STScI | SPACE TELESCOPE SCIENCE INSTITUTE

Instrument Science Report WFC3 2016-16

Reprocessing WFC3/IR Exposures Affected by Time-Variable Backgrounds

Gabriel Brammer

November 2, 2016

Abstract

The background seen in WFC3/IR observations frequently shows strong time-dependent behavior above the constant flux expected for zodiacal continuum light. This is often caused by an emission line of helium at 1.083 μm excited in the sun-illuminated upper atmosphere, when seen in the filters (F105W, F110W) and grisms (G102, G141) sensitive to the feature. The default behavior of the `calwf3` *pipeline assumes constant source-plus-background fluxes when it performs up-the-ramp fitting to identify cosmic rays and determine the average count rate within a MULTIACCUM IR exposure.* `calwf3` *provides undesirable results in the presence of strongly variable backgrounds, primarily in the form of elevated and non-Gaussian noise in the* `FLT` *products. Here we describe methods to improve the noise properties of the reduced products. In the first, we simply turn off the* `calwf3` CRCORR *step, treating the IR detector as if it were a CCD, i.e., accumulating flux and reading it out at the end of the exposure. Next, we artificially flatten the ramps in the IMA products and then allow* `calwf3` *to proceed as normal fitting the ramp and identifying CRs. Either of these procedures enable recovery of datasets otherwise corrupted beyond repair and have no discernible effects on photometry of sources in deep combined images.*

1. Introduction

The background seen in the IR channel of the Wide-Field Camera 3 instrument is a combination of zodiacal light, scattered light from the bright Earth limb and line emission at 1.083 μm from helium atoms excited by sunlight in the day-side upper atmosphere (Brammer et al. 2014). The strength of the first component, the zodiacal light, depends on the orientation of the target with respect to the sun (i.e., as a function of ecliptic latitude and Sun angle; Pirzkal 2014), which varies throughout the year but is effectively constant within a given exposure/orbit/visit. The scattered light and line emission components can vary *within* an orbit and even within a single exposure. Additional scattered light can be seen as the target-to-limb angle decreases throughout an orbit (Giavalisco et al 2002), occasionally even causing a "blowout" with a strong scattered light pattern seen on the left side of the detector, primarily at very low limb angles $< 20^\circ$ (Hilbert & McCullough 2009). The He emission line component is seen when the spacecraft leaves the Earth shadow and enters the illuminated atmosphere, and the He line backgrounds can reach count rates exceeding 4 e^-/s (Brammer et al. 2014), i.e., many times the zodi count rate (cf. $\sim$1 e^-/s). The strength of the He 1.083 μm background likely depends on the observed path-length through illuminated atmosphere, and as a diffuse source it is not expected to show spatial structure across the WFC3/IR field of view.

WFC3/IR exposures are taken in a "MULTIACCUM" sequence of non-destructive reads specified by the user from a variety of pre-defined configurations with both constant (SPARS) and increasing (STEP) time intervals between subsequent reads. At the time of each read, the accumulated charge on each pixel is recorded and this charge accumulation history for a given exposure—the "ramp"—is recorded in the RAW data files. While not affecting the charge on the pixel, the read process is inherently noisy at the level of $\sim$20 e^-/read (i.e., the detector "read noise"). Fitting a linear function to the multiple reads of the ramp decreases the effective read noise of the final exposure: Hilbert & McCullough (2009) find that the read noise is decreased from 20 e^- to 12 e^- when fitting a ramp of 15 samples of a SPARS200 sequence.

Incident cosmic rays (CRs) strike essentially instantaneously and only affect a single read of the read sequence. The `calwf3` pipeline identifies these CRs as discontinuities in the linear ramp, first identifying and excluding the discontinuous read and then fitting and averaging linear ramps before and after the discontinuity. A key assumption of the CR identification algorithm implemented in `calwf3` (i.e., the CRCORR pipeline flag) is that, aside from CRs, the pixel ramps are linear—or, equivalently, that the count rates are constant over time. Strong time variation in the background can trip the CR thresholds, with most or all of the image identified as a CR at a given read. Furthermore, the background variation is fairly smooth from read to read so the algorithm that tries to iteratively identify clean reads before

and after a CR hit will likely fail.

