Calibrating electron multiplication gain for the Roman Coronagraph detector

Peter Williams, David Nemati, Bijan Nemati, Guillermo Gonzalez
Tellus1 Scientific
2025-JAN-24

1 Introduction

Accurate knowledge of the electron multiplication (EM) gain is necessary for obtaining a correct estimate of the EMCCD generated images processed as 'analog' (proportional) frames. In these cases, the collected images are divided by the known EM gain to obtain the image, in electrons, as generated on the sensor. To a lesser extent, photon-counted frames also depend on EM gain. This is because the thresholding efficiency is a function of the applied EM gain, so correcting 'back' requires calibration of EM gain. The task is to be able to report the actual EM gain in a frame to within 10% of its true value. Due to the complexities of both the hardware and calibration process, accomplishment of this goal is nontrivial. We will walk through all the necessary considerations and the details of the solution that we will implement.

2 Gain multiplication

2.1 Physics

EXCAM and LOCAM both use electron multiplying charge-coupled devices (EMCCDs) to detect signals. Electron multiplication occurs in a region of the detector called the gain register. This specific area has a high voltage applied across it, creating an electric field that accelerates the incoming electrons. The increased energy of each electron is enough to cause it to knock other electrons free from their medium, resulting in a cascade effect where the original number of input electrons is multiplied by some factor >1 to designate the final number of output electrons. The analog voltage is specified by a digital value, as in all modern controls, and thus a digital-analog converter (DAC) is needed. Thus, the input parameter in our EM gain calibration is called high voltage DAC (HV DAC).

2.2 Calibration parameters

The calibration needed is a translation of HV DAC value to the electron count at the output of the gain register. This relationship is also temperature dependent. The equation which defines this relationship is

$$\ln(G) = \left(\frac{a_2 - T}{a_2 - T_{cal}}\right)\left(a_1 + a_4 e^{a_3\,DAC} + a_5 e^{2a_3\,DAC}\right)$$

Six independent constants are present, a_1, a_2, a_3, a_4, a_5, and T_{cal}. The five a_i values are determined through least squares fitting, while the T_{cal} value is simply reported; it is the temperature at which one of the fit stages is performed. These six all-important values are the only data product produced by the EM gain calibration. It should be noted that this functional form is not unique, and more complex relationships can be chosen. We have observed in at least two data sets now that this is not a perfect equation; there are small systematic errors which could be fixed with more parameters. However, one key feature preserved by this simple form is that it is algebraically invertible;

$$DAC = \frac{1}{a_3}\ln\left(\frac{-a_4 + \sqrt{a_4^2 - 4a_5\left(a1 - \left(\frac{a_2 - T}{a_2 - T_{cal}}\right)\ln(G)\right)}}{2a_5}\right)$$

And as we will see, the systematic errors appear to be within 3%, giving plenty of margin for other error sources within our overall 10% budget.

To understand how this conversion is used in practice, imagine a user wants to take a frame at some nonunity gain G. They command a gain of G, and then these six parameters plus the measured temperature are applied in the equation above to compute an HV DAC value. That HV DAC value is the voltage applied in the detector's gain register, creating the electron multiplication and producing a frame with gain G'. If G' is within 10% or less of G, we have successfully calibrated the EM gain.

3 Nonlinearity

3.1 Hierarchy problem

There are two independent variables relevant to this calibration that combine to produce the measured electron count at each pixel, incoming electrons and HV DAC value. As with any function of two orthogonal inputs, the output can be neatly visualized by a surface in 3D.

