The Statistical Properties of Dark Ramps for the Roman-WFI Detectors

Sarah Betti, Timothy Brandt, Stefano Casertano, Richard Cosentino, Rachael Beaton, Javier Sanchez

October 6, 2025

Abstract

We characterize the dark signal in the detectors of the Wide Field Instrument (WFI) on board the Nancy Grace Roman Space Telescope using data from the first and second Thermal Vacuum Integration and Testing (TVAC) campaigns. We calculate the average dark signal rate in each detector and find that it is consistently low ($\approx$0.02 e^-/s) and below the operational requirement for the detectors. We statistically show that the variation in dark signal is stable both up-the-ramp and across exposures; however, many thousands of exposures are needed to measure the signal pixel-by-pixel with substantial signal-to-noise. However, when averaging over millions of pixels, we uncover an exponentially decaying bias over the first few reads, with a typical amplitude of $\approx$0.5 DN (0.8 e^-) and an e-folding decay time constant of $\approx$23 seconds. Finally, we uncover evidence that the read noise is correlated from one read to the next, so that the read noise decreases in long ramps somewhat less than it would for uncorrelated read noise.

1 Introduction

The major aim of the Nancy Grace Roman Space Telescope is to perform the most sensitive and wide-area near infrared observations of the sky through three major surveys. These surveys, the High Latitude Wide Area Survey, the High Latitude Time Domain Survey, and the Galactic Bulge Time Domain Survey, will be carried out using the flagship Wide Field Instrument (WFI), and will require precise photometric measurements at high signal-to-noise, thereby necessitating a robust understanding of the noise properties and structures of the WFI detectors.

Significant work has gone into understanding and characterizing different noise properties and detector performance of the WFI, including measuring the dark current in the Application Specific Integrated Circuit (ASIC) for Control And Digitization of Imagers for Astronomy (Petric et al. 2023, ACADIA), fixing the increase in dark current compared to component level testing that was observed in several detectors (Mosby et al. 2024), and characterizing $1/f$ noise (Betti et al. 2024). Dark signal is one of the low level noise sources

in the 18 detectors (Mosby et al. 2020) that make up the focal plane system in the WFI. It is composed of the internal detector current, thermal background, and residual biases. Dark signal, other than bias offsets, represents accumulating charge and is subject to Poisson shot noise (Mosby et al. 2020).

Due to the stringent requirements needed for the Roman WFI proposed surveys, we need a thorough understanding of the overall dark signal down to 0.05 e^-/s precision in each detector (Mosby et al. 2024). The characterization of the dark signal in WFI ramps is essential to provide a robust means to remove this noise during science observations.

The WFI detectors use non-destructive readouts and record accumulated signal, which includes dark signal, over time, called a ramp. When we use the term "dark signal," we mean the sum of true dark current and external backgrounds, which both accumulate charge on the detector. During the data reduction process, the dark signal will be subtracted off from each resultant in the "dark" stage of the Roman pipeline. The dark calibration reference files utilized in `romancal`, the Roman Space Telescope calibration pipeline for WFI data (Roman Calibration Pipeline Team 2025), to perform this subtraction are created from dedicated dark signal exposures, wherein no illumination is present. The baseline expectation is that individual ramps from multiple exposures will be combined resultant-by-resultant to form a dark cube reference file. Therefore, it is necessary to fully characterize both exposure-to-exposure and resultant-to-resultant variation in the dark signal that will make up these initial reference files.

This document aims to robustly characterize the stability in dark signal across different dark exposures and assess the variability in dark signal up-the-ramp. From here on in, up-the-ramp refers to the signal accumulated on single read resultants, which we refer to as reads or frames. In Section 2, we present the data used for this analysis. In Section 3, we measure the dark signal rate to assess its overall structure. In Section 4, we characterize the stability and variation in dark signal up-the-ramp, while in Section 5, we look at the dark signal across exposures. In Section 6, we estimate the number of exposures needed to produce a robust dark cube reference file. In Section 7, we access the correlated read noise and its effect on ramps. Finally, in Section 8, we summarize our results.

2 Data

To characterize the dark signal, we use datasets from the first and second Thermal Vacuum Testing (TVAC1 and TVAC2) for the WFI instrument from the "Total Noise" (OTP00639) and "Dark" (OTP00644) testing. TVAC consists of the initial tests of WFI with all components included, such as the fight detectors, controllers, and harnesses, with the flight Focal Plane control electronics, and full coupling to other WFI systems like the simplified Relative Calibration System.

The Total Noise dataset in TVAC1 was taken on 21 October 2022 in the Nominal Operations plateau (operations-like environmental plateau). This dataset is comprised of 100 55-frame dark exposures with a frame time of 3.04 s, in which each of the reads (and one reset read) are saved as single read resultants, and a total exposure time of 170.3 s. The Total Noise dataset in TVAC2 was taken on 18 April 2024 in the Nominal Operations plateau, and similarly is comprised of 100 55-frame dark exposures with a frame time of 3.16 s and a

total exposure time of 177.1 s. The frame time increased between TVAC1 and TVAC2 due to a change in the operation of the guide windows. As for TVAC1 data, each of the reads (and one reset read) are saved as single read resultants. The Dark dataset in TVAC2 was taken on 18 April 2024 in the Nominal Operations plateau and is comprised of 4 350-frame dark exposures of 3.16 s each. These dark files are considered "Long Darks" as their total exposure time is ~ 1110 s.

3 Dark Signal Rate

We calculate the dark signal rate for each detector by running a set of two long dark exposures corrected for $1/f$ noise (Betti et al. 2024) through the Roman reference file pipeline `dark` step (Roman Reference File Pipeline Team 2025), wherein the dark exposures are sigma-clipped, stacked, and median combined frame-by-frame, and then collapsed into a rate image. Only two exposures out of four were used due to computational limitations. We show the dark rate signal measured for each detector in Fig. 1. These maps are consistent with dark signal maps found by Mosby et al. (2024) using the semi-flight FPS test data taken in 2022, including the similar detector substructure seen in Mosby et al. (2024). The median dark signal in all detectors is relatively consistent with each other, as shown in Fig. 2, where the median dark signal is 1×10^{-2} DN/s (see Table 1). Using the same average conversion gain used by Petric et al. (2023) of ~ 0.6 DN/electron, we find the average dark signal is 0.018 ± 0.004 e^-/s. This is far below the requirement of a median dark current below 0.05 e^-/s for each detector. Additionally, we find that 99.7% of all pixels meet the per pixel dark signal operability requirement of 0.5 e^-/s. We note the spike at 0 DN/s in Fig. 2 is due to dead and masked pixels.

4 Dark Signal Up-the-Ramp

4.1 Repeatable Dark Signals

We now turn to possible repeatable dark signals in dark ramps. For this section we use the 55-read dark ramps (OPT00639) taken as part of TVAC1 and TVAC2 testing, for a total of 200 exposures. We first process these ramps using the Improved Roman Reference Correction (IRRC, Betti et al. 2024) and then fit a linear accumulation of signal to the 20th through 55th reads. For a given pixel, we define the residuals r_i from these linear fits as

$$r_i = y_i - a - bt_i \tag{1}$$

where y_i is the measured value at read i, t_i is the time of that read, and a and b are constants fitted using $i \geq 20$. The r_i form the basis of our analysis. The goal of this fit and subtraction is to highlight any transient signals that primarily affect the early reads.

As a first step, we exclude all pixels with a root-mean-square (rms) scatter above 10 DN as potential bad pixels over the full ramp. This is a little less than double the median read noise of $\approx$6 DN. We then take the mean of the remaining pixels' residuals read-by-read, further averaging 50 exposures. We use detector WFI17 as an example.

detector	median dark signal (DN/s)	% of pixels with abs(dark signal−median) $< 3\sigma$	% of pixels with dark signal < 0.5 e^-/s
WFI01	0.0100 ± 0.0059	93.42	99.26
WFI02	0.0108 ± 0.0035	97.00	99.74
WFI03	0.0099 ± 0.0065	87.69	99.74
WFI04	0.0102 ± 0.0046	95.22	99.55
WFI05	0.0177 ± 0.0049	96.37	99.63
WFI06	0.0102 ± 0.0048	96.29	99.46
WFI07	0.0143 ± 0.0052	96.02	98.96
WFI08	0.0096 ± 0.0041	95.96	99.79
WFI09	0.0101 ± 0.0035	96.73	99.75
WFI10	0.0067 ± 0.0040	95.67	99.25
WFI11	0.0166 ± 0.0060	97.36	99.74
WFI12	0.0122 ± 0.0048	95.26	99.85
WFI13	0.0071 ± 0.0043	94.72	99.51
WFI14	0.0103 ± 0.0044	95.04	99.67
WFI15	0.0090 ± 0.0034	96.80	99.80
WFI16	0.0100 ± 0.0039	96.23	99.80
WFI17	0.0094 ± 0.0049	94.94	99.85
WFI18	0.0104 ± 0.0052	94.37	99.60
all	0.0108 ± 0.0028	95.28 ± 2.15	99.61 ± 0.24

Table 1: Dark signal statistics for WFI detectors.