The main result in the resulting FLT products in the presence of strong background variations is increased noise, as only a fraction of the available exposure time (perhaps just one or two reads out of 15) was identified as valid. Furthermore, the distribution of background pixel values frequently shows multi-modal non-Gaussian shapes as different parts of the image trip and confuse the CR algorithm in different ways. *These FLT products are unusable and should be discarded or reprocessed!* In the sections below, we describe techniques and tools for identifying and reprocessing exposures affected by variable backgrounds.

2. Example affected exposures

If the user does not specify the "SHADOW" special requirement in the proposal planning software, nearly every orbit-long visit using the F105W or F110W filters and/or the G102/G141 IR grisms will see the He 1.083 μm emission line at some point during the orbit when the spacecraft is out of the Earth shadow. The He 1.083 μm emission can be seen either at the beginning or end of an orbit, depending on the shadow phasing, and the background count rates will be seen to decrease or increase, respectively, throughout an exposure. Due to scheduling constraints around SAA passages, the notional orbit planned during Phase II preparation in APT can be split over subsequent actual orbits of the spacecraft, and it can also happen that every or no exposures of a given visit will be affected by the 1.083 μm line, depending on the phasing of the target visibility and the shadow passages. Enforcing that two exposures planned to be executed sequentially within an orbit requires setting the "Non-Int"terruptible sequence flag in APT.

Here we demonstrate some of the background behavior using exposures from the HUDF12 GO program (GO: 12498, PI: Ellis), which obtained many multi-orbit visits with F105W exposures. Two of these exposures are shown in Fig. 1. The first was taken with the spacecraft in the Earth shadow and had a relatively flat background dominated by zodiacal light. The second exposure, taken in the same orbit and which started just as the spacecraft left the shadow, shows a dramatic increase in the background count rate during the exposure starting with 0.5 e^-/s zodi at the beginning and reaching nearly 3 e^-/s by the end of the exposure dominated by the 1.083 μm line. Brammer et al. (2014) discuss in more detail statistics of the 1.083 μm backgrounds seen in archival observations.

Fig. 2 shows cutouts from the default `calwf3` FLT products for the two exposures whose ramps were shown in Fig. 1. The second exposure, which suffered the strong increase in the background flux from the He line, is much noisier than the first. The increased noise is partially a real effect, owing to the larger photon noise from the increased background flux (the exposures are background noise limited). However, the right panel shows the