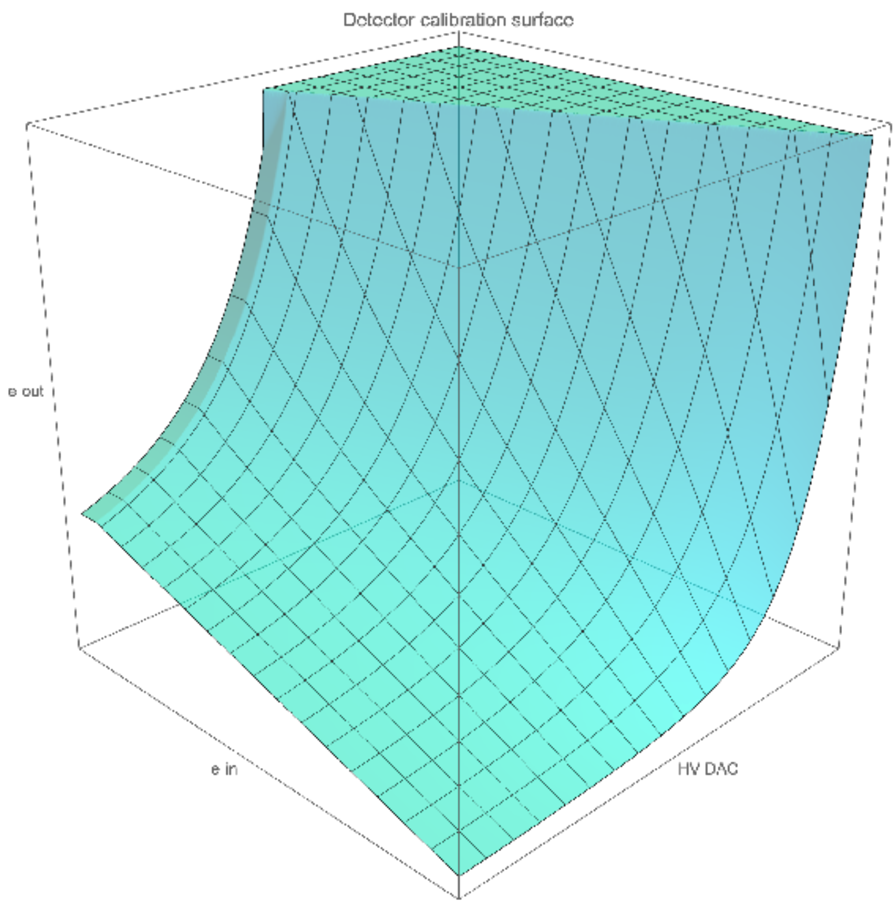

Figure 1: Surface displaying the output electrons as a function of input electrons and HV DAC value. The horizontal axes are electron in (left) and HV DAC (right), and the vertical axis is electrons out. The qualitative shape is being shown without numerical values on the axes because each detector and read out sequence is different.

This shape is characteristic for many detectors including EXCAM. The top shelf is the saturation value of the gain register, and the lip along the left side is the full well capacity of the CCD. Isolines in HV DAC are similar to photon transfer curves (electrons in vs. electrons out). Lines of constant "e- in" are used to calibrate EM gain since they quantify electrons out vs. HV DAC value. Thus, an inherent hierarchy problem becomes evident. Prior to any calibrations, raw frames containing output electron counts are received. If the number of input electrons could be known, the gain value could be then calculated (since the illumination is known, the ambiguity in the number of input electrons comes from detector nonlinearity.) On the other hand, if the gain value was known, the number of input electrons could then be calculated. But having neither quantity calibrated, it appears that no progress can be made on the other. This dilemma is encapsulated in Figure 2 below.

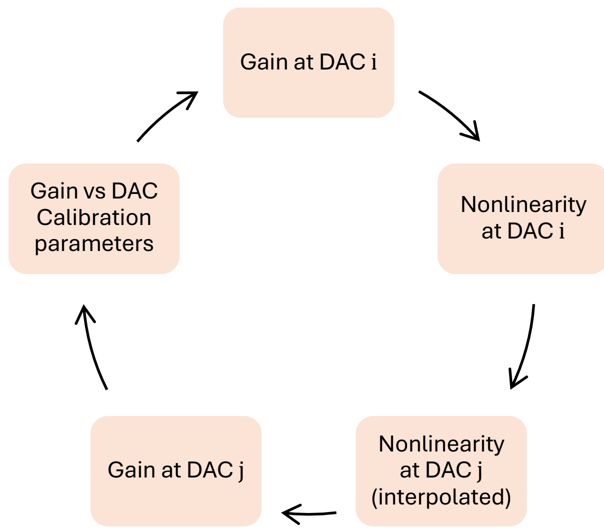

Figure 2: Loop of calibration steps illustrating a cyclical dependence.

Two quick notes: first, "Nonlinearity" here is a stand in for both the nonlinearity and k gain calibrations which are done together, and second, there are two analogous diagrams that could be made for dark current and clock induced charge (CIC) corrections, which we will discuss later. Now, it also must be understood that this hierarchy problem could in principle be abolished with more data. Notice that in the diagram there are two sets of frames represented, those at gain i and those at gain j. The EM gain calibration parameters are what allow us to interpolate between datasets i and j. If every single HV DAC value used for calibration was also given an entire corresponding photon transfer curve, dark current map, and CIC map, then i=j, and there is no problem, merely a clearly prescribed set of calibration steps. However, a data taking endeavor of that size is orders of magnitude beyond the already somewhat laborious protocol we use, and entirely impractical. Thus, we must find a more creative way to interject truth into the loop.

3.1 Unity nonlinearity approximation

The solution is to make an assumption somewhere in the loop. That is, to specify some information *a priori* to the calibration that is technically needed to know that information. We want to choose that assumption in such a way that the error on the final EM gain calibration is minimized. It turns out that the best assumption to make is that the nonlinearity is 1 for nonunity gains, and that can be justified in two ways. While it is true that there is only one proper calibration of nonlinearity, it's important to remember that the camera hardware is tuned prior to taking data. One of the parameters optimized in the procedure is the overall nonlinearity. This provides only a rough estimate, as the engineers only look at 2 gain values besides unity and they only use a handful of different illuminations. But what this preliminary data can tell us is that the nonlinearity values are going to be on the order of 1±1%. The second justification for the assumption is that after the loop closes and the calibrations are finished, we can check the values we get. Sure enough, they are all within a few percent of unity. And thinking critically about the process as you continue to read, you can convince yourself that the exact nonlinearity would be compounded away from unity through the calibration process. This means that the values we end with are even closer to 1 than the true values we assumed to be 1 at the beginning. I will call this assumption of unity nonlinearity *approximation 1.*