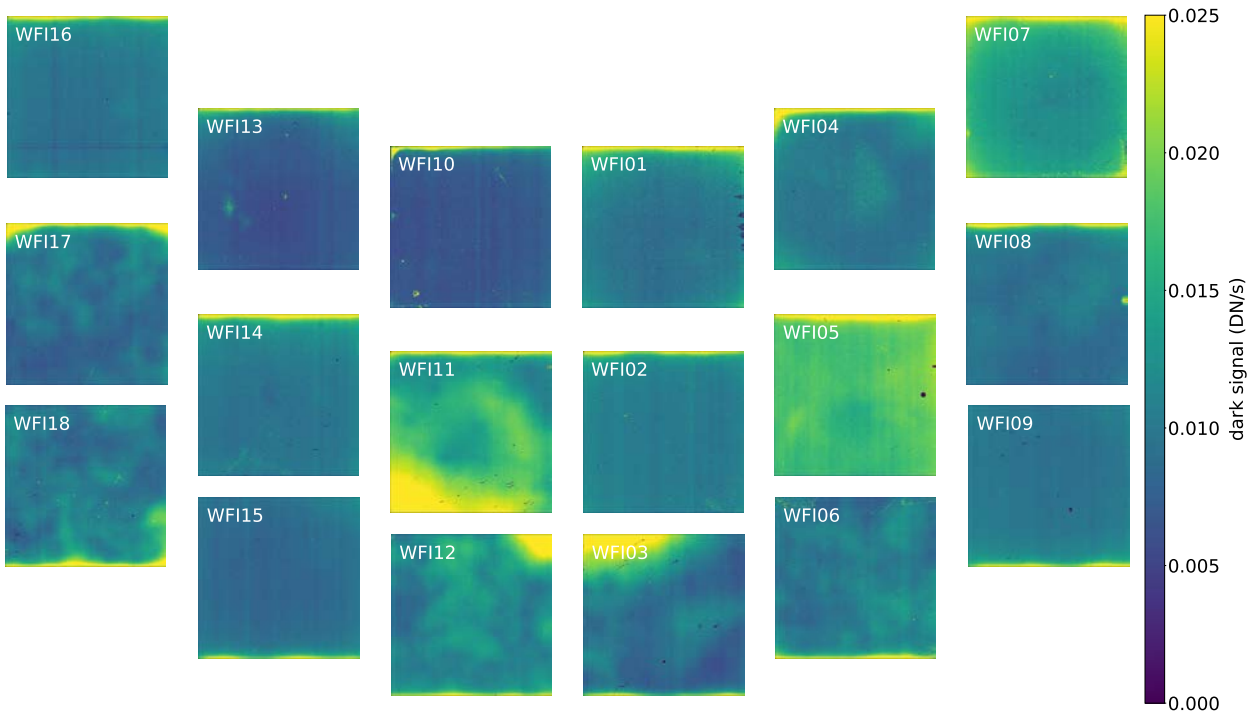

Figure 1: Map of dark signal rate for the WFI detectors using long dark TVAC2 datasets with 1/f noise removed.

Figure 3 shows the results as blue dots. We fit a single exponential to the residuals from Read 2 onward. The residuals are from dark frames after fitting a line to reads 20-55, so we fit and subtract a line from the exponential in the same way. As a result, the average slope of the adjusted exponential is zero between reads 20 and 55, and the line increases by the end of the ramp. This adjustment does not add an additional degree of freedom to the fit. The single exponential provides an excellent description of the measured points; it has an e-folding time constant of about 26 seconds and an amplitude of about 0.7 DN at reset. The first read is slightly discrepant. Future work will address the stability of the first read, especially in light of a possible change in the number of resets performed between exposures. The TVAC data were taken with three resets between frames, while a minimum of four or more so-called idle frames is under discussion at the time of this publication.

We have searched for variations in the behavior shown in Figure 3 with position on the detector. We find no evidence of variation with readout channel. There does appear to be some variation with position on the detector: the signal is slightly larger where the detector is read out earlier. This is consistent with the hypothesis that the signal comes from the readout electronics rather than from the pixels themselves. Figure 4 shows this for WFI17, the same detector plotted in Figure 3. The mean time difference between the top third and bottom third of the detector corresponds to two-thirds of the frame time in Figure 3. This dependence on time since reset will not impact the functional form of the exponential dark decay signal, but will make its amplitude a weak function of a pixel's position on the detector.

Detector	Amplitude (DN)			Decay Time Constant (s)		
	TVAC1	TVAC2	TVAC1+2	TVAC1	TVAC2	TVAC1+2
WFI01	0.13 ± 0.05	0.24 ± 0.06	0.15 ± 0.03	17.7 ± 11.9	29.2 ± 11.8	24.6 ± 10.1
WFI02	0.44 ± 0.03	0.43 ± 0.04	0.43 ± 0.02	25.0 ± 4.2	18.1 ± 3.6	22.1 ± 2.7
WFI03	0.56 ± 0.10	0.43 ± 0.06	0.50 ± 0.07	20.6 ± 8.9	31.6 ± 8.4	23.1 ± 8.2
WFI04	0.34 ± 0.02	0.39 ± 0.03	0.36 ± 0.01	20.4 ± 3.2	18.8 ± 3.8	19.6 ± 2.5
WFI05	0.40 ± 0.03	0.47 ± 0.08	0.42 ± 0.03	26.8 ± 4.4	33.0 ± 9.7	28.8 ± 4.5
WFI06	0.62 ± 0.03	0.61 ± 0.06	0.61 ± 0.02	20.6 ± 2.7	25.5 ± 6.2	22.1 ± 2.5
WFI07	0.23 ± 0.04	0.26 ± 0.06	0.22 ± 0.03	24.8 ± 10.5	15.6 ± 9.8	20.2 ± 7.5
WFI08	0.38 ± 0.05	0.35 ± 0.06	0.36 ± 0.04	34.3 ± 8.1	27.6 ± 7.3	31.4 ± 5.6
WFI09	0.45 ± 0.02	0.57 ± 0.06	0.48 ± 0.02	20.9 ± 2.9	24.3 ± 5.7	21.9 ± 2.8
WFI10	0.14 ± 0.03	0.34 ± 0.07	0.19 ± 0.02	14.2 ± 5.0	32.3 ± 13.0	22.1 ± 7.6
WFI11	0.37 ± 0.04	0.43 ± 0.04	0.37 ± 0.02	29.9 ± 7.5	17.6 ± 5.5	23.3 ± 4.3
WFI12	0.36 ± 0.03	0.46 ± 0.07	0.39 ± 0.03	28.8 ± 5.3	31.7 ± 8.1	29.5 ± 4.5
WFI13	0.25 ± 0.03	0.29 ± 0.05	0.26 ± 0.03	26.4 ± 7.0	30.1 ± 11.4	27.5 ± 6.7
WFI14	0.41 ± 0.02	0.41 ± 0.06	0.40 ± 0.02	24.5 ± 3.6	29.8 ± 9.3	25.4 ± 3.7
WFI15	0.40 ± 0.02	0.47 ± 0.06	0.42 ± 0.02	21.8 ± 3.2	26.1 ± 6.5	23.2 ± 2.9
WFI16	0.35 ± 0.04	0.41 ± 0.04	0.36 ± 0.02	26.6 ± 7.3	18.5 ± 5.6	22.7 ± 4.3
WFI17	0.65 ± 0.03	0.70 ± 0.06	0.67 ± 0.03	24.9 ± 3.0	26.9 ± 5.3	25.2 ± 2.6
WFI18	0.54 ± 0.05	0.53 ± 0.05	0.53 ± 0.04	22.9 ± 4.8	18.6 ± 5.0	21.3 ± 3.4

Table 2: Fitted properties of the dark offset in each detector from TVAC1, TVAC2, and combined data sets.

