knots. With the help of Dr. Lage we ran the linearizer so that it used **10 evenly-spaced knots and around the 50000 ADU bump it added 5 extra knots between the others**. Figure 12 shows the results obtained and figures 13 and 14 show the deviation from linearity for all 16 amplifiers for 4 different detectors.

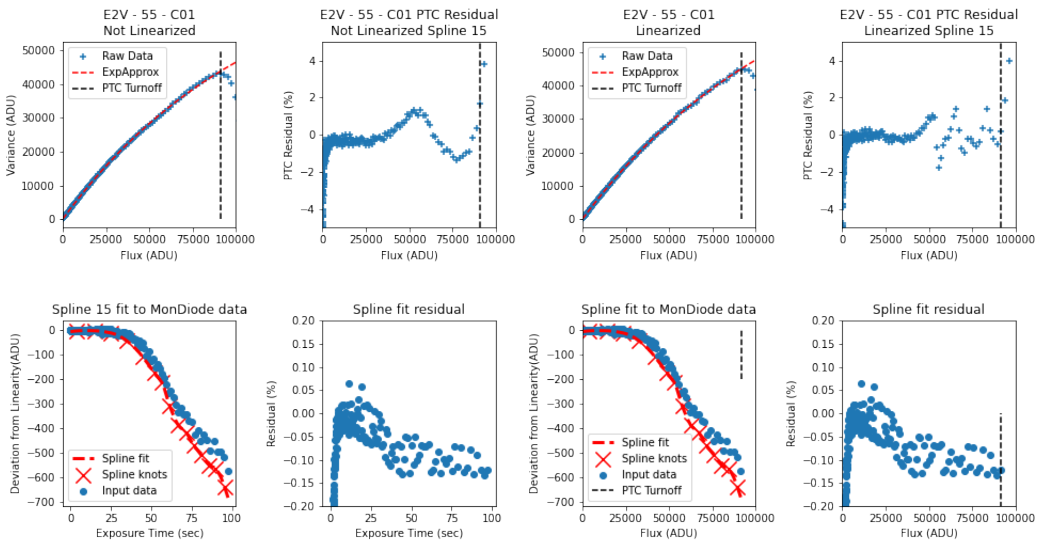


Figure 12: *Linearized PTC data for 10 equally spaced knots and 5 extra knots around the 50000 ADU bump.*

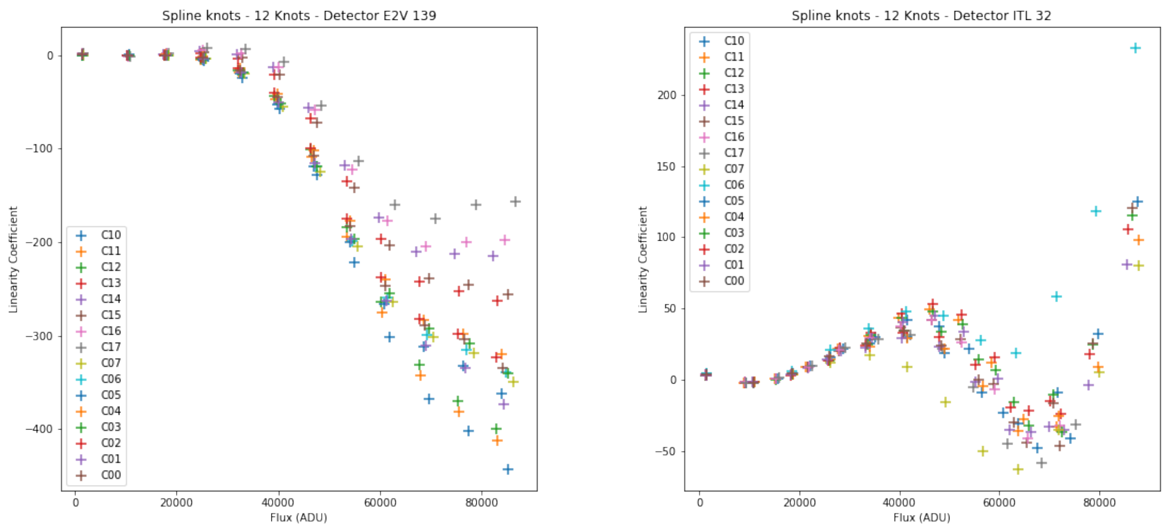


Figure 13: *Deviation from linearity for the 16 amplifiers of detectors E2V 139 and ITL 32.*

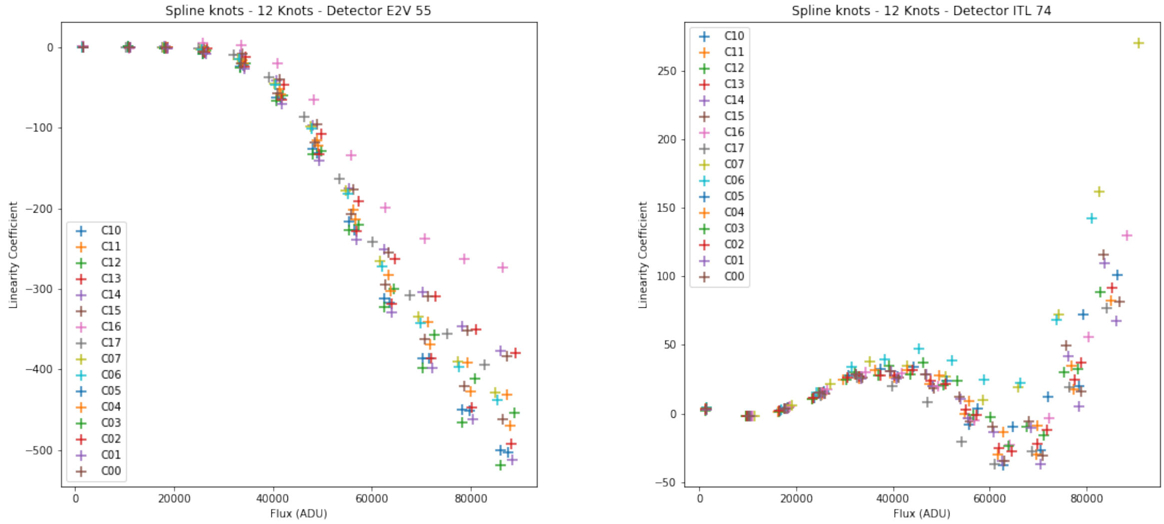


Figure 14: *Deviation from linearity for the 16 amplifiers of detectors E2V 55 and ITL 74.*