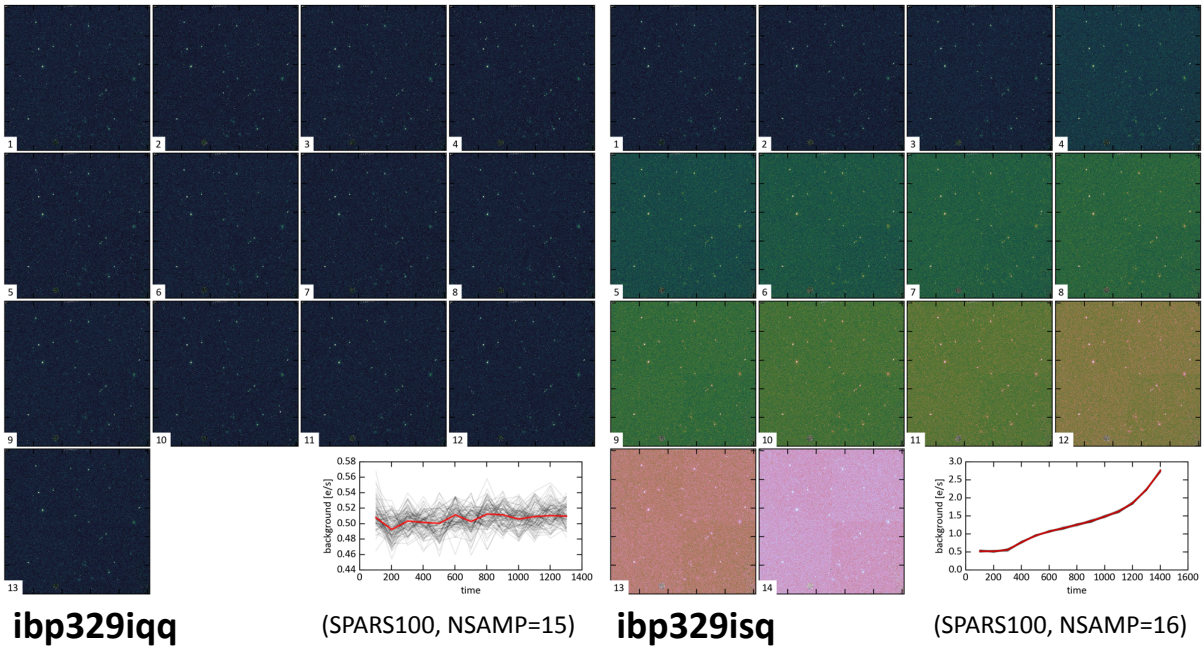

Fig. 1.— Each set of panels shows the individual reads of two back-to-back F105W exposures from the HUDF12 GO program. The small inset panels show the average background count rate as a function of time throughout the exposure, with samples recorded every 100 s for the adopted SPARS100 sequence. The background count rate is essentially constant for the first exposure, `ibp329iqq`, while the spacecraft was in the Earth shadow. The spacecraft left the shadow about at read #2 of the second exposure of the orbit, `ibp329isq`, at which point the background from the atmospheric He 1.083 μm emission line increases dramatically. Note that the read numbering here is such that read #1 is the first 100 s read after the 2.9 s reset read, which is actually the third read recorded in the RAW/IMA files.

distributions of pixel values for the indicated cutouts: the distribution is not only broader for the second exposure, as expected, but it is also multimodal and non-Gaussian, which is not expected. For this exposure, the `calwf3` ramp fit was compromised by the rapid increase of the background flux during the exposure and the FLT product as-is is corrupt and unusable (a loss of half an orbit of telescope time).

3. Mitigation

3.1. "Last-minus-first"

As it is the breakdown of the ramp fit that corrupts the FLT data products such as for the exposure shown in Fig. 2, a trivial mitigation strategy is simply disabling that step of the `calwf3` processing by setting the header keyword CRCORR='OMIT' in the RAW exposure file, as demonstrated in the code snippet below. This procedure discards the information accumulated throughout the exposure read sequence and treats the WFC3/IR detector more like a CCD that accumulates charge that is read out only at the end of the exposure. In this case for WRC3/IR, the observed count rate is determined by simply subtracting the first from the last reads of the detector ("last-minus-first") and dividing by the time elapsed between them.

Code 1 "Last-minus-first" processing

```
import astropy.io.fits as pyfits
import wfc3tools

raw_file = 'ibp329isq_raw.fits'
img = pyfits.open(raw_file, mode='update')
img[0].header['CRCORR'] = 'OMIT'
img.flush()

wfc3tools.calwf3.calwf3(raw_file)
```

Fig. 3 shows the cutouts and pixel statistics after disabling the CRCORR pipeline processing. The distribution of pixel values in the compromised exposure now shows a single peak and the noise properties in the image itself are clearly better behaved. Assuming read noise $RN = 20$ e$^-$, using

$$(\sigma t)^2 = RN^2 + b \cdot t \tag{1}$$

predicts $\sigma = 0.025$ e$^-$/s ($\sigma = 0.034$ e$^-$/s) for the pixel distributions, almost exactly as observed, given the observed median background count rates, b, and the integration times, $t = 1302$ s ($t = 1402$ s), for the two exposures.

The cutouts of Fig. 3 highlight the primary drawback of the strategy disabling the CRCORR fit: cosmic rays that accumulate throughout the exposure are not identified and not flagged at the pipeline level. If multiple dithered exposures are available, the cosmic rays can be identified with AstroDrizzle as for CCD exposures with ACS and WFC3/UVIS (e.g., §4.2.7 of Gonzaga et al. 2012). Another consequence of the "last-minus-first" method is the loss of dynamic range for bright sources that saturate during the read sequence and whose fluxes can otherwise be recovered by considering only the early reads where the pixels are not saturated.