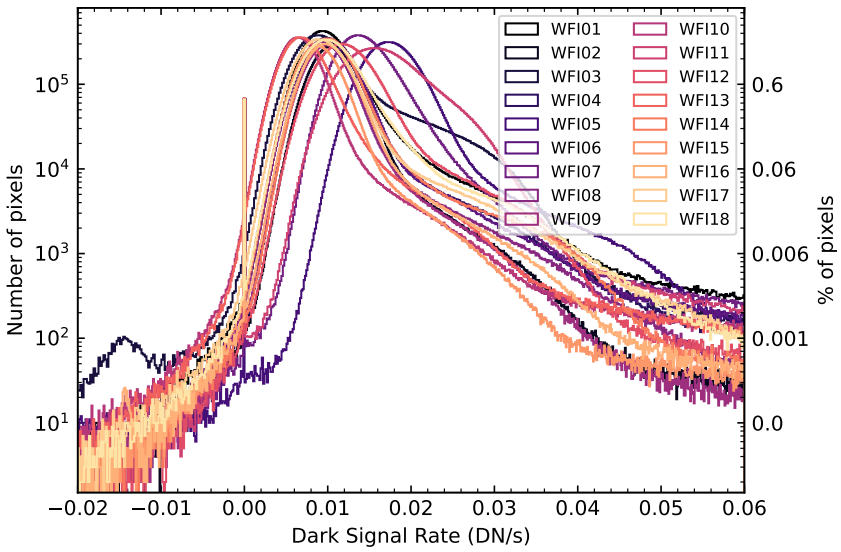

Figure 2: Dark signal rate histogram for the WFI detectors shown in Fig. 1. The average dark signal shifts between detectors but is consistently below the operational requirement. The spike at 0 DN/s in all detectors is due to dead and masked pixels. We note that the Total Noise test can produce ramps with negative slope even with dark signal present over long exposures due to the nature of read noise.

We next perform the same calculations on the rest of the detectors for data sets from both the TVAC1 and TVAC2 test campaigns. We use 100 55-read dark frames from each campaign. We compute the results for TVAC1 and TVAC2 separately, and then for the combined data set of 200 exposures per detector. In all cases we derive means and uncertainties using bootstrap resampling.

Table 2 lists the results of our exponential decay fits for each detector, while Figure 5 provides a visualization of the same data from the combined TVAC1+2 data set. The detectors show differences of a factor of ≈3-4 in the amplitude of the dark offset, with hints that it may be very slightly larger in TVAC2 than in TVAC1. The time constants, however, are all consistent with one another. They have an inverse-variance-weighted mean of 23.2 s, with the residuals from this single value providing a χ^2 of about 11.6 for 17 degrees of freedom (eighteen detectors minus one free parameter). The amplitudes are completely inconsistent with a single value across the entire array; the equivalent χ^2 value for the amplitudes from their weighted mean is nearly 500.

The signal shown in Figure 3, if uncorrected, will impose a small bias on the nonlinearity correction and on the inferred count rates at low illuminations. For certain science applications these effects will be important. This process can likely be tested with illuminated science data to determine if the additive component is unchanged. If the apparently electronic signal remains steady in time and manifests as a residual from a constant signal accumulation rate, it can be continually monitored, and whether through an exponential model fit or through empirical dark cubes from ground testing and/or on-orbit data, `romancal` (Roman

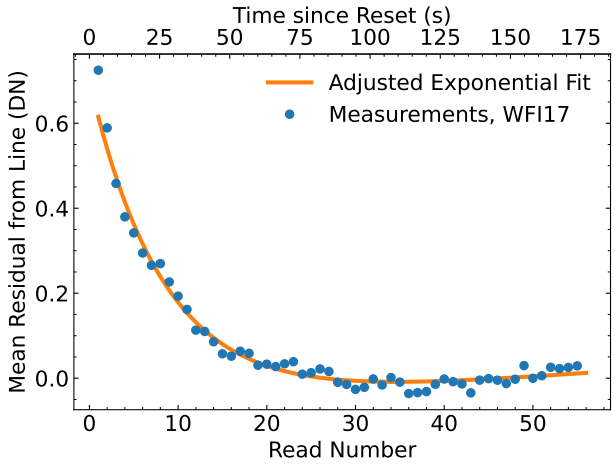

Figure 3: Read-by-read residuals from a linear fit to reads 20-55, averaged over pixels and over 50 exposures. The solid line shows an exponential fit with what would have been the best-fit linear component to reads 20-55 subtracted. Because the slope of the orange line is (on average) adjusted to zero between reads 20 and 55 to match the processing of the blue points, the orange line increases towards the end of the ramp. A single exponential provides a good fit to all but the first read.

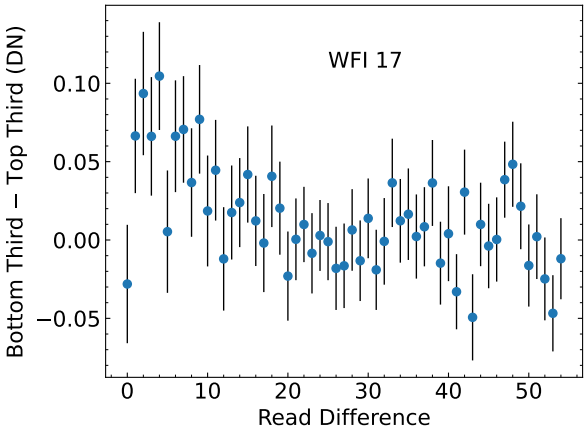

Figure 4: Difference between the mean residual from Figure 3 between the bottom third of the detector (which is read out first) and the top third (which is read out later). There is a 2/3 frame mean delay between reading out the top and bottom of the detector. At early times, Figure 3 shows a ≈0.1 DN offset difference between reads. That suggests that we should see 2/3 as much, i.e. a ≈0.07 DN top-bottom difference in this figure, if the signal is due to the readout electronics. This ≈0.07 DN top-bottom offset at early times is indeed what we observe.

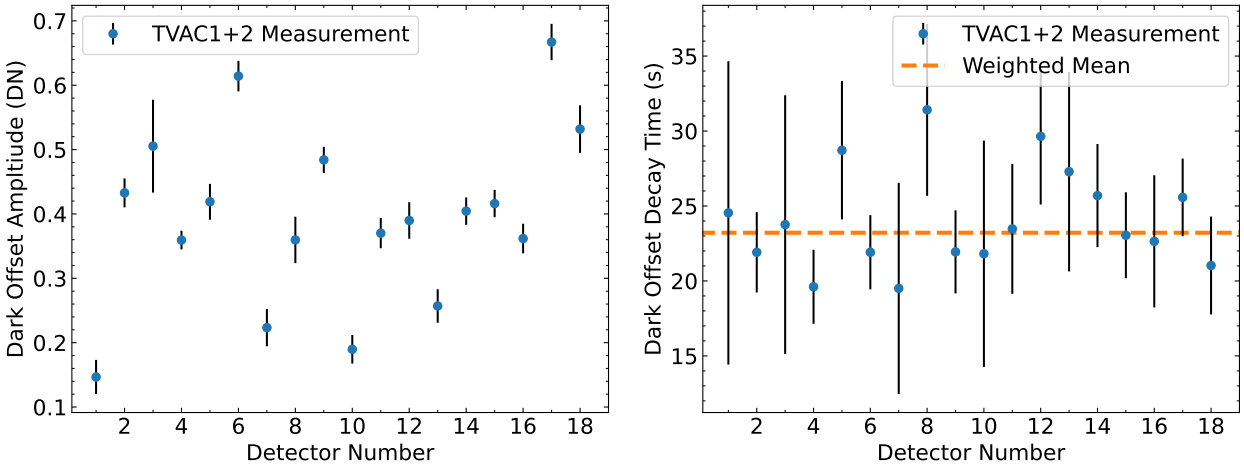

Figure 5: Visualization of the exponential decay properties of the bias in dark frames (Table 2) using the combined TVAC1+2 data set, with 100 55-read dark frames each from TVAC1 and TVAC2. The uncertainties are derived from bootstrap resampling. The decay times are all consistent with a value of about 23 seconds, with $\chi^2 \approx 11.6$ for 17 degrees of freedom (18 detectors minus one fitted quantity).

Calibration Pipeline Team 2025) processing will then likely need to account for it.

4.2 Linearity of the dark signal in long exposures

The previous section tested the stability and repeatability of the dark signal in the beginning of an exposure (up to $\sim 150s$). We find that the mean residuals from linear signal accumulation follow an exponential decay with an e-folding time constant of ≈23 seconds (≈7 frames), indicating that there is some offset from a linear modeled ramp that is most pronounced in the first frames.

We now look at the end of the long dark exposures, when the dark signal has significantly accumulated. We aim to determine if the mean residuals and linear fit to the data are sufficient to describe the ramp over a long accumulation of signal. We calculate the slope up the ramp using: (1) a full linear model across the whole ramp, and (2) a rolling slope calculated up the ramp using a ~ 200 s (65 frame) rolling window. The rolling slope is found using ordinary least squares and solving for the best fit slope and intercept within the 65 frame window. In the left column of Fig. 6, we show the ramps for four types of pixels identified: (a) linear good pixels, (c) non-linear good pixels, (e) telegraph pixels, and (g) RC pixels[1], from top to bottom. The black line shows the best fit linear model to each pixel. In the right column, we show the best fit slope for the 65 frame window rolling fit for each type of pixel. As evident visually, panels (f) and (h) show that a linear model does not accurately fit the telegraph and RC pixels, unlike panel (b) which clearly shows how a linear fit is well fit up-the-ramp. These bad pixels will be masked and not used in further analysis. In panels c) and d), we show that a linear fit through the whole ramp will underestimate the rate by

[1]RC pixels show an exponentially decaying transient signal, similar to the behavior of an RC circuit. They are believed to be due to equilibrating charge traps.

