SPACE TELESCOPE SCIENCE INSTITUTE
Operated for NASA by AURA

Instrument Science Report WFC3-2004-06

Non-Linearity Correction Algorithm for the WFC3 IR Channel

B. Hilbert
May 13, 2004

ABSTRACT

WFC3 IR channel data are obtained in the form of multiple, non-destructive readouts, grouped into "ramps", similar to the method employed by NICMOS. This strategy facilitates the characterization of the non-linearity of the IR Focal Plane Array (FPA), as the detector's response to a steady signal can be measured multiple times, until the detector becomes saturated. Using test data collected in this manner, the determination of correction coefficients for non-linearity in the WFC3 IR channel is performed through polynomial fits to the measured and ideal signal levels.

Introduction

As with all focal plane array detectors, the measured signal on the IR detector used for WFC3 is not directly proportional to the number of incident photons, especially when the measured signal is large. The goal of this study is to measure the signal level at which each pixel's response became non-linear, and to correct for this effect.

Theory

Each pixel in the WFC3 IR detector is based on the combination of positively- and negatviely-doped layers of HgCdTe, creating a p-n junction. This p-n junction acts as a capacitor, where the capacitance is determined by the voltage across the junction. At the beginning of each exposure, a reverse bias voltage is applied across the junction. This reverse bias is, by convention, negative. Equation 1 shows that each pixel has a capaci-

tance that decreases as the reverse bias voltage becomes more negative. The junction capacitance, C_{jn}, is given by Equation 1 (McCaughrean 1988), where V_{bi} is the diode built-in potential, and V is the reverse bias.

$$C_{jn} \alpha (1 - V/V_{bi})^{-1/2} \quad (1)$$

As photons impinge on the detector, they create electron-hole pairs, which are separated by the intrinsic electric field in the p-n junction. These photo-electons act to discharge the capacitance, and raise the voltage across the junction towards zero volts. In theory, measuring the voltage remaining at the end of the integration, and comparing this voltage with the bias (reset) voltage at the beginning of the integration should allow the calculation of the number of photo-electrons generated during the exposure. However, the captured photo-electrons have a second effect on the diode. As noted above, the capacitance of a diode is a function of the voltage across that diode. Thereore, as photo-electrons accumulate, discharging the diode and decreasing its voltage towards zero, the capacitance of the diode actually increases.

In general, the stored charge on an open circuit is equal to the product of the capacitance and the voltage on the circuit. (Equation 2) This implies a linear relationship between the measured voltage and the charge accumulated in the circiut. However, if part of the capacitance is a function of the voltage across the circuit, the relationship between charge and voltage is no longer linear.

$$Q = C * V \quad (2)$$

The non-linear effects are most apparent when the difference between the initial bias voltage and the voltage at the end of an observation is large, causing a large change in the capacitance. This large voltage difference translates into a large measured signal where the non-linearity is most visible. Therefore, in this study, data were taken with high-flux illumination sources, such that the difference between the initial reverse bias voltage and the final voltage was maximized.

Data

The linearity data used to derive and verify the correction coefficients, were collected at Goddard Space Flight Center's Dectector Characterization Laboratory (DCL). For the flight detector (FPA64), the linearity data consisted of 3 ramp files. (D12_linearity_16f_1300nm_01.fits, D12_linearity_16f_1300nm_02.fits, D12_linearity_16f_1300nm_03.fits) These three ramps were created with the detector being illuminated with 1300 nm monochromatic light. Three ramps were also created using 900 nm light, as well as three ramps using 1600 nm light. However, with different

filters and illumination levels at each wavelength, the three datasets could not combined. When tested independently, all datasets produced identical non-linearity corrections. For the remainder of this document, the non-linearity correction discussed is that resulting from the use of the 1300 nm data.

Method

Upon collection of the data, the first step in the data reduction was to run the files through the WFC3 IDL pipeline. The pipeline performed the initial steps of data reduction, including subtraction of the first read from all subsequent reads (to remove bias levels), subtraction of reference pixel values, and removal of any cosmic ray effects on the data.