3.2 Results and Conclusions

If we determine the average residual of the deviation from linearity in each case (no linearization, several spline knots with even spacing, and not evenly spaced knots) we can make plots like the ones shown in figures 15 and 16 in which we see how 12 Spline knots seems to be the right configuration for best linearization in both E2V and ITL detectors. Although we see that the difference between the 12-knot case and the others is not very significant, and mainly we can conclude that just **applying the linearizer with any of this configurations is much better than not linearizing at all**.

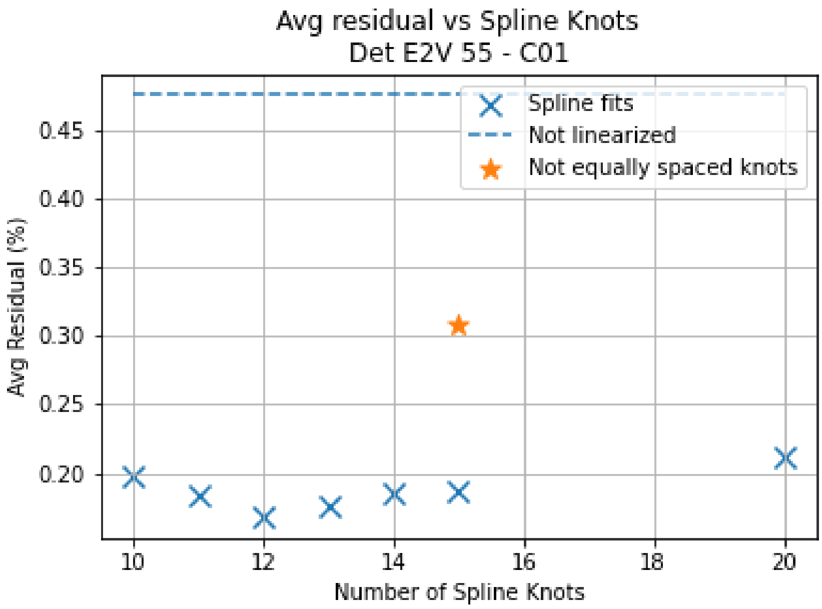


Figure 15: *Average deviation for different fit types for detector E2V 55.*

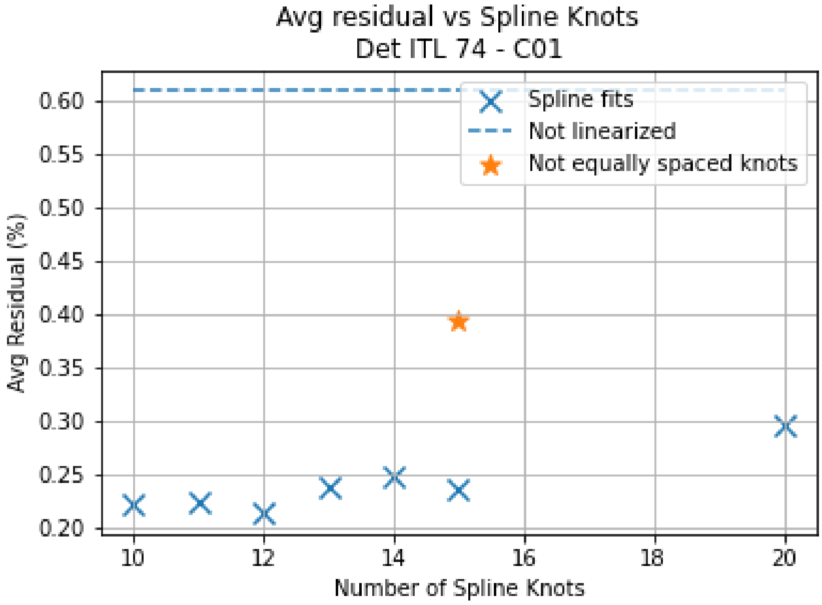


Figure 16: *Average deviation for different fit types for detector ITL 74.*

We also note that the way we ran the linearizer with not-evenly-spaced knots resulted in overfitting and seems to have affected the quality of the linearization, we suggest this approach to be tested using maybe **more sofisticated algorithms that can optimize the position of the knots** to get below the 0.01% deviation mark. One possibility would be to explore algorithms like LOESS[4] that use weight functions to smooth data with noise and get better fits.

During the presentation of this results to the DESC-SAWG meeting of the Rubin Observatory scientific team, Dr. Dan Weatherill pointed out an open problem about the fit residual plots (see fig. 10) where indifferent laboratories, using different detectors **a bimodal behavior has been observed consistently** and is still not very well understood. This could be another interesting topic or future research in linearity for the LSSTCam.

References

Astier, P., Antilogus, P., Juramy, C., Le Breton, R., Le Guillou, L. & Sepulveda, E. (2019), 'The shape of the photon transfer curve of ccd sensors', *Astronomy & Astrophysics* **629**, A36.

[4]LOESS `https://towardsdatascience.com/loess-373d43b03564`