16% (for this pixel) due to the decrease in counts at t$\sim 900-1000$s. We find that the rolling slope differs from the linear slope by over 3σ by $t = 1000$s.

We then determine if these "non-linear" good pixels are prevalent in the long dark exposures in order to understand their effect on measuring accurate dark signal using rate images. After masking out known bad pixels, we calculate the rolling slope for all pixels within the central 100×100 pixel region[2]. Table 3 gives the percent of good pixels flagged as "non-linear" in the central region. We find, on average, $< 0.5\%$ of pixels within the central region show this feature.

Finally, as the above analysis was performed using only one exposure, we determine the "non-linear" good pixels for all four long dark exposures for WFI06 to see if the "non-linear" pixels are static or random across exposures. As shown in Fig. 7 and Table 4, while the percentage of pixels flagged as "non-linear" is relatively constant ($\sim 0.36\%$ in the central region), the location of the flagged pixels is random, indicating this class of pixels is small and variable. They may simply represent unfortunate realizations of read+photon noise. Our 3σ threshold is expected to have a false positive rate of 0.27%, similar to rate of flagging pixels for nonlinearity.

Overall, the "non-linear" pixels could overestimate the dark signal on a per-pixel basis by a small amount; however, this "non-linearity" was only seen for t$>$ 900s for all flagged pixels across all detectors. Therefore, for shorter exposures, such as those that will be observed for the four main WFI surveys, this will likely not be a major concern, and a linear fit will accurately model the ramp.

5 Dark Signal across Exposures

5.1 Reproducibility of the Dark Signal

As a dark cube is key for the subtraction of dark signal in the data reduction process, it is essential to understand the underlying noise properties and frame-to-frame effects of dark exposures. In order to investigate these effects, we measure the observed signal in the read (frame) differences up-the-ramps in the Total Noise data. By looking at the frame differences up-the-ramp, we are able to measure variation in the underlying noise structure.

To measure the noise in individual Total Noise exposures, we first perform a reference pixel correction as outlined in Betti et al. (2024) to remove $1/f$ noise. We then mask bad pixels using bad pixel masks[3]. Then, for each pixel, we determine the ramp, found by fitting a line to the signal at each read as a function of frame's integration time. We subtract off the best fit ramp to determine the frame-by-frame residuals, as

$$\Delta y_i = y_i - (y_0 + mt_i), \tag{2}$$

where y is some pixel, i is the frame index, m is the slope, and t is the time measured at each frame. Finally, for each pixel, we compute the residual difference between consecutive

[2]As this experiment is computationally expensive, we can only look at a subset of the detector at a time. However, performing this same experiment over other portions of the detector find similar results.

[3]taken from `roman_wfi_0023.imap` at https://roman-crds-tvac.stsci.edu

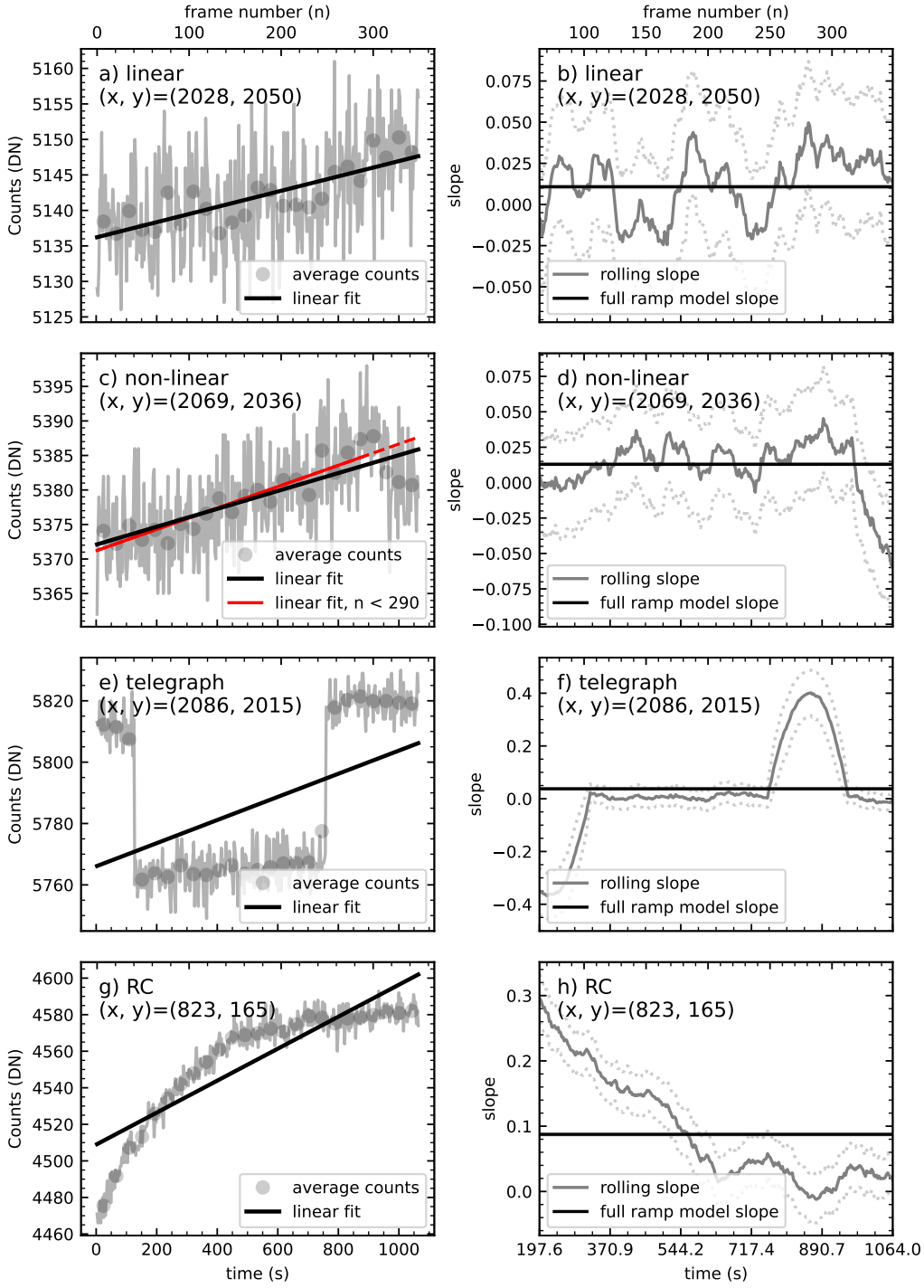

Figure 6: Types of pixels and their linear fits. Each row shows an example of a different type of pixel, with its accumulated signal as a function of time and best linear fit on the left and the rolling slope value up the-the-ramp on the right. **a-b):** An example of a regular linear pixel. **c-d):** An example of a non-linear regular pixel with a drop in counts after $t > 900$s. The red line on the left panel shows the best linear fit for only the first 290 frames. **e-f):** An example of a telegraph pixel. **g-h):** An example of an RC pixel.

detector	% flagged in central region
WFI01	0.27
WFI02	0.80
WFI03	0.14
WFI04	0.18
WFI05	0.77
WFI06	0.42
WFI07	0.53
WFI08	0.33
WFI09	0.65
WFI10	0.30
WFI11	0.55
WFI12	0.40
WFI13	0.14
WFI14	0.58
WFI15	0.46
WFI16	0.39
WFI17	0.38
WFI18	0.58
all	0.43 ± 0.19

Table 3: The percentage of regular pixels flagged with a nonlinear slope up-the-ramp after $t > 900$s for the central 100×100 pixel region of each detector.

exposure	% flagged in central region
2024-04-18 06:52:50	0.42
2024-04-18 07:11:30	0.36
2024-04-18 07:30:09	0.30
2024-04-18 07:48:52	0.38

Table 4: The percentage of regular pixels flagged with a nonlinear slope up-the-ramp after $t > 900$s for the central 100×100 pixel region for each exposure taken for WFI06.