Once the initial data reduction was completed, the pixel-by-pixel characterization of the non-linearity of the detector was performed using the following steps. First, a median non-linearity ramp was created by taking the median, on a pixel-by-pixel and read-by-read basis, of the three input ramps.

Next, each pixel's signal was examined up the ramp, in order to identify the extent of its non-linearity. Looking at the 16 reads comprising the ramp for a given pixel, a straight line was fit to reads 3 through 6. In this way, the fit was performed at a point in the ramp prior to any measureable non-linearity effects, while still allowing several initial reads for any instability effects to be minimized.

The differences between the best-fit line and the actual ramp were then examined in order to determine the point at which the ramp's signal deviated by 5% from the line. The exact signal at which the deviation was 5% was determined through interpolation between the existing reads. This point, for a given pixel, was designated as the "saturation" point for that pixel. All points in the measured ramp with signals higher than this saturation point were discarded for all subsequent calculations. An example of the best-fit for a given pixel is shown in Figure 1, where the fit, which was made using reads 3 through 6, has been extended to the end of the ramp.

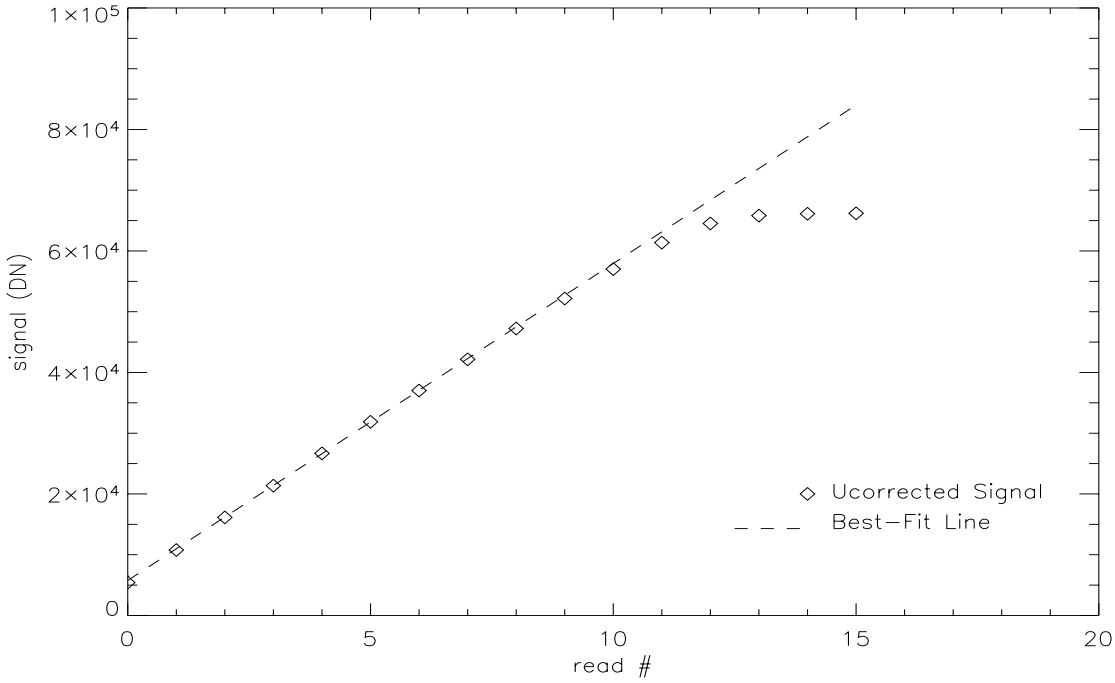

Figure 1: Signal measured in the original ramp (diamonds) versus read number within the ramp, for pixel (400,400). The best-fit line from a fit to reads 3 through 6 is plotted as the dashed line. The saturation level, where the measured signal deviates from the best-fit line by 5%, is calculated to be 63,840 DN.