3.2 Estimated gain assumption

There is a second possible solution. Going back to the above diagram, what if the gains in datasets i and j were not identical, but overlapped? In other words, what if we obtained nonlinearity calibrations for just a few of the same HV DAC values used for EM gain calibrations? Could we just interpolate the rest of the actual gain values as a first pass? It turns out that this method will be used in one application; it is the only way to do the dark current and CIC map corrections. It could also in principle replace approximation 1, but it would have more error. We would be interpolating a highly nonlinear function (gain vs. HV DAC) across just a few data points. This is not the same as the nonlinearity interpolation (referenced in Figure 2) that is done in the final data pipeline, because that function remains nearly constant. I will call this assumption of interpolated gain value *approximation 2.*

4 Calibration

Although it is all part of the EM gain calibration, it is useful to organize the steps into two main parts, calibration proper and fitting. In calibration proper, we start with a raw frame and end with the actual gain value associated with that frame.

4.1 G<1000 Calibration

There are two completely separate methods for calibration proper. The first is used for estimated gains less than 1000, and the vast majority of HV DAC values used for data taking produce frames in this range. The three steps with asterisks denote that they are optional; we perform the calibration with and without them and compare at the end.

1. L1 frames are received
2. These are processed to L2a with the official pipeline
    a. Cropped to 1024x1024 image region
    b. Bias and prescan subtracted row by row
    c. Bad pixels and cosmic rays healed
    d. *Nonlinearity correction is applied *only* to unity gain frames. This is where approximation 1 is used on nonunity gain frames. The reason we can correct for nonlinearity exactly at unity is because we have a photon transfer curve at unity.
3. L2a frames are then processed to L2b with a modified version of the pipeline
    a. dN converted to electrons using the k gain value. This step is not gain dependent because the k gain is reported as a detector-wide constant.
    b. Fixed pattern noise map subtracted. This is also not gain-dependent.
    c. Dark current and CIC maps subtracted. These are provided at unity gain and must be multiplied by the frame's gain value. Obviously, the true gain value is not yet known, so here it is guessed. *The value can be varied. This is where approximation 2 is used.
    d. Desmeared

4. *Any residual background counts are then subtracted. For reasons beyond the scope of this explanation, after being processed to L2b there were still net positive counts in the unilluminated region of the frame.

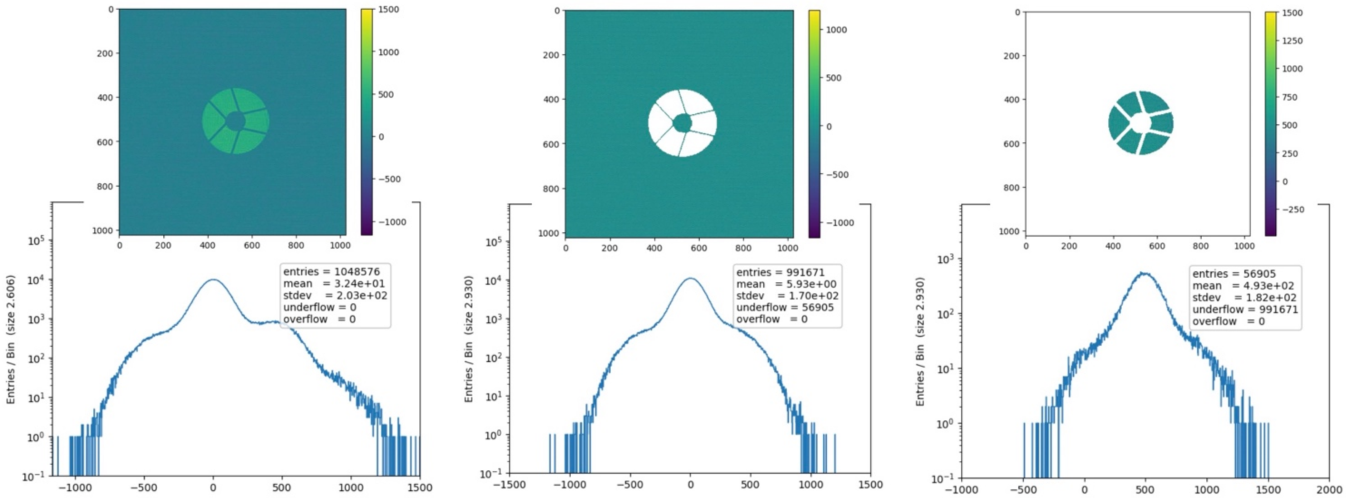

Figure 3: Electron counts for an example L2b EXACM pupil image with no mask (left), a background mask (center) and an image mask (right) applied.