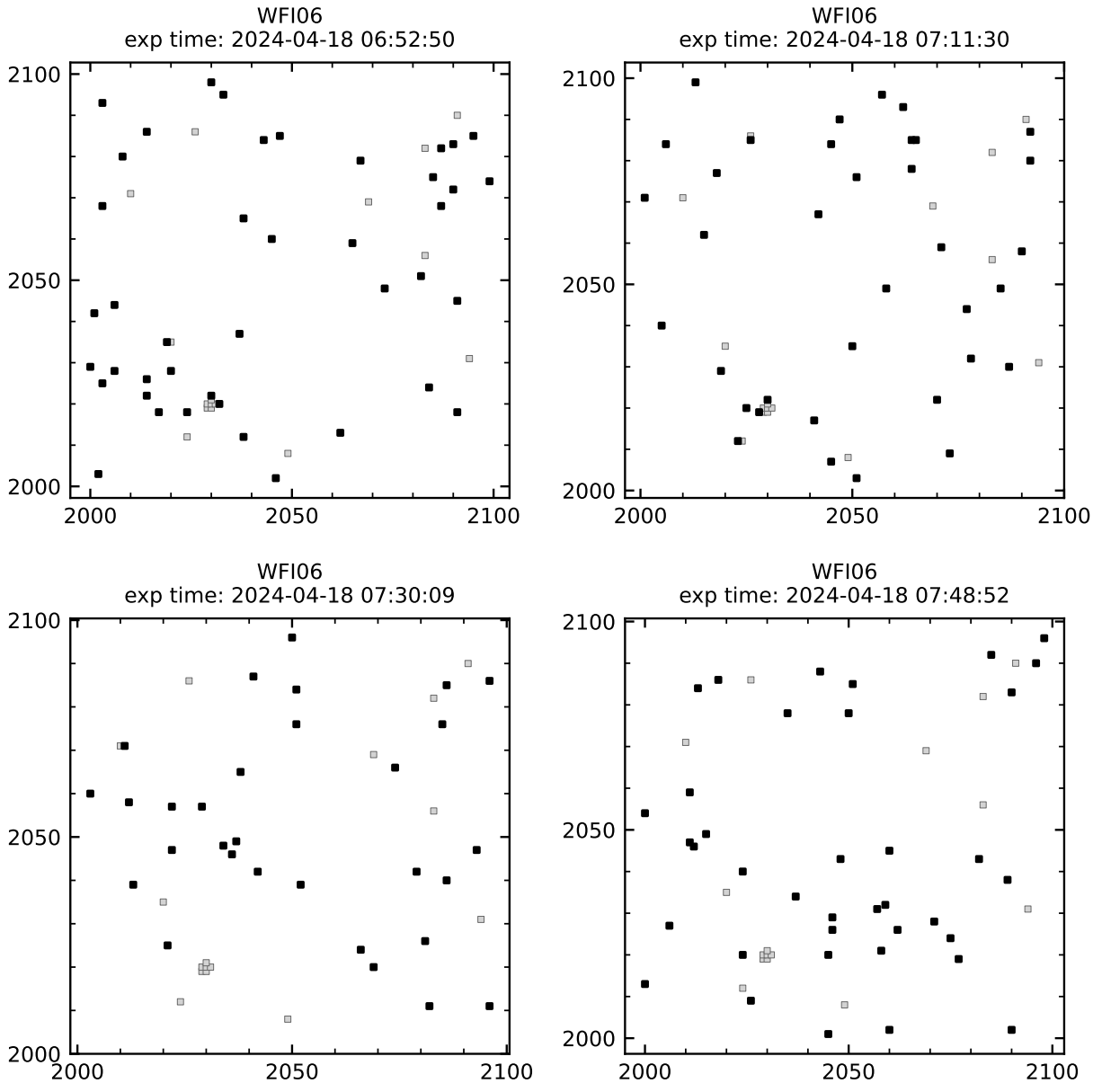

Figure 7: The spatial distribution of regular pixels (black) flagged with a nonlinear slope up-the-ramp after $t > 900$s for the central 100×100 pixel region for each exposure for WFI06. The masked bad pixels are shown in gray.

reads as,

$$\Delta y_{i,1} = \Delta y_{i+1} - \Delta y_i. \tag{3}$$

To quantify the level of deviation in residual signal, we first determine the reproducibility in residuals across exposures. Using our residual frame differences, we extract the Frame 3 - Frame 2 read for all 100 exposures in the Total Noise 55 single read resultant dataset for each detector. We determine the distribution and normality in the population of frame differences for each pixel across the 100 exposures using the Shapiro-Wilk Gaussianity test with a p-value of 5%. We show the distribution of residual pixels in Fig. 8 for a stable pixel with Gaussian residuals and for an unstable pixel. This tests the null hypothesis that the sample comes from a normally distributed population. For a p-value < 0.05, we can reject the null hypothesis, which indicates the residual frame differences are not stable exposure-to-exposure.

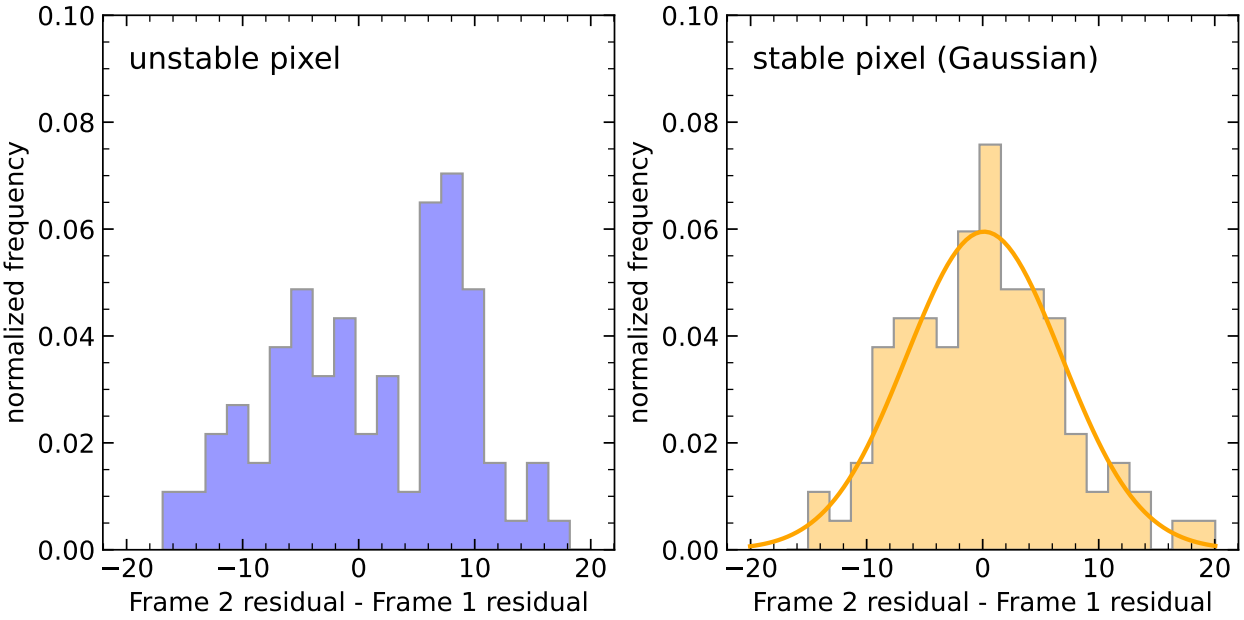

Figure 8: Distribution of residual difference between frames in 100 exposures for a pixel with an unstable pixel (left panel) and a normally distributed and stable population (right panel) in WFI02. From this population, we perform a Shapiro-Wilk test to measure the normality of the pixel residuals across exposures.

We then apply this procedure to all 18 detectors to search for clusters of unstable pixels. We find that $\sim 5\%$ of the pixels are unstable/reject the null hypothesis. Given our p-value threshold, this is at the level expected if the vast majority of our pixels have residuals that *are* normally distributed. Then, we calculate the average nearest neighbor ratio to quantify the spatial correlation. To do this, we calculate the ratio of the observed mean distance between each unstable pixel and the nearest unstable pixel and the expected mean distance for each pixel assuming it is randomly distributed, which in the continuum limit is given by:

$$ANN = \frac{\Sigma_{i=1}^{n} d_i/n}{0.5/\sqrt{n/A}}, \tag{4}$$

where d_i is the distance between pixels, n is the total number of unstable pixels, and A is the total area over which n pixels are distributed. A ratio < 1 indicates pixel clustering, while ~ 1 is consistent with a random distribution. Small corrections of a few percent are needed because of position discretizations at small separations and edge effects. As shown in Table 5, we find an average ANN of 1.047 ± 0.004, roughly consistent with a random distribution of unstable pixels across the detector, with no clustering. This is again consistent with the hypothesis that the vast majority of these pixels had unfortunate (but Gaussian) noise realizations that gave them low p-values by chance.

Finally, we account for sampling error and random sample-to-sample variation when using all 100 exposures. Some pixels will be flagged by chance even if the errors are all Gaussian; we wish to statistically correct for this. We retest our assumption of Gaussianity using our threshold of 5% by splitting our exposures into two sets of independent data. If all pixels have Gaussian noise, then 5% will fail the test at $p = 0.05$ significance, but this 5% will be random. Since we have two independent data sets, we expect 5% of 5%, or 0.25%, to be flagged in both subsets. However, we find that $\sim 0.34\%$ of the pixels across all detectors reject the null hypothesis across both subsamples, slightly more than the 0.25% expected for purely Gaussian errors. We subtract 0.25% from all jointly flagged pixel fractions to correct for random flagging. Our results indicate that the residual difference between frames 3 and 2 is not fully Gaussian across exposures, but that $\approx$0.1% of pixels have noise properties inconsistent with normality.