The next step was to determine the coefficients of a polynomial such that, when the measured signal is multiplied by the coefficients, the best-fit, linear signal results. In order to accomplish this, the linear signal was plotted versus the measured signal. The resulting curve was then fit with a polynomial. Unfortunately, the decrease in measured signal rate relative to the linear signal rate at the end of the ramp made fitting with a single polynomial difficult. As a result, the curve was split into two sections, divided at 75% of the maximum linear signal. Each section was fit with a 3^{rd}-order polynomial. For each pixel, the coefficients for both polynomials, as well as the cutoff signal level at which the two fits meet, were saved in a FITS file, for later application to other data files.

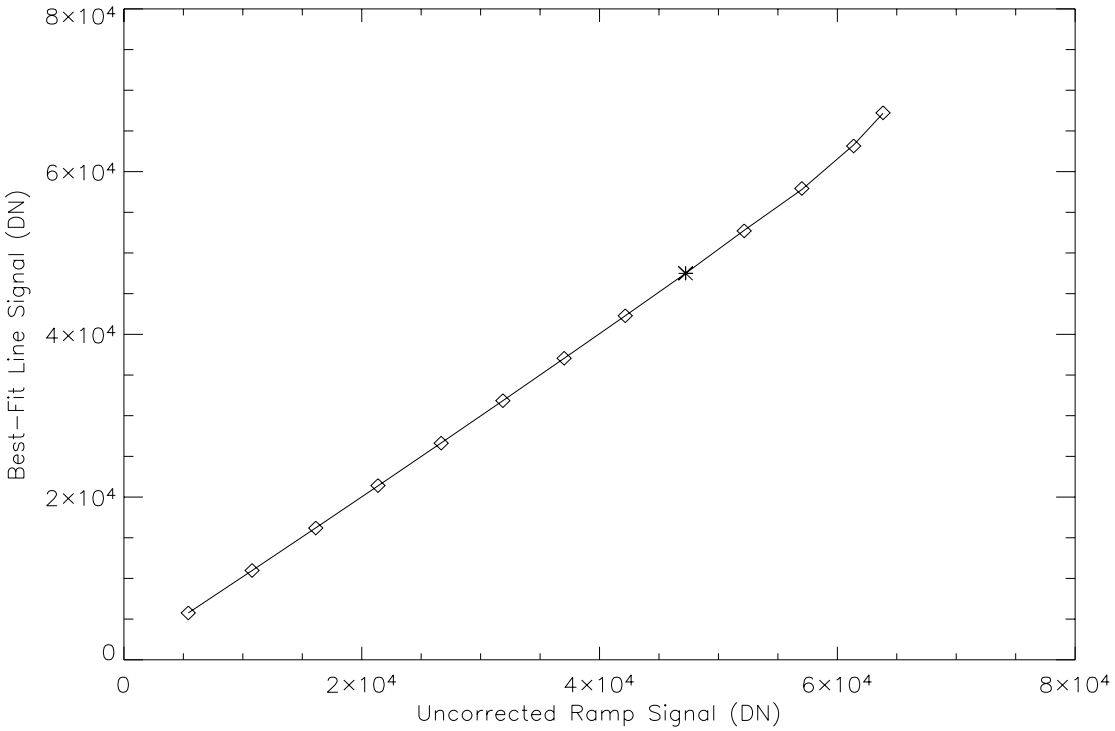

Figure 2: The two curves plotted in Figure 1 are plotted here as a single curve (diamonds): The original measured signal is plotted versus the best-fit linear signal. After fitting this curve with two 3rd-order polynomials, the combined fits were overplotted (solid line). The maximum error of the polynomial fits' signals relative to the best-fit linear signal is 0.2%. In this case, the 8th read, plotted here with a star, is closest to the 75% singal level, and is the read at which the two polynomial fits meet. Attempting to use a single 3rd order polynomial to fit the entire ramp resulted in errors of several percent.