3.2. "Flatten-ramp"

As mentioned above, the up-the-ramp `calwf3` treats each pixel separately and fails when the pixel ramps become non-linear in the presence of the time-variable sky backgrounds. Under the assumption that the background of the entire image moves up and down as a pedestal (i.e., the effective background is the same for each pixel), we can "flatten" the non-linear ramps by subtracting the variable background of each read averaged over many pixels, b_i, and adding back a constant for the overall average background count rate $\langle b \rangle$.

In practice, we first run `calwf3` up to the point of generating the calibrated IMA files but omitting the cosmic-ray identification step. We flatten the ramp in the IMA files then feed the flattened IMAs back to `calwf3` for the final step of identifying the cosmic rays. The full dynamic range of the image is recovered using the `calwf3` ramp fit that only considers pixel measurements before they reach nonlinear/saturation levels. Example Python code is provided below demonstrating this technique.

Code 2 "Flatten-ramp" processing

```
import numpy as np
import astropy.io.fits as pyfits
import wf3tools

### Generate the IMA file
raw_file = 'ibp329isq_raw.fits'
wfc3tools.calwf3.calwf3(raw_file)

### Work on the IMA product
ima = pyfits.open(raw_file.replace('raw', 'ima'), mode='update')

### Subregion for stats, here just the whole image
stats_region=[[0,1014], [0,1014]]
slx = slice(stats_region[0][0], stats_region[0][1])
sly = slice(stats_region[1][0], stats_region[1][1])

### Subtract per-read median count-rate scalar and add back in
### full exposure count rate to preserve pixel statistics
total_countrate = np.median(ima['SCI',1].data[sly, slx])

for i in range(ima[0].header['NSAMP']-2):
    med = np.median(ima['SCI',i+1].data[sly, slx])
    ima['SCI',i+1].data += total_countrate - med
    print '%s, read #%d, background:%.2f' %(raw_file, i+1, med)

ima[0].header['CRCORR'] = 'PERFORM'
ima[0].header['DRIZCORR'] = 'OMIT'

### Write the updated IMA
ima.flush()

#### Run calwf3 on cleaned IMA
wfc3tools.calwf3.calwf3(raw_file.replace('raw', 'ima'))
```

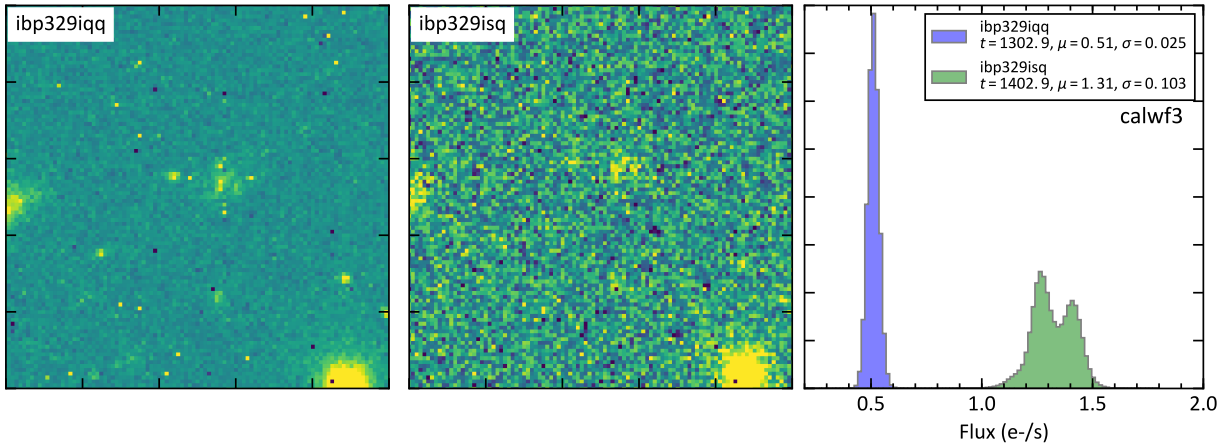

Fig. 2.— The left and middle panels show 200×200 pixel cutouts of the HUDF12 F105W exposures shown in Fig. 1, and the right panel shows the distribution of pixel values in those images. The `calwf3` pipeline has failed to properly account for the strong variation in the background levels of the second exposure, and the result is significantly increased non-Gaussian noise in the FLT pipeline product.