In Table 5, we list the percentage of pixels that are unstable by our metric, along with the percent of pixels after correcting for the 0.25% expected by chance. Overall, $< 0.15\%$ of pixels per detector can be considered unstable and vary over the 100 exposures. Most of the pixels that do not have consistent residuals appear random, with no set pattern or region prone to unstable pixels, consistent with their origin by chance noise realizations. We do see small clusters in WFI06 and WFI08; however, those groups align with known areas of bad pixels.

We next look to see if there are any pixels or groups of pixels whose residual frame differences remain consistently high exposure to exposure, which could indicate unstable dark signal in that pixel, as the Frame 3 pixel value would always be consistently larger than the Frame 2 pixel value. For each pixel, we determine the number of exposures where the residual frame difference count is >15 DN ($\sim 2.5\times$ median read noise). We first look at WFI08 and find that less than 1% of the pixels have more than 15 DN out of the 100 exposures with high residual frame counts; fewer than 1000 pixels exceeded the 15 DN pixel difference in more than 50 exposures. Spatially, we find that these high residual frame count pixels are random, with no groups of bad pixels. We then apply this to all of the detectors, as shown in Fig 9, and find similar proportions of high residual pixels. For example, we find that ramps with $\sim 1\%$ of high residual pixels are on average only found in 20 out of the 100 exposures. This indicates $> 99\%$ of all pixels are stable across frames and up-the-ramp, as low residual differences result from pixels consistent with a linear fit frame-to-frame.

5.2 Stability of the Dark Signal

In this section, we discuss frame-to-frame effects and their impact on the stability of the dark signal across exposure, expanding upon Section 4.1. Using the procedure from Section 4.1,

detector	ANN	percent of pixels with $p < 0.05$	percent of pixels with $p < 0.05$ corrected for chance
WFI01	1.040	0.58	0.33
WFI02	1.048	0.34	0.09
WFI03	1.050	0.38	0.13
WFI04	1.043	0.39	0.14
WFI05	1.045	0.37	0.12
WFI06	1.045	0.40	0.15
WFI07	1.038	0.50	0.25
WFI08	1.050	0.42	0.17
WFI09	1.051	0.34	0.09
WFI10	1.047	0.54	0.29
WFI11	1.048	0.36	0.11
WFI12	1.051	0.31	0.06
WFI13	1.048	0.39	0.14
WFI14	1.049	0.36	0.11
WFI15	1.052	0.37	0.12
WFI16	1.051	0.38	0.13
WFI17	1.051	0.36	0.11
WFI18	1.050	0.36	0.11

Table 5: Average nearest-neighbor ratio of unstable pixels along with the percent of pixels unfavorable noise properties (p-values < 0.05 in two independence data sets) found using the Shapiro-Wilk test. The last column is found by subtracting 0.25% – probability of pixels with p-values < 0.05 due to chance – from the total percent of pixels.

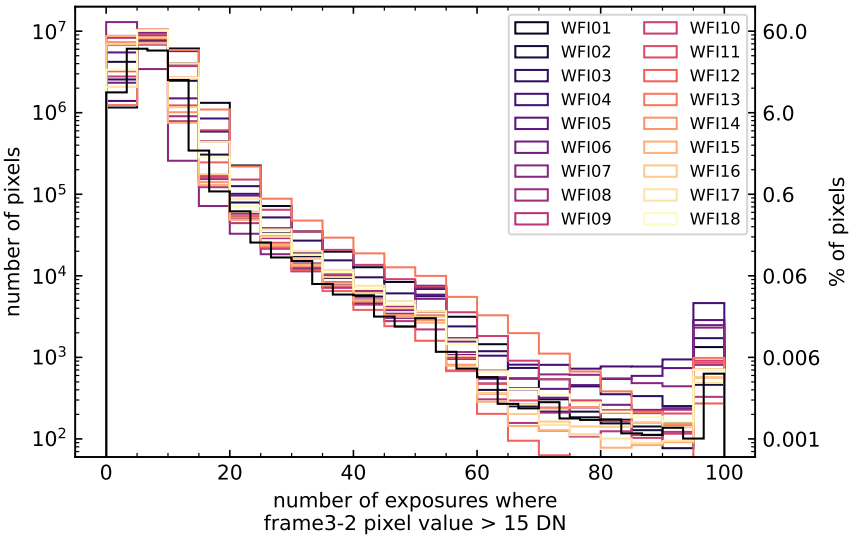

Figure 9: Histogram of the number of pixels that remain consistently high as a function of added exposures for all 18 detectors. 99% of pixels have less than 20 exposures that are consistently high and therefore have varying pixel counts up-the-ramp.

we calculate the average residual from the line up-the-ramp for 100 exposures in the 55-read dark frames in the TVAC1 dataset (OPT00639).

After removing the exponential decay term due to a change in the bias found in Section 4.1 as well as potential bad pixels described above, we fit the linear accumulation of signal up-the-ramp to each pixel. We then subtract off this best fit from each pixel to give the residual counts per frame, allowing us to variation in the underlying noise properties. For each exposure, we determine the mean of the pixels' residuals read-by-read. Finally, we repeat this procedure for all 100 exposures in dataset. The distribution of these average counts per frame highlights any change in signals over the duration of the exposure(s).

The top row of Fig. 10 shows the distributions in the first four frames for WFI09. We find that the distributions are largely centered around 0 DN, as expected due to the removal of the exponential decay. By looking at each distribution up-the-ramp, we find a standard deviation of ~ 0.09 DN for each frame. Expanding to all detectors, we find that the distributions for each frame are consistent within 1% of each other, with an average standard deviation across detectors of $\sim 0.1 \pm 0.01$ DN.

We then look to see how the residual counts across exposure vary at different parts of the ramp. We find the average all the residual pixel counts within 10 frames for each exposure. As shown in the bottom row of Fig. 10, we look at distribution of residual counts for Frame 1 up to Frame 10, Frames 16 to 25, Frames 31 to 40, and Frames 46 to 55, left to right, for WFI06. In this case, the residual means are stable to 0.026 DN, calculated from the standard deviation of the individual distribution standard deviations, and consistent with expectation that 0.09 DN $/\sqrt{10} \sim 0.028$. All together, these distributions show that the average frame residual pixel counts follow the results found in Section 4.1. We find residual counts up-the-ramp that are within 1% of each other across detectors and these counts remain stable up

to 0.07 DN across exposures.

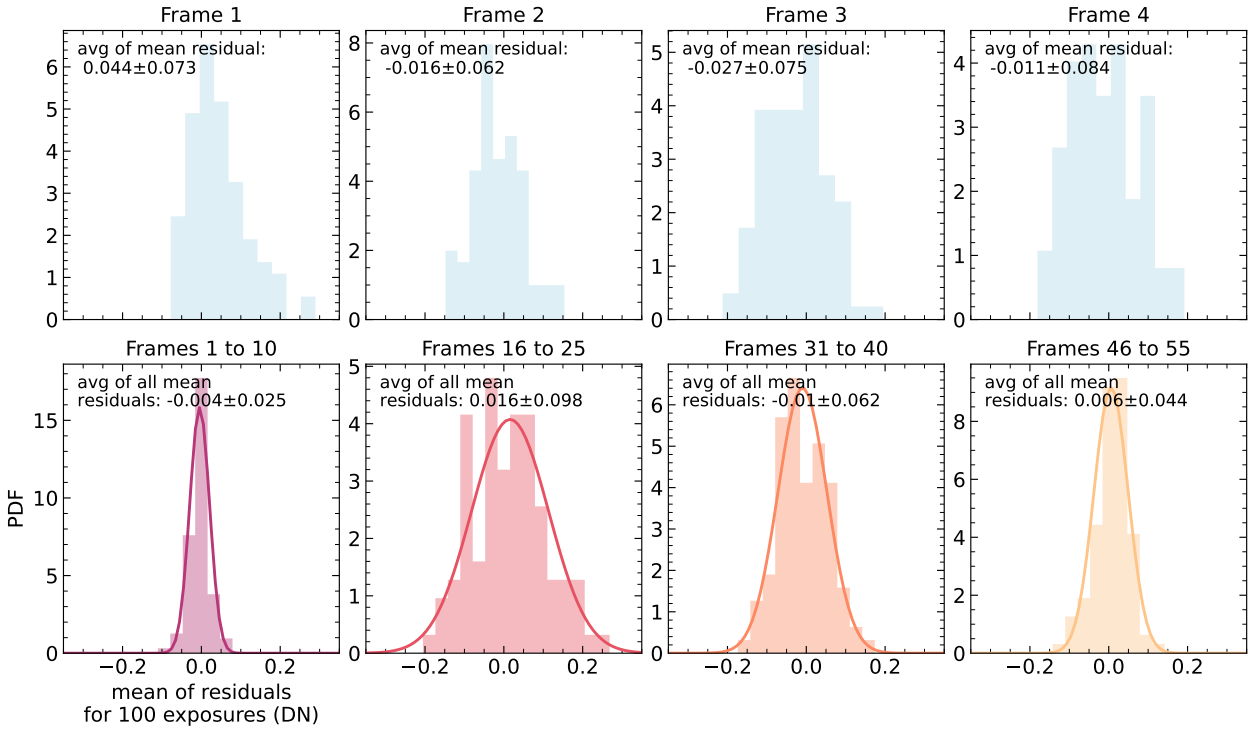

Figure 10: Distribution of the average frame residual counts over 100 exposures. **top row:** Individual frame residuals for Frame 1, 2, 3, and 4, left to right respectively. **bottom row:** Average residuals over all pixels in 10 frame: 1 to 10, 16 to 25, 31 to 40, and 46 to 55, left to right respectively. A gaussian fit is overlaid.