Figure 3 shows an example L2b frame at unity gain, evenly illuminated within the Roman pupil region. Notice that there are still 5.93 average electrons per pixel in the background (center). This residual worsens as we move to nonunity gains. We take as the residual background the mean only in an annular region around the pupil rather than the mean of the entire background, as the pupil boundary is fuzzy and there are other effects near the edges of the frame. Figure 4 shows the region used.

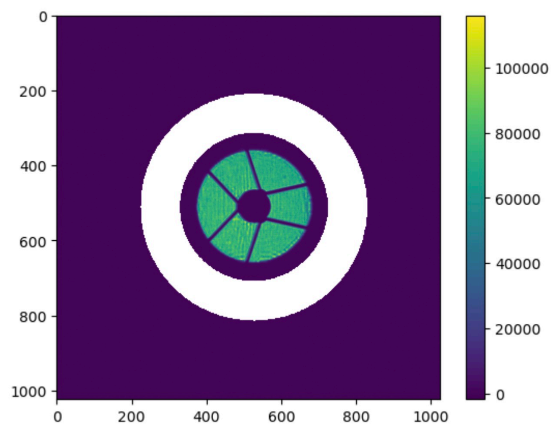

Figure 4: The white region of a sample pupil image is the annulus used to compute the residual background counts.

5. The illuminated region is then masked out and the mean electron counts computed
6. That value is divided by the frame's exposure time, leaving one number per frame that accurately reflects its electrons out / pixel / second; call that number R.
7. The R values of all frames of the same illumination, HV DAC, and temperature are averaged.
8. Steps 1-7 are applied to a batch of frames at unity gain, producing R_{unity}, and a batch at some nonunity gain value, producing R. The actual gain value is R/ R_{unity}.

4.2 G>1000 Errors

Prior to the fitting section, it's worth looking at the possible errors that we have incurred so far. The systematic errors are estimated by altering several of the calibration steps across the entire data set. The numerical values reported here are from the TVAC calibration. Turning off the nonlinearity correction to unity gain frames decreases the average gain by 1.2%. Turning off the residual background subtraction step decreased the average gain by 1.1% And finally, adjusting the value of the gain we guessed for the dark current and CIC maps affects the average gain almost negligibly. A 25% swing in either direction from our guessed gain results in a ±.001% change to the final result. Turning now to random errors, the noise was computed across R and R_{unity}, then added in quadrature. For well illuminated unity gain frames, this error source contributed around 2%. For unity gain frames that had low counts, the worst case was as poor as 5% error.

4.3 G>1000 Calibration

The simplest way to obtain the gain in this regime is to use the prescan part of the raw frames from CGI. These appear on the raw frame as an extended portion of the CCD image area but are really readings of the part of the serial register that need to be flushed out before each row of the CCD could be read out.

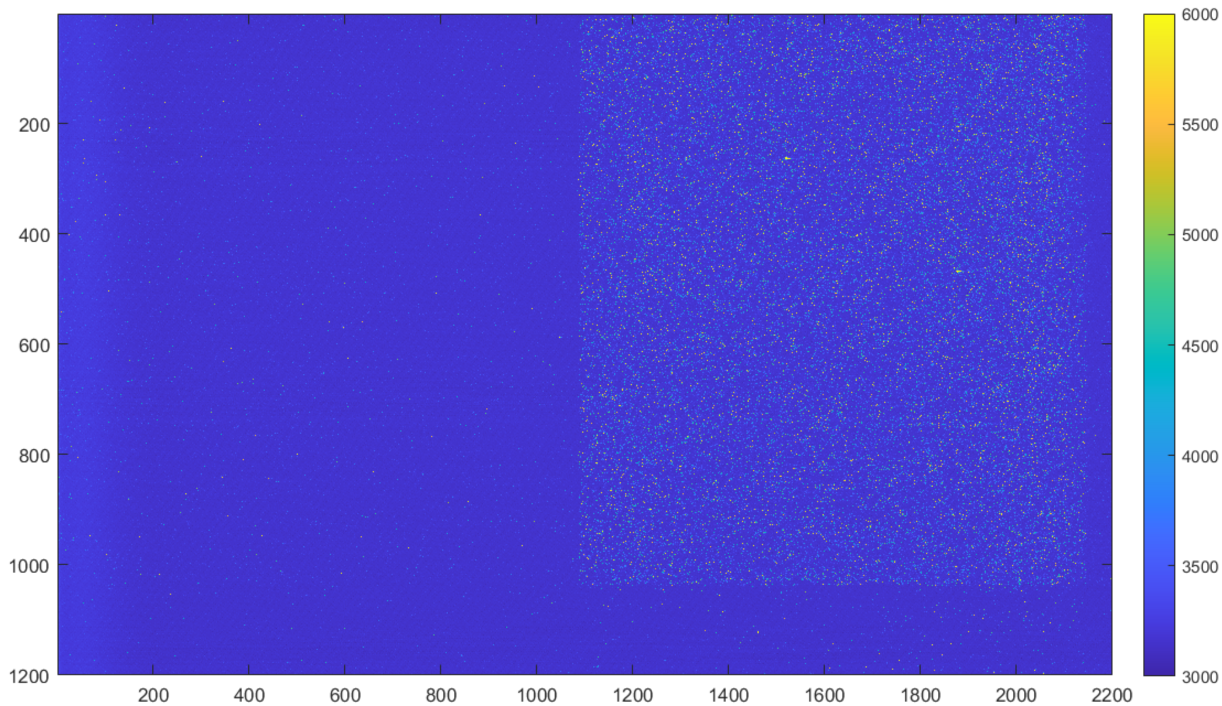

Figure 5: Unprocessed L1 dark frame with 100-second exposure time taken with EM gain of about x7815. CIC events are visible in the prescan region (left).