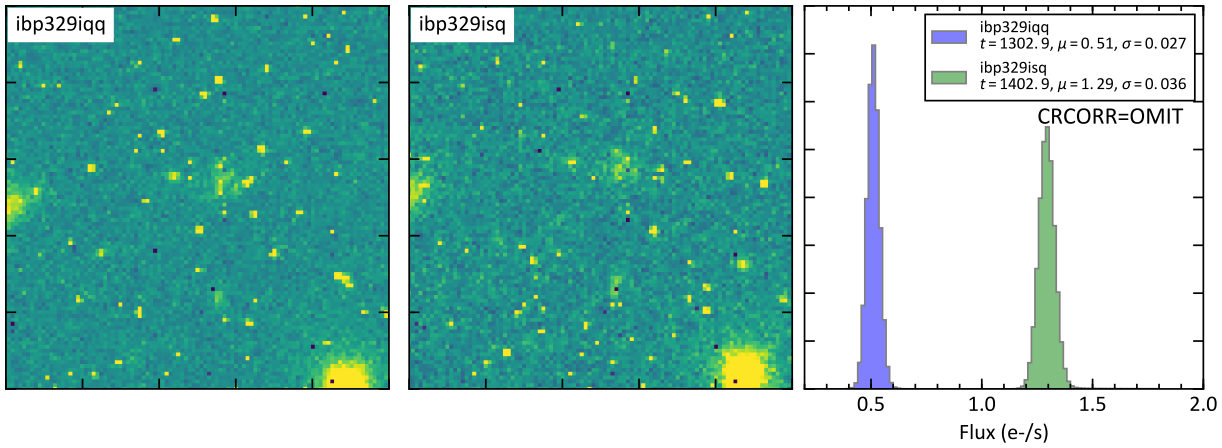

Fig. 3.— Reprocessing `ibp329isq` with CRCORR=OMIT (§3.1) significantly improves the statistics of the FLT product. The noise is somewhat higher in the second exposure with the higher total background counts, but the Gaussian distribution of the pixel values is recovered. This improvement comes at the obvious cost of leaving the exposures peppered with cosmic rays, which must now be identified by some other means.

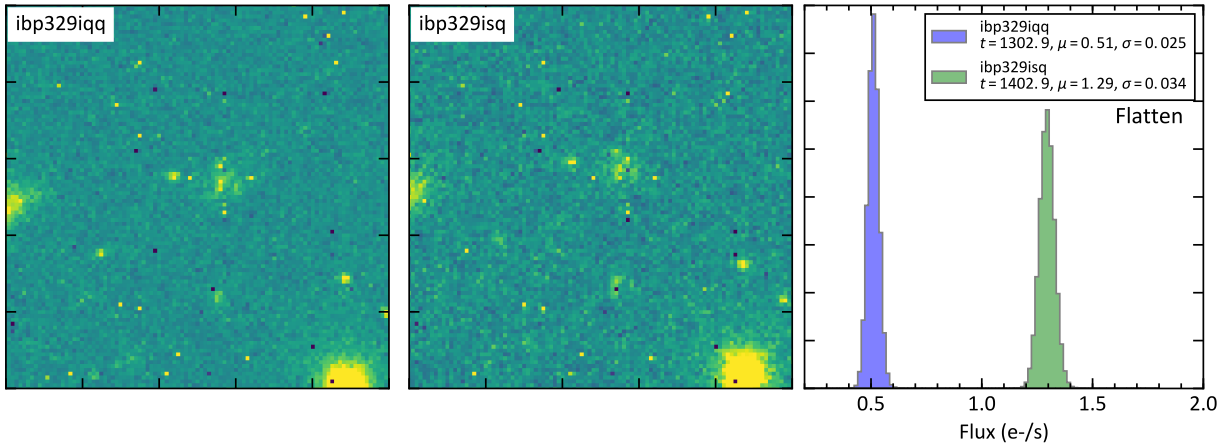

Fig. 4.— Result of reprocessing exposures with the "Flatten-ramp" algorithm in §3.2. The statistics of the processed image are indistinguishable from taking the result directly from "CRCORR=OMIT", but now with the benefit of having the cosmic rays identified by `calwf3` for the individual exposure.