Analysis

Self-Correction

The first test of the non-linearity correction coefficients was to correct the individual non-linearity files themselves. In order to apply the correction to a given pixel in the data files, the original measured signal levels in the pixel were multiplied by the coefficients from the 3rd-order fits. Signals below the cutoff level were multiplied by the coefficients from the lower 3rd -order fit, while signals above the cutoff level were multiplied by coefficients from the upper 3rd-order fit, as shown in Equations 3 and 4, where x is the measured signal, c_0 through c_3 are the correction coefficients for the lower signal range, c_4 through c_7 are the correction coefficients for the higher signal range, and y is the corrected signal. Ideally, the signal levels after this multiplication process should be linear for the

duration of the ramp up to the saturation level. The coefficient containing the zero point for the lower signal range (c_0) is subtracted from Equation 2 in order to assure that the high signal fit matches the low signal fit at the signal level where the two fits meet. c_0 is not used in Equation 1 to prevent pixels with very little signal being "corrected" up to c_0.

$$y = (c_1 * x) + (c_2 * x^2) + (c_3 * x^3) \quad (3)$$

$$y = c_4 - c_0 + (c_5 * x) + (c_6 * x^2) + (c_7 * x^3) \quad (4)$$

A more sensitive way to check the linearity of a ramp, and therefore the success of the non-linearity correction, is to look at the signal rates across a ramp. Figure 3 shows the signal rate both before and after the non-linearity correction for pixel (400,400).

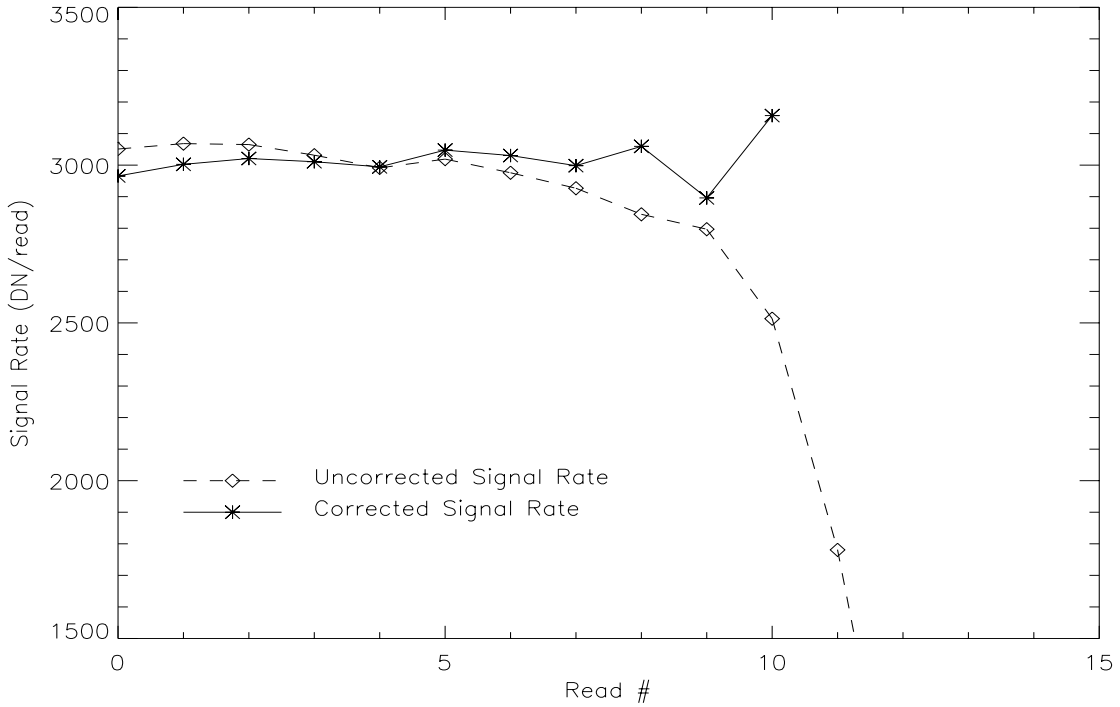

Figure 3: Corrected and uncorrected signal rates for pixel (400,400) in one of the three ramps used to calculate the correction coefficients. A perfectly linear detector would have a constant signal rate, and therefore a horizontal line in this plot. The corrected signal rates extend only to read 11 because the signal reached saturation at this point.