The clocking of this serial register occasionally creates CIC, at a rate much less than 1 electron per pixel per frame. These CIC electrons go through the gain register and result in an EM amplified signal. Since the EM gain process has a distinct probability distribution function, fitting a histogram of the observed serial prescan CIC counts to the known distribution can provide both the EM gain and the average CIC rate. CIC becomes noticeable on a histogram of pixel values from a dark region when the counts are scaled logarithmically. (See Figure 6 below.) The peak is read noise, and the 'shoulder' just above the peak (in counts) is the so-called partial CIC, explained below. The rest of the histogram is dominated by CIC events initiated before the EM gain register, which is what we fit to obtain the EM gain. According to the model, the line fit to CIC in the histogram has a slope of $-1/G$, enabling us to infer the actual EM gain value and compare it to the commanded value.

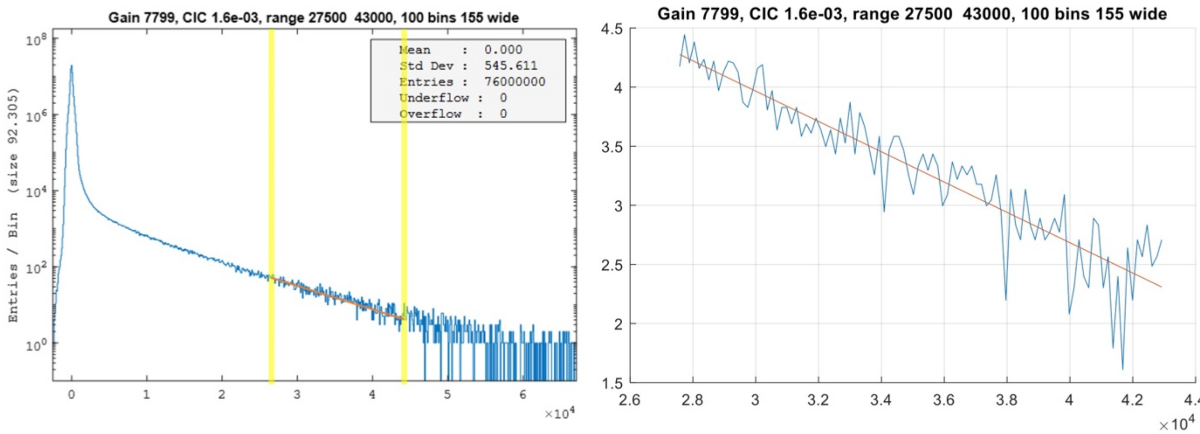

Figure 6: Histogram of counts (electrons), in log scale, from the prescan area. On right is a zoomed in region (also in log vertical scale) showing the linear fit, whose slope give the EM gain.

4.4 Competing distributions

As alluded earlier, complicating the linear fit are two additional distributions that can distort the result. The first is read noise, which is essentially a zero-mean Gaussian. Most of the serial prescan pixels have no CIC electrons and no EM amplification, and these pixels only show read noise. The relevant limitation is that the CIC histogram only becomes sufficiently disambiguated from the Gaussian read noise curve when the actual EM gain is very high, at least above ~1000. Thus, this method is only used in that regime.

Correcting for the second distribution, the partially amplified CIC ("partial CIC") is more challenging. This phenomenon occurs when unilluminated pixels acquire electrons after making some partial progress through the gain register. These events, which almost always are single-electron at origin, get amplified partially, and thus are called partial CIC events. Since they do experience some EM gain, partial CIC events have this tail that extends into the high-counts region. As a result, a fit to the high-counts part of the histogram that would otherwise have been purely serial CIC events with the same expected gain is now polluted with partial-gain events. This causes a systematic error in the calibration of the EM gain. A simple model of partial CIC assumes that the probability of the random CIC event is uniform across the entire EM register. There are 604 'gain stages' in the EM register. The first stage of the EM register virtually sees the full EM gain, and the last stage sees almost no EM gain, with all the intervening pixels seeing intermediate gains. A normalized probability distribution made of the sum of all these 604 probability distributions is a first approximation to the partial CIC probability distribution.