4. Comparing reprocessed products to `calwf3`

The "flatten-ramp" algorithm described above modifies the read sequences in the IMA files, but it should not affect photometry of resolved sources in images that have been reprocessed in this way. Here, we check the reliability of the reprocessed photometry by generating two versions of the HUDF F105W mosaic: the first consisting only of exposures with flat ramps dominated by the zodiacal light throughout and the second consisting of highly variable ramps from the 1.083 μm sky emission line that have been reprocessed as described above. The count-rate curves for each of these two subsets of exposures are shown in the left panel of Fig. 5. Many of the the variable ramps shown here have corrupted default pipeline FLT images as in Fig. 2 that would essentially be unusable without reprocessing.

We measure photometry on the deep combined images with the SExtractor software and compare the "MAG_AUTO" magnitudes derived from the two mosaics in the right panel of Fig. 5. As expected, there is no discernible effect on the photometry of the reprocessed images; all residuals are consistent with zero within the photometric uncertainties shown in the shaded red band.

5. Excising Individual Reads

Some WFC3/IR exposures are impacted by events that only occur during individual reads of the detector, such as satellite trails that fly through the field of view but only affect

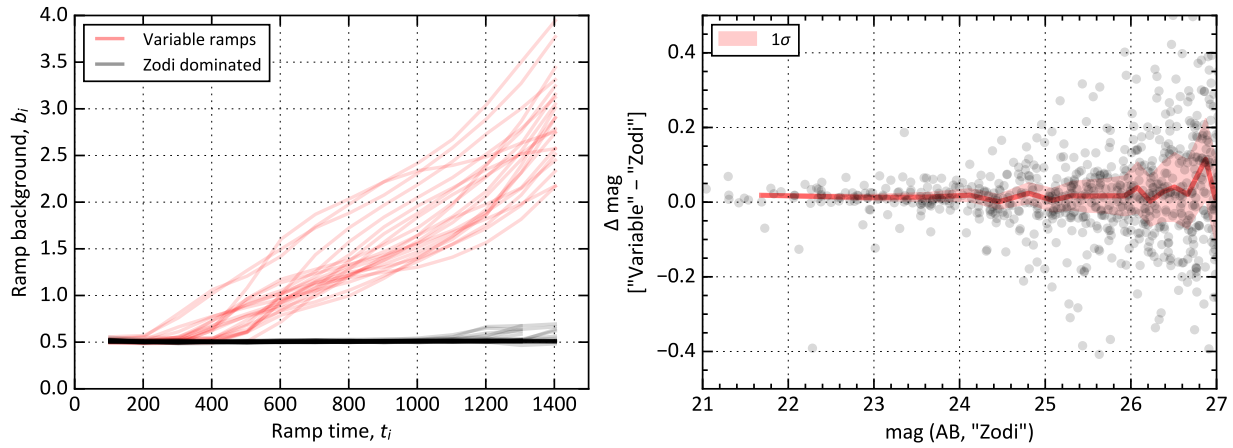

Fig. 5.— Photometric comparison between exposures with flat, zodi-dominated ramps (black curves) and exposures with highly variable, 1.083 μm dominated ramps (red curves). The right panel shows residuals of photometry generated from deep stacks of the subset of images whose background ramps are shown at left. There are no systematic trends or offsets suggesting that the "flatten-ramp" algorithm preserves photometric quality while enabling recovery of otherwise corrupted calibrated images.

single reads of a MULTIACCUM exposure sequence. While they are conceptually similar to cosmic rays, `calwf3` struggles to identify these types transient events that can affect many adjacent pixels, and with standard processing residuals of the short event will often be seen in the final calibrated image. The affected pixels must be masked if the exposure is to be used for subsequent processing with, e.g., AstroDrizzle.

The up-the-ramp `calwf3` processing does not provide a mechanism for excising individual reads of a MULTIACCUM sequence before computing the average calibrated count rate of an exposure. However, if we treat each individual read of the MULTIACCUM sequence as a short independent exposure, we can remove any individual reads affected by transient events as necessary. An example of an exposure with a transient satellite trail is shown in Fig. 6. These satellite trails can be identified by visual inspection of "ramp" images such as that shown in the figure. We remove the affected read #9 from the sequence and the final cleaned calibrated image is shown at right.