Independent Data

The second test of the non-linearity correction algorithm was to apply the calculated correction coefficients to data other than those used to calculate the coefficients. For this purpose, the correction coefficients were applied to gain data taken two days prior to the linearity data. The results from applying the correction coefficients to pixel (400,400) are

shown in Figure 4. These data were at a lower signal level than the linearity data, resulting in a more subtle correction in the signal rate.

An interesting aspect of the correction of the gain data, as well as that of the linearity data, was the relatively large correction performed on the beginning of the ramp, where the signal was low and should not have been affected by a non-linear detector response. This non-linearity is a result of the instability present in the FPA. By calculating the coefficients of the correction using a best-fit line based on reads 3 through 6, the contamination from the instability was minimized. When these coefficients were applied to data in which the instability was present, such as the gain data, the non-linearity coefficients acted to remove some of the instability. Not all of the instability effects however, were removed, as seen in Figure 5. In an attempt to study the effects of the instability, non-linearity correction coefficients were also calculated through the use of best-fit lines to reads 1 through 4 in the original data, rather than reads 3 thorugh 6. Fitting the earlier reads in the ramp makes the best-fit more susceptible to instability effects, and produces a different correction. Figure 5 shows that both versions of the signal correction produce ramps with constant, but not identical, signal rates. The correction for instability will be discussed in a future ISR.

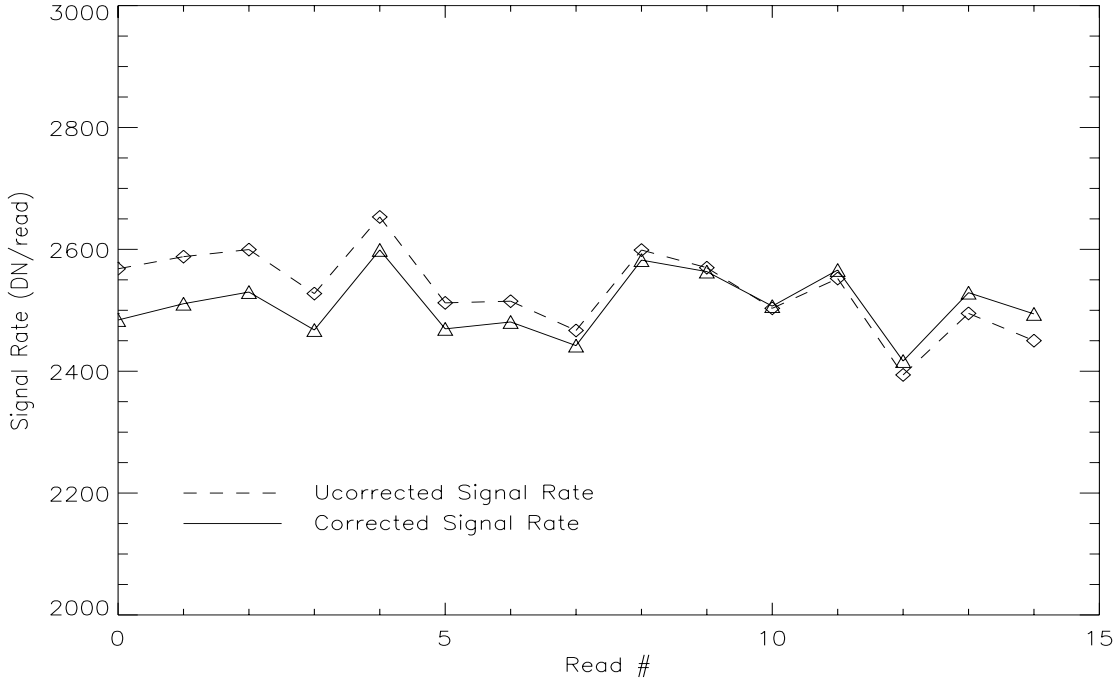

Figure 4: Corrected versus uncorrected signal rates for pixel (400,400) in gain data taken independently from the linearity data shown in Figures 1 through 3. The correction here is more subtle, but still shifts a ramp with an overall decreasing signal rate to one with a constant signal rate.