6 Detectability of Dark Signal

Currently, the dark current reference file for the WFI detectors is calculated on a per-resultant basis, and will be used to subtract the L1 science data on a pixel-by-pixel and resultant-by-resultant basis. The dark reference file is produced, in essence, by combining multiple dark exposures frame-by-frame to create a super dark cube. This cube can then be collapsed into a dark rate image. Therefore, we assess the number of exposures needed to measure the dark signal up-the-ramp for an individual pixel. This includes signal that is stable from exposure-to-exposure in a given read but differs from a linear accumulation of signal with time. We showed in Section 3 that the dark signal in a rate image is both easily detectable and below the required threshold for operability.

To determine the threshold for the number of exposures, we calculate the cross correlation in the read difference across ramps. We define the read difference $d_{i,n}$ as the difference between reads n and $n+1$ in ramp i. The quantity $d_{i,n}$ is thus defined for every pixel in a ramp at a given read; expectation values are taken over pixels and may also be taken over

ramps. We define our cross correlation metric as

$$\text{corr} = \frac{\langle d_{i,n} d_{j,n} \rangle}{\sqrt{\langle d_{i,n}^2 \rangle \langle d_{j,n}^2 \rangle}} \text{ for } i \neq j. \tag{5}$$

A corr = 1 indicates the dark signal is perfectly correlated across ramps and can be completely removed by a dark cube. A corr = 0 means that the dark signal is perfectly uncorrelated across ramps; there is no repeatable signal that could be captured by a dark cube. In this case the dark signal behaves as read noise.

We compute cross-ramp correlation coefficients (Equation (5)) using 55-read dark ramps. We divide the full set of ramps into pairs i, j. We then compute, e.g., the expectation of the first read difference in ramp i times that in ramp j for the numerator of Equation (5). The denominator is computed from the variance of the same read differences within ramps i and j separately. The numerator gives the square of the repeatable dark signal, while the denominator gives the read noise variance assuming a very low dark accumulation rate. Taking Figure 5 as an example, a repeatable difference of 0.1 DN between reads with a read noise of 5 DN would yield a value corr $\approx 4 \times 10^{-4}$.

Fig. 11a shows the correlation coefficient as a function of the read difference number for WFI04. Besides Frame 1-0 (read difference number = 0), which has a small transient in darks, the correlation for all reads is ≈ 0, indicating the dark signal is nearly uncorrelated across ramps. Then, we can determine the number of dark exposures needed for a dark cube that is detected at 3σ as $N \sim 9/\text{corr}$. In Fig. 11b, we plot the number of ramps for a 3σ dark cube as a function of read difference number. For this test, we find we need $\sim 10^5$ exposures to produce a dark cube where the signal is detected per-frame at 3σ. However, we note that this test is based on TVAC2 data and would need to be re-evaluated during commissioning to determine if a smaller number of exposures or a lower S/N is sufficient to detect the dark current on the resultant-level. Finally, we note that Fig. 11a is entirely consistent with Figures 3-5. The exponentially decaying dark signal has a difference of $\lesssim$0.1 DN between reads, decreasing with frame number, yielding a corr value (Equation (5)) that rapidly falls below 10^{-4}.

Our analysis shows that we are unlikely to obtain sufficient data to measure the electronic effect identified in Section 4.1 with high signal-to-noise ratio on a single pixel. A dark for each resultant would be noise-dominated at the level of individual pixels, though it could still be valuable if it removes the systematic offset identified there. The only way to construct a resultant-by-resultant dark that is not read noise dominated would be to fit a parametric model, e.g., one in which the bias decays exponentially with a fixed initial amplitude per detector.

7 Correlated Read Noise

The most important source of noise in all but the longest dark exposures is read noise. The most favorable scenario for read noise, and the one typically assumed when fitting rates to ramps, is that it is Gaussian and uncorrelated from read to read. Correlated noise will not average down as quickly, and can increase the effective noise on the slope measurement in

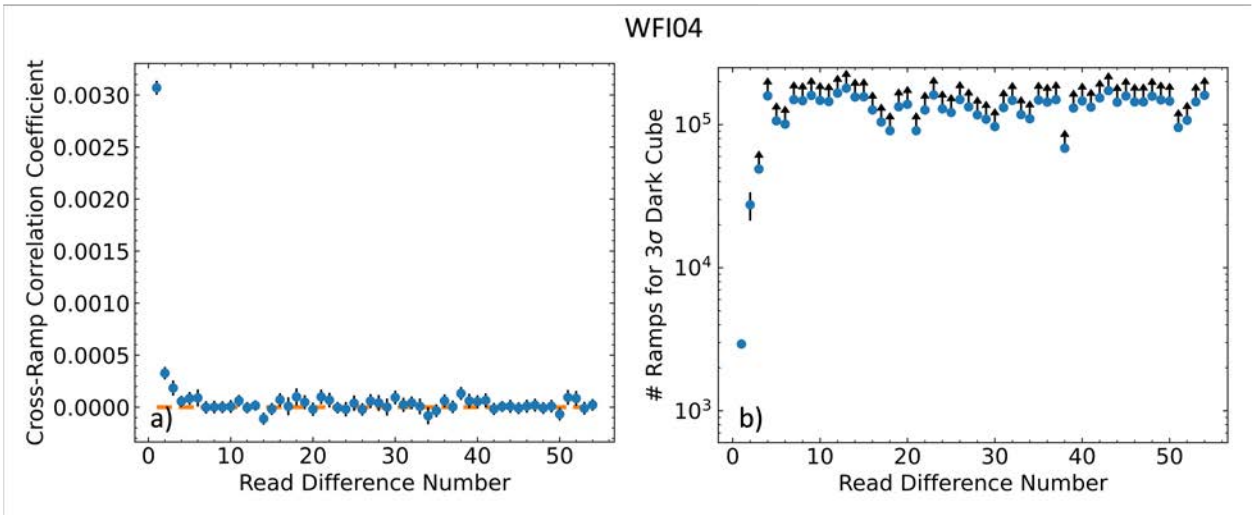

Figure 11: **a):** Correlation coefficients for the read difference number for WFI04. **b):** The number of ramps to detect the dark signal at each frame at 3σ in WFI04.

long ramps compared to uncorrelated noise. In this section we test the null hypothesis of Gaussian, uncorrelated read noise. We quantify the read-to-read correlations of read noise, not the pixel-to-pixel correlations that are due to $1/f$ noise and partially removed by the IRRC.

We base our analysis on 350-read dark frames after applying the IRRC. Our basic procedure is to fit a line to the accumulated counts in every pixel, to compute the residuals r_i from this linear fit for read i, and to compute the variances

$$V_0 = \langle r^2 \rangle \tag{6}$$

and

$$V_i = \langle (r_j - r_{j+i})^2 \rangle \quad \text{for} \quad i \geq 1 \tag{7}$$

where the expectation value is taken over pixels and, in Equation (6), also over reads. If the only noise source is uncorrelated read noise, the variance V_0 will be the read noise variance and $V_i = 2V_0$ for $i \geq 1$.

Because we used the data themselves to fit for a line before computing residuals, neither the statement that V_0 is the read noise variance nor that $V_i = 2V_0$ is strictly true. We calibrate them empirically using synthetic data with only read and photon noise. To make the synthetic data, we use read noise, a dark count rate, and a gain representative of the detector WFI06. We then compute the percentage difference between $V_i/2$ and V_0 as

$$100 \times \left(1 - \sqrt{V_i/(2V_0)}\right). \tag{8}$$

If the read noise is positively correlated between frames, V_i will generally be less than $2V_0$, so larger values of this metric indicate more correlated read noise.