In practice we first generate a "last-minus-first" total fluence image (§3.1) from a calibrated IMA file by running `calwf3` and omitting the cosmic ray identification step. This image of course contains all of the cosmic rays accumulated throughout the exposure as well as the satellite trail. We then compute intermediate difference images of each read that we wish to excise from the stack and subtract them from the total, also subtracting the excised

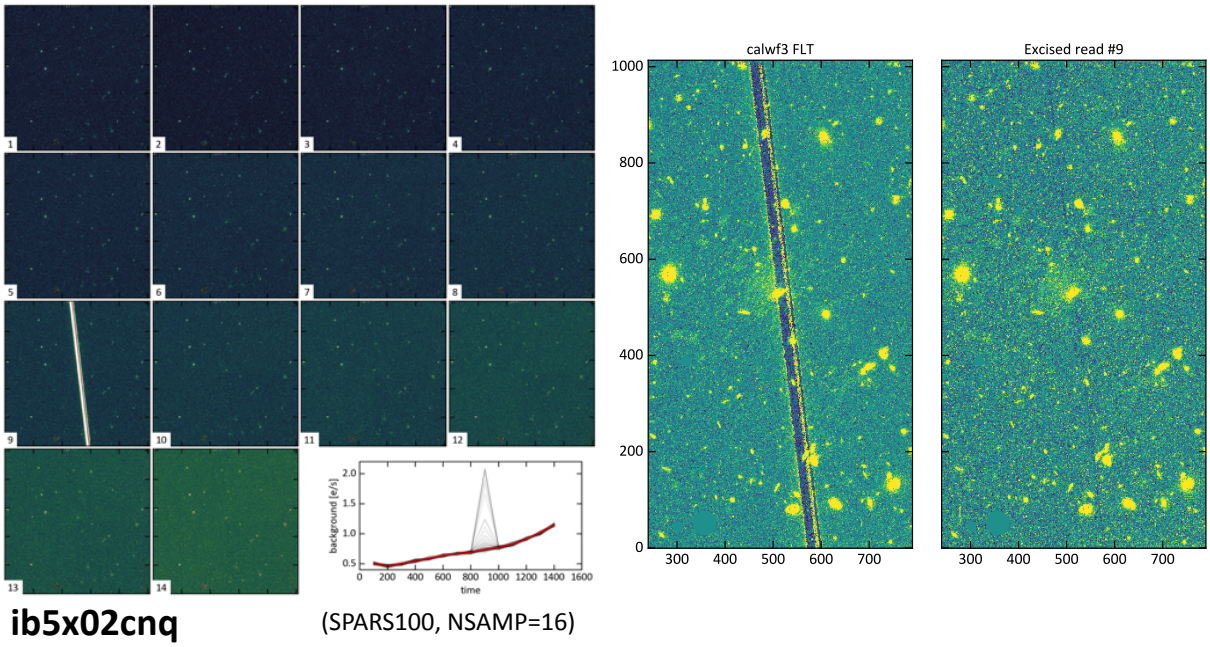

Fig. 6.— Left: ramp diagnostic image showing each read of the MULTIACCUM sequence. A satellite trail flies through on read #9 and only affects that read. Right: Comparison of the `calwf3` pipeline FLT image and the reprocessed calibrated image with the single affected read excised from the sequence as in §5. Note there are some slight negative residuals along the track of the satellite trail, which are likely negative persistence from the bright flux of the satellite.

read times from the total exposure time of the sequence. The final count rate image is then the cleaned difference image divided by the remaining exposure time. The final image shown in Fig. 6 is cleaned of the bright satellite trail, but, as before, the cosmic rays must be identified with, e.g., AstroDrizzle combining additional dithered exposures from the same visit.

5.1. Variable backgrounds and S/N

The signal-to-noise of faint objects in broad-band WFC3/IR exposures is typically background limited. Through a given exposure the combined variance from the background shot noise and detector read noise evolves following Eq. 1. The number of detected electrons of a non-variable source with count rate f grows linearly with time ($S = f \cdot t$), and, therefore the signal-to-noise with *constant* background evolves approximately as S/N $\propto \sqrt{t}$.