4.5 Independent distribution approximation

However, since the partial CIC model is not mature, as a first pass, this effect can be ignored by using only the higher counts region of the histogram, where the statistics are not as good. This allows us to make a reasonably accurate estimate of EM gain. Trial and error indicates that it is desirable to use sufficient data that, with bin numbers & widths both between 80 and 200, the bins with fewest counts should have at least 10 entries. Varying the number of bins as well as the precise location of the left edge allows an estimate of the level of certainty. For example, concatenating 90 frames with a commanded EM gain of 620, and varying the bin width while moving the left edge from 27650 all the way to 35550 (with a constant right edge of 50000), yielded an average estimate of 7955 ± 70. In TVAC, six other commanded gains were calibrated likewise, with 40 frames each, for a total of seven of these G>1000 or "high gain" values. During Roman commissioning we propose to double the number of high gain values used.

5 Data distributions

5.1 HV DAC value

It is simple to read the temperature and HV DAC value from the metadata of each frame. An HV DAC value could technically be computed from the commanded gain value and the calibration parameters that were in place when the frame was taken, however this is not recommended in any circumstance. After examining the two methods, they are not guaranteed to match. The latter can give an HV DAC value up to 2 increments away from the former (the true value). The reason for this is that there is always error in the fitting procedure for the a_i, and thus every time a_i are invoked to convert a gain to a DAC there is room for error. So, in summary, a frame's HV DAC value must always be read directly from its metadata.

5.2 Concept of operations

We are now armed with data points in HV DAC, gain, and temperature. This data embeds into a 3d space for fitting, but how is it distributed? This is of course determined by the actual data taking procedure that is chosen. A significant amount of forethought must go into designing the concept of operations for the data taking campaign. We can compare plans by simply sketching them in the HV DAC vs temperature plane.

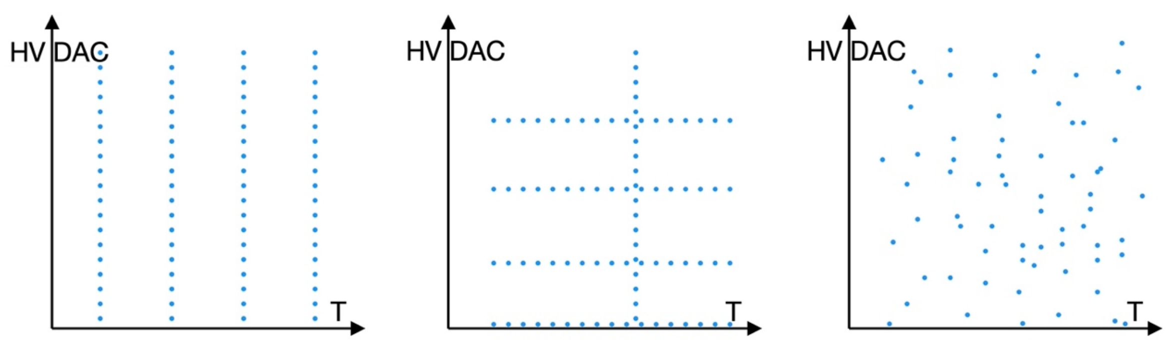

Figure 7: Notional distributions of planned data acquisition in the HV DAC – temperature plane. Shown are isotherms (left), isogains (center), and a random general distribution (right).

There are infinitely many possible patterns. The plot on the right illustrates the most general case. But only a few patterns make any logical sense to actually use. The merits and drawbacks of each were studied at length during the preparation for TVAC, and a few of those points are worth discussing here.

5.3 Core isotherm

If there were no temperature dependence, and we were only fitting DAC vs gain, our procedure would consist of only one vertical line. Meaning, all the data would fall on an isotherm. Thus, the "core" of any calibration scheme should have at least one well-sampled isotherm. You can see this in both the left and center plots above. Notice also that the points shown are evenly spaced in HV DAC. Because the user controls only commanded gain during the data taking procedure, we must specify the HV DAC values indirectly. To hit a target HV DAC value, we convert backwards to commanded gain. This conversion requires us to know in advance what temperature we will be operating at, as well as the values of the six calibration parameters that are loaded into the system at the time of the data acquisition. The end product is a list of gain values to command which are nowhere close to evenly spaced but that result in evenly spaced sampling of the parameter space. This is crucial, because the gain vs DAC curve is very steep. Sampling evenly in commanded gain would massively overbias any fitter to only the low HV DAC values. In TVAC, we decided to counteract this bias even more by doubling the sampling rate for the high gain frames.

The final question in designing this core isotherm is how many data points are needed? This was estimated for EXCAM by doing an extinction study. The five-parameter fit was performed with fewer and fewer data points subsampled as evenly as possible. A qualitative estimate is that around n=70 or so the convergence flatlines, meaning that random errors shrink below systematic. Therefore, a thorough calibration should most likely take at least 70 data points along this isotherm.

5.4 No thermal fit

With this established, we now consider altering the temperature. There is more subjectivity in this decision. There are three possible categories of choices that make sense. The first is trivial: take no thermal calibration data. As absurd as it seems, the reason for this choice is simply time and effort. Taking data at different temperatures requires additional documentation, the thermal team to be involved in the calibration, and it's only so fast that the temperature can be altered, so it adds time to the calibration. Recall that only the a_2 parameter requires data at different temperatures, the rest of them are temperature independent in the form of the calibration curve we chose. Additionally, the gain is much less sensitive to changes across the expected range of temperatures than it is to the HV DAC value, so temperature dependence is somewhat less critical to calibrate. In the scenarios where we choose to exercise this option, the value of a_2 is simply grandfathered in from a previous calibration.