Figure 12 shows our results for two 350-read dark frames on detector WFI06, for the difference of these two frames, and for a synthetic 350-read ramp with similar read noise

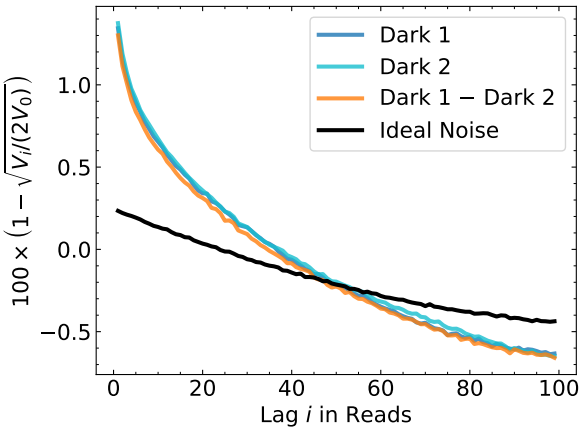

Figure 12: Read noise correlation measure, as defined by Equation (8), for 350-read dark ramps. Larger values indicate more correlated noise. The ideal noise curve is not flat, as photon noise will induce correlations in the noise. Our measurement of a larger-than-expected value at a lag of zero indicates that the detector's read noise is correlated from one read to the next. The close match between the dark frames and the difference of two dark frames shows that little of this is due to the repeatable dark signals discussed in Section 4. The "Dark 1" and "Dark 2" lines are indistinguishable on the figure.

and photon noise levels, but with ideal noise properties. The actual dark frames show a significantly larger signal with a small lag in reads than the ideal synthetic data, i.e., V_i increases with i. This indicates that the read noise is correlated from one read to the next. Most of the noise is uncorrelated; V_i differs from $2V_0$ by just $\sim$1%. For short ramps, the noise on a slope measurement will decrease nearly as quickly as it would for ideal, uncorrelated noise. As the uncorrelated part of the noise averages down in a longer ramp, the correlated part that does not average down as fast will become increasingly prominent. As a result, the correlation shown in Figure 12 will have little effect on short ramps, where uncorrelated noise dominates, but will significantly affect the noise properties of long ramps with very low count rates. This is quantified later in this section and shown in Figure 13.

Section 4 uncovered evidence of a small, repeatable dark signal that appeared similarly to an exponential decay in the bias. Such a signal could impact V_i/V_0. To remove this possibility, we take the difference between two unilluminated ramps read-by-read and compute the correlation metric V_i/V_0 on this difference ramp. The statistical properties of the noise will be nearly identical, while any repeatable signal should vanish. The resulting metric, shown by the orange line in Figure 12, is nearly identical to that seen in either 350-read dark individually, showing that this read noise correlation is not due to repeatable dark signals. Instead, it appears to represent a fundamental limitation of the detector.

Having found evidence that the read noise is correlated between reads, we test whether this correlated read noise has a spatial correlation analogous to the $1/f$ noise. We measure the noise in the read-by-read difference between two 350-read dark frames. By taking the difference between these frames, we remove any stable dark signal while retaining the noise properties (both read noise and photon noise) in an individual frame. We derive count rates

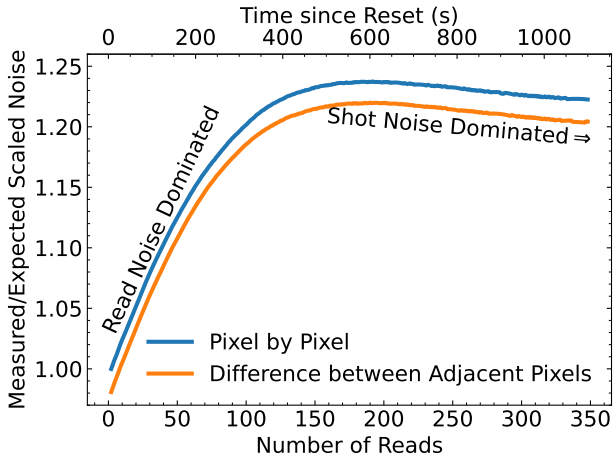

Figure 13: Scaling of the noise, computed as the root-mean-square scatter of derived slopes, compared to the noise expected from ideal read and photon noise. We compute noise using the standard deviation of the derived count rates in the read-by-read difference between two unilluminated frames. The blue curve measures the scatter directly, while the orange curve measures the scatter in the slope differences between adjacent pixels to remove spatially correlated noise. For long ramps, shot noise becomes important and the ratio of measured to expected noise begins to fall.

from the reads using the coefficients appropriate in the read noise limit. We then compute the standard deviation of these count rates across the detector, limited to well-behaved pixels (those with root-mean-square scatters of all of their reads below 14 DN). This threshold retains 99% of pixels while giving a clean measurement.

In the limit of read and photon noise, the scaling of the scatter of count rates may be computed from the read noise and the typical count rate. We assume a characteristic gain of 1.8 e^-/DN and use the typical dark count rate of 0.03 DN/read to compute the expected noise. For the read noise, we adopt the scatter of the slope computed from just two reads divided by $\sqrt{2}$. The actual count rate and differs pixel-by-pixel, so this theoretical noise value is only approximate.

Figure 13 shows the ratio of the noise computed as the root-mean-square scatter of derived slopes to the value expected from ideal read and photon noise. The orange line shows the scatter of the slope *differences* between adjacent pixels divided by $\sqrt{2}$. It is smaller due to the fact that some of the read noise is spatially correlated and does not appear in the difference between adjacent pixels. The observed noise decreases more slowly than expected in the read noise limit, before finally falling towards the expected scaling at long exposure times. Figure 13 does not extend to sufficiently long times to be shot-noise-limited for all pixels, while count rates in dark frames are low enough that correlated read noise remains significant even in very long ramps.

The difference between the orange and blue lines in Figure 13 arises because of the removal of some of the spatially correlated noise. This spatially correlated noise is residual $1/f$ noise remaining after applying the IRRC. If the read noise that is correlated from read-to-read is *also* spatially correlated, the orange line will tack closer to the ideal behavior than the

blue line. However, the difference between the orange and blue lines is nearly independent of the number of reads. This is consistent with the hypothesis that the great majority of the temporally correlated noise shown in Figure 12 is not spatially correlated and remains even when taking the difference between adjacent pixels. The figure looks visually identical, and the conclusion holds, regardless of whether we use pixels that are adjacent along the fast or the slow readout direction. Spatially correlated noise–the difference between the orange and blue lines in Figure 12–remains similar as a fraction of total noise at all integration times. Residual $1/f$ noise will appear even in very long exposures, but it will not be any more prominent in these long exposures than it is in shorter exposures.

8 Conclusion

We present a characterization and analysis of the TVAC1 and TVAC2 dark signal and noise properties. We calculate the dark signal rate in "long" dark exposures and find that the rates (~ 0.018 e^-/s) are below the operational requirements of 0.05 e^-/s. We then test for and find that the dark signal is highly stable across frames and exposures. It is nearly linear and up-the-ramp, but shows evidence of a small and apparently electronic bias offset that exponentially decays over the first several reads. This signal can be monitored on-orbit and may need to be removed by a dark reference cube (a read-by-read dark). Using the TVAC2 data, we find that for a 3σ dark signal detection in individual reads of individual pixels will require 10^5 exposures, though we note this is dependent on the TVAC2 dataset and is likely to change during commissioning. A lower signal-to-noise ratio may be acceptable if systematics can be reduced, or a model could be constructed and fit for the decaying bias offset.

Overall, we find that the dark signal is very low and stable in the WFI detectors at nominal operations. While a dark rate image likely describes the dark signal over most resultants due to its stability, a dark cube reference file may be needed for the initial reads. Additional testing is also needed to determine if the dark signal is stable over different temperatures and illuminations.

Finally, we measure the properties of the read noise and find that it is correlated read-to-read in an individual pixel. As a result, the read noise in long ramps is larger than would be expected from uncorrelated noise. This should be accounted for in the ramp fitting step if possible, and in the exposure time estimates.

9 References

Betti S. K., Casertano S., Brandt T., Cosentino R., Desjardins T., Sanchez J., 2024, Application and Characterization of IRRC on Roman/WFI TVAC1 Data, Nancy Grace Roman Space Telescope Technical Reports, 16 pp.

Mosby G., et al., 2020, Journal of Astronomical Telescopes, Instruments, and Systems, 6, 046001

Mosby G., et al., 2024, in Coyle L. E., Matsuura S., Perrin M. D., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 13092, Space Tele-

scopes and Instrumentation 2024: Optical, Infrared, and Millimeter Wave. p. 130923K, doi:10.1117/12.3020366

Petric A., et al., 2023, Characterization of Roman Detectors Darks Data Acquired with Leach and ACADIA Controllers, Nancy Grace Roman Space Telescope Technical Reports, 26 pp.

Roman Calibration Pipeline Team 2025, Roman Calibration Pipeline, https://github.com/spacetelescope/romancal

Roman Reference File Pipeline Team 2025, Roman Reference File Pipeline, https://github.com/spacetelescope/roman-wfi-reference-pipeline