Fig. 7 shows the estimated S/N of an arbitrary source throughout the read sequences

of the exposures shown in Fig. 1, where the signal, S_i, and noise, N_i, of read i are computed using

$$\begin{aligned} S_i &= \sum_0^i f \cdot dt_i \\ N_i^2 &= (\sum_0^i b_i \cdot dt_i) + \mathrm{RN}_i^2, \end{aligned} \tag{2}$$

where dt_i is the time elapsed between reads $i-1$ and i and RN_i is the effective read noise of the ith read following Fig. 5.21 of the WFC3 Instrument Handbook.

In the presence of the strongly increasing background, the S/N of exposure `ibp329isq` improves more slowly with time than for the optimal case of constant background. However, even in this example where the background at the end of the sequence was nearly six times the background at the start, the optimal S/N is obtained integrating throughout the entire sequence. In more extreme examples, the S/N curve will turn over such that integrating through an additional reads with high backgrounds actually decreases the S/N. In these cases, one may wish to excise the affected reads with extreme backgrounds using the same technique as that described above to excise reads affected by satellite trails, and this decision can be made objectively by measuring the background ramps as in Fig. 1 and computing Eq. 2 for the exposure read sequence.

6. Summary

We have identified algorithms for reprocessing WFC3/IR exposures with variable backgrounds that violate the assumption of linear pixel count rates made by the default `calwf3` pipeline. The simplest of these algorithms is simply treating the WFC3/IR detector like a normal CCD and measuring the total accumulated charge during the exposure without considering the MULTIACCUM ramp at all. As with CCDs, these images will contain cosmic rays that must be identified by other means. We describe another algorithm to manually flatten the evolution of the background in the IMA files, which allows us to recover the `calwf3` cosmic ray rejection. The last method allows the user to reject individual reads from the middle of the MULTIACCUM sequence that have been affected by satellite trails or other transient phenomena.

Simple code is provided here, and the full code for generating the ramp diagnostic figures as in Fig. 1 (`show_MultiAccum_reads`) and for the reprocessing and excising of individual reads (`make_IMA_FLT`) can be found at https://github.com/gbrammer/wfc3/blob/master/reprocess_wfc3.py. It has been tested to work with the latest astroconda distribution, which provides `calwf3` and other Python tools for *HST* data analysis.

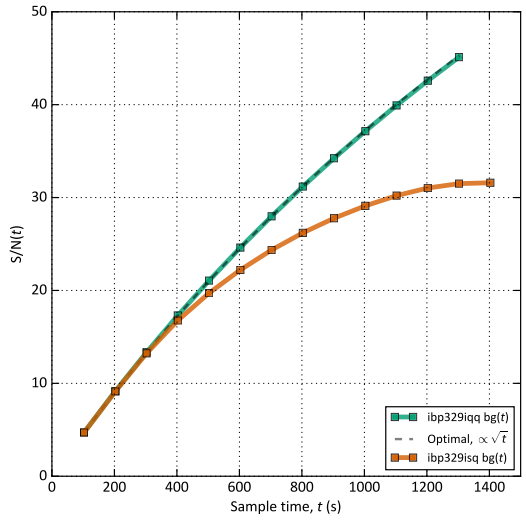

Fig. 7.— Background-limited S/N of a source with count rate $f = 1e^-/s$ for the exposures shown in Fig. 1. For the exposure with constant background, the S/N increases roughly as $\sqrt{t}$. For the exposure where the background increased significantly throughout the read sequence, S/N increases more slowly to the point where including the last 100 s read of the exposure did not improve the S/N at all.

I would like to thank Nor Pirzkal for providing a careful review of this manuscript.

References

Brammer, G. B., et al., 2014, WFC3 ISR 2014-03 "Time-varying Excess Earth-glow Backgrounds in the WFC3/IR Channel" [PDF]

Giavalisco, M., et al., 2002, WFC3 ISR 2002-12, "New Estimates of the Sky Background for the HST Exposure Time Calculator" [PDF]

Gonzaga, S., et al., 2012, "The DrizzlePac Handbook" [URL]

Hilbert, B. & McCullough, P.R.M, 2009, WFC3 ISR 2009-21, "WFC3 SMOV Results: IR Channel Dark Current, Readnoise, and Background Signal" [PDF]

Pirzkal, N., 2014, WFC3 ISR 2014-11, "The Near Infrared Sky Background" [PDF]