5.5 Isogains

The next quickest option is to do the thermal calibration using a few lines of constant commanded gain. This corresponds to the plot in the center. At each chosen G>1, the temperature is swept from the minimum to the maximum, taking data along the way. Note that only the commanded gain is truly constant in this scenario, since the HV DAC and actual gain values are temperature dependent. The center plot above approximates these as horizontal lines, but they have a slight negative slope. This method is the fastest way to take data across a spread of temperatures because it does not require holding in any equilibrium conditions. Once a temperature extremum is crossed, the operator immediately reverses the thermal gradient. Roman ground software (GSW) has never been equipped to calibrate the gain with data taken this way; it has only been done with informal tools developed by the Tellus1 team. There are no plans to implement this ability into GSW in the future, so this data pattern will not be used for commissioning.

5.6 Isotherms

At last, we come to the most complex (within reason) but thorough data distribution, corresponding to the leftmost plot above. The basic principle is that we collect data along the same isotherm pattern that we used for the "core," but we do it at 3 or 4 different temperature values. This is by far the best dataset to have for calibration, and it is the intended use case for Roman GSW. The downside to this concept of operations is that it is very lengthy; it requires the detector to come to thermal equilibrium at each chosen temperature value. Equilibrium takes around 2 hours, and then the data collection can begin, so with no other delays the data collection time is around 3 hours per isotherm. To date, this calibration has only been performed once, on EXCAM during TVAC in hot thermal balance (HTB). (Hot and cold thermal balance refer to the temperature of the proximity electronics, and no such distinction will be made during the commissioning calibrations.) At the time, GSW was not prepared to do that calibration so internal

Tellus1 tools were used. GSW has now been updated, and this is the pattern that will be used for Roman commissioning on both LOCAM and EXCAM.

It is worth noting that outside of these basic patterns, there has not been success fitting a more general distribution. Fits to something like the righthand plot above have been simulated and struggle to converge at all. The precise reasons for this have not been studied in depth. The key information from this section is presented in Table 1 below.

Table 1: Summary of the three data distributions used during TVAC

Pattern	Time	TVAC GSW	Current GSW	When used
No thermal fit	+0 hours	yes	yes	TVAC HTB LOCAM TVAC CTB EXCAM
Isogains	+1-2 hours	no	no	TVAC CTB LOCAM
Isotherms	+8-10 hours	no	yes	TVAC HTB EXCAM Commissioning LOCAM Commissioning EXCAM

6 Fitting

6.1 Procedure

The characteristic shape of the HV DAC vs gain curve was seen in the right boundary of the calibration surface (Figure 1). Here we see a set of data at constant temperature proving this.

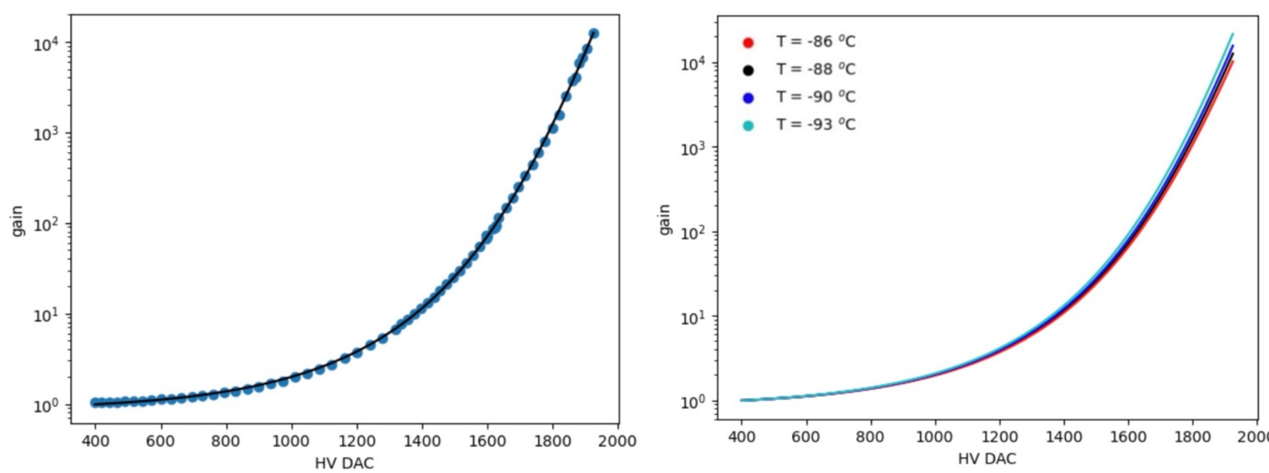

Figure 8: Actual data taken at calibration temperature overlayed with best fit (left), and modeled curve at 4 temperature values (right). The black model at -88 deg C is the calibration temperature.