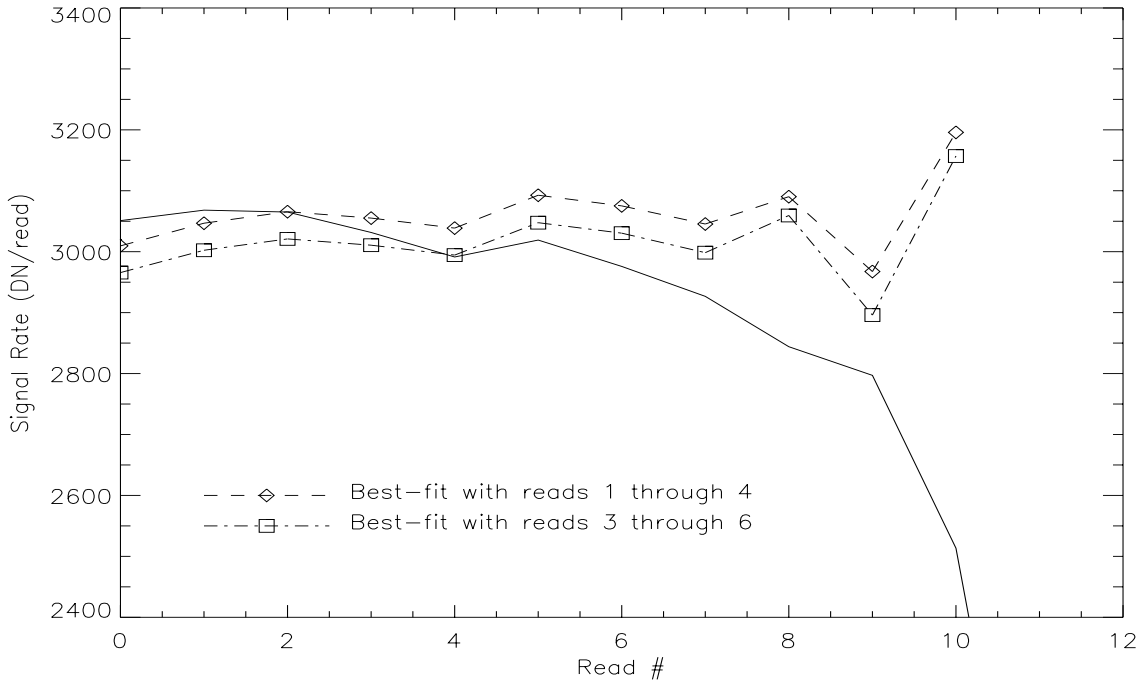

Figure 5: Same plot as Figure 3, with signal rates before and after different non-linearity corrections. Squares show the same correction as in Figure 3. Traingles show the signal rate with correction coefficients based on best-fits to reads 1 through 4 , rather than reads 3 through 6.

Conclusions

The inherent non-linear response of the WFC3 IR channel can be reliably corrected through effective polynomial-fitting. The use of two 3rd order polynomials to approximate the curve of measured versus corrected signal results in a more accurate fit to the data, which yields a better non-linearity correction. In addition, by fitting a straight line to reads 3 thorugh 6 of the non-linearity data, most of the effects of the detector's instability can be avoided. Applying the correction coefficients generated with this method to independent FPA64 data results in data ramps with linear signal rates.

More linearity data will be taken during WFC3's Thermal Vacuum testing, at which point this method will be used to create a more accurate calculation of the correction coefficients.

Acknowledgements

Many thanks to Massimo Robberto and Megan Sosey for discussions on the finer points of infrared array detector operations, and non-linearity corrections.

References

McCaughrean, M. J. *The Astronomical Application of Infrared Array Detectors*. University of Edinburgh, Doctor of Philosophy thesis, 1988.