Temperature adjusts the pitch of this curve, as modeled on the right. As discussed, this effect is not dominant; however, note the logarithmic scale in gain. At high gains, temperature fluctuations do cause the value to swing by multiples, emphasizing the need for this calibration.

Fitting is done in two stages. First, all the data is fed into a least squares fit for

$$\ln(G) = b_1(a_2 - T)e^{b_3\,DAC}.$$

Unlike a general fit to the actual relationship, fitting to this simpler equation allows for convergence. It is from this fit that we obtain the a_2 value. We also use the b_3 value as an initial guess for the a_3 value in our final fit. We throw away the b_1 value. Next, only the "core" series data is fed into the second least squares fit. Recall that the core was a series of evenly sampled values in HV DAC at some constant calibration temperature T_{cal}. The functional form for this fit is

$$\ln(G) = a_1 + a_4 e^{a_3\,DAC} + a_5 e^{2a_3\,DAC},$$

which is the full relationship, only with the temperature dependent factor absent. This yields a very nice convergence, with rms residuals at the 3% level. Applying the parameters back to data at all isotherms, RMS residuals are up to only 6%. You can see that most of the error is still systematic, and as discussed, that is because of the form we have chosen to model the HV DAC vs. gain relationship.

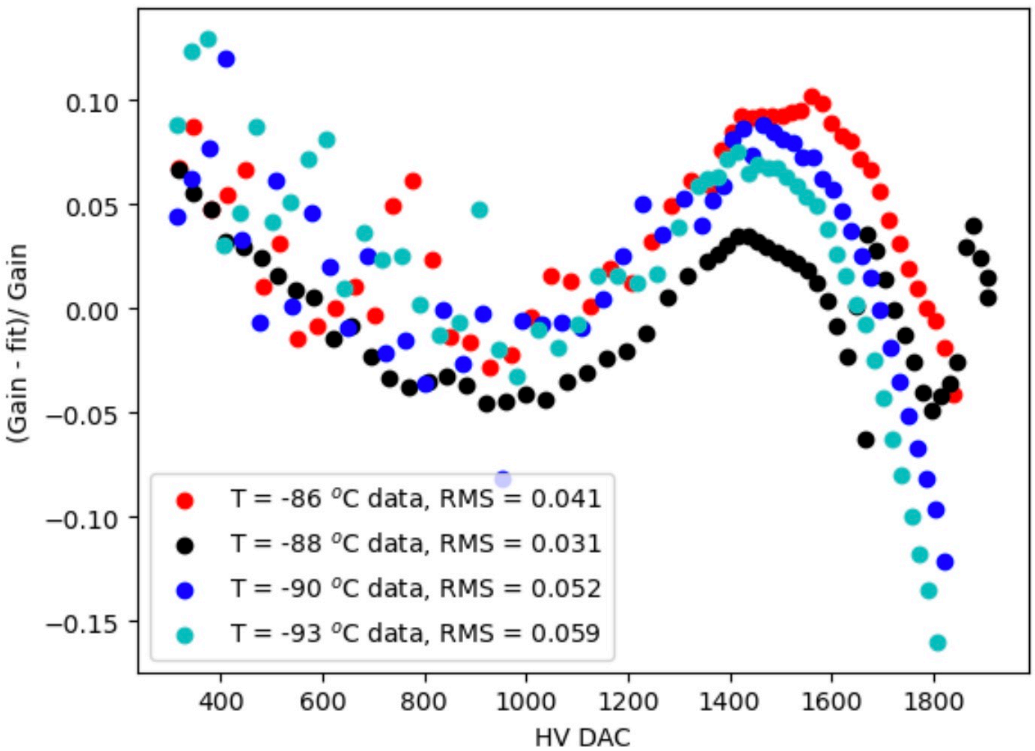

Figure 9: Fit residual at all temperatures using the best fit parameters obtained in a two-stage least squares fit.

6.2 Total error

The worst-case scenario occurs in extremely low gain frames (1<G<2) due to the low illumination in calibration data. In this regime we had ~5% random error in the determination of their gain, and less than 1% systematic error. The fit then had errors on the level of 10%. Thus, if we are being honest, the best we've done there is calibrate to within 11% or so. But we never expect to use this regime for normal Roman operations, and things improve rapidly as gain increases. Our calibration proper random errors reduce to 2%, systematic stays at 1%, and the rms residual is 3% at nominal temperature. This means that for the normal use case of Roman, we expect to be within 4%. Unless the EM gain is being used in some nonstandard regime (really close to unity but not unity, and multiple degrees outside the nominal temperature), a user will be able to specify a desired gain and receive a frame with an actual gain within 4%. Therefore, we have met and exceeded the opening criteria of accomplishing this task with 10% or less error.

References

Bijan Nemati, "Photon counting and precision photometry for the Roman Space Telescope Coronagraph," Proc. SPIE 11443, Space Telescopes and Instrumentation 2020: Optical, Infrared, and Millimeter Wave, 114435F (13 December 2020); https://doi.org/10.1117/12.2575983