

Simple Improved Reference Subtraction for H4RG, H2RG, and H1RG Near-infrared Array Detectors

Bernard J. Rauscher^{a,*}, Dale J. Fixsen^b, Gregory Mosby Jr.^a

^aObservational Cosmology Laboratory, NASA Goddard Space Flight Center, 8800 Greenbelt Rd., Greenbelt, USA, 20771

^bCRESST/UMd/GSFC, Greenbelt, USA, 20771

Abstract. Teledyne’s H4RG, H2RG, and H1RG near-infrared array detectors provide reference pixels embedded in their data streams. Although they do not respond to light, the reference pixels electronically mimic normal pixels and track correlated noise. In this paper, we describe how the reference pixels can be used with linear algebra and machine learning to optimally reduce correlated noise. Simple Improved Reference Subtraction (SIRS) works with common detector clocking patterns and, when applicable, relies only on post-processing existing data. The resulting reference correction is optimal, in a least squares sense, when the embedded reference pixels are the only references available. We demonstrate SIRS using H4RG ground test data from the Nancy Grace Roman Space Telescope Project. The SIRS software is freely available for download.

Keywords: H4RG, H2RG, H1RG, Reference.

*Bernard J. Rauscher, Bernard.J.Rauscher@nasa.gov

1 Introduction

We describe a simple and effective post-processing technique for reducing the correlated noise in many near-infrared (NIR) astronomy images. Simple Improved Reference Subtraction (SIRS) uses the reference columns that are built into all Teledyne HxRG $\in \{\text{H4RG}, \text{H2RG}, \text{H1RG}\}$ NIR arrays to efficiently suppress correlated noise.

SIRS builds on the same mathematical foundation as the James Webb Space Telescope’s (JWST) Near Infrared Spectrograph’s (NIRSpec) Improved Reference Sampling and Subtraction (IRS²; pronounced “IRS-square”),^{1,2} although it does not require the specialized NIRSpec clocking pattern. Instead, SIRS relies on traditional clocking patterns and post processing. As such, SIRS is even applicable to archival data. Like IRS² however; SIRS uses linear algebra and machine learning to make reference corrections that are linear, deterministic, and optimal with least squares as the figure of merit.

1.1 When Should SIRS be Considered?

SIRS’s main benefit is reducing correlated noise. To quickly determine whether SIRS might help in a particular situation, the key diagnostic is visible banding in slope images (see Fig 4; we defer detailed discussion of this figure to Section 5). If banding is visible, and if the typical band has a height that is more than two lines, then SIRS should help. A “slope image” is the result of least squares fitting straight lines to up-the-ramp sampled IR array data.

More quantitatively, SIRS works when there is temporal correlation in the pixel time series that is longer than the time required to read two lines. The factor of two is needed because the reference columns must critically sample correlated noise to remove it. In Fourier space, this becomes $f_{\text{noise}} < f_{\text{Ny, line}}$, where

$$f_{\text{Ny, line}} = \frac{1}{2t_{\text{line}}}. \quad (1)$$

Here, t_{line} is the time required to read one line including any overheads for starting the next line.

Quickly returning to Fig 4: (1) there is visible banding and (2) the height of a typical band appears to be a few lines. This is a situation where SIRS helps.

1.2 Paper Outline

The rest of this paper is structured as follows.

Section 2 presents the scientific rationale for why reducing correlated noise is important. This includes a discussion of why correlated noise is particularly important to Roman.

In Section 3, we briefly describe the technical features of HxRG detectors that are most important for this paper. These include the reference pixel layout and a description of the conventional clocking pattern that was used. Teledyne's user manuals describe several clocking options. This section explains which options were used here.

We then describe the theory in Section 4. This includes the key SIRS equations. This section presents a new derivation that can be easily extended to incorporate additional references as they become available.

SIRS's development was motivated and enabled by the large data sets that the Goddard Detector Characterization Laboratory (DCL) is producing for Roman. In Section 5 we use these data to show how SIRS works in practice.

Finally, Section 6 describes future work. Although the current data do not contain the H4RG's reference output, in the near future we expect to have data that do. Here we describe how SIRS will be modified and validated once these new data become available.

We close with a summary.

2 Science Case

Mostly, SIRS reduces correlated noise. Although there is some reduction in the (uncorrelated) variance, since it is only a few percent it is barely significant in practice. However, correlated noise is more pernicious. There are many scientific studies that could be adversely affected by the types of correlated noise that SIRS removes.

At the most basic level, nearly all NIR astronomy uses more than one pixel to make measurements. When this is done, whatever the measurement, one must consider the covariance within the set of pixels that is used. By definition, uncorrelated noise falls along the diagonal of the covariance matrix and one can follow the familiar rule of adding uncertainties in quadrature. But, this breaks down when correlation is present and simply adding the uncertainties in quadrature risks badly under-estimating the true uncertainty.

The pernicious effects go far beyond propagation of errors however. Roman's weak lensing survey provides an example. It uses the gravitational distortion faint field galaxy shapes to probe the growth of structure in the universe. The correlations addressed here cause banding, and tend to raise or lower groups of lines in the detector. This in turn tends to exaggerate the extent of galactic images in the rapid scan direction. For galaxies at the edge of visibility this can be a significant distortion. Even for galaxies that are well above the noise, the correlated noise can still complicate measuring the shear field. Although weak lensing groups are expert in these things, we believe it is nevertheless better to just remove the correlated noise if possible before the careful scientific analysis starts.

A second example is provided by correlations between galaxies. A small amount of correlated noise can enhance galaxies just enough to push them over the detection threshold. Since the noise lines appear at distinct intervals on the detector, they will appear at some angle on the sky. That in turn will manifest at some preferred spatial angle on the sky.

There can be many other examples where the subtle effects of correlated noise that appear at or even below the random noise level, but by combining many images or integrating over many sources can be enhanced, just like the signal, to become the dominant source of noise. For 2nd order measurements like correlations, galaxy distortions, *etc.*, correlation not only generates noise –it can generate bias.

3 Technical Description of HxRG Detectors

Teledyne’s HxRG detector arrays are among the most widely used astronomy sensors today. All HxRGs have a photosensitive mercury-cadmium-telluride (HgCdTe) detector layer bonded to a silicon readout integrated circuit (ROIC). The ROIC routes control voltages and clocks to pixels and multiplexes the many individual pixel outputs to a much smaller number of detector outputs. For easy reference, Table 1 summarizes some of the key parameters that relate to SIRS. Readers who desire a more thorough introduction to HxRGs may wish to see Mosby *et al.*³ and references therein.

Table 1 HxRG Formats

Parameter	Unit	H4RG	H2RG	H1RG
Pixels readout	pixels	4096×4096	2048×2048	1024×1024
HgCdTe pixels	pixels	4088×4088	2040×2040	1016×1016
Number of outputs	#	1, 4, 16, 32 or 64	1, 4, or 32	1, 2, or 16

Every instrument will be different and the appearance of visible banding in slope images remains the best indicator for when SIRS might be beneficial. Broadly speaking though, we would expect SIRS to work best with short frame readout times. At the 10^5 pixels s^{-1} output $^{-1}$ readout rate that is typically used, we would expect SIRS to potentially benefit H4RG and H2RG systems using ≥ 32 outputs and H1RG systems using 16 outputs.

When fewer outputs are used, there are too few references and they are too far apart. This is what motivated us to develop IRS², including its specialized clocking pattern, for JWST which reads its H2RGs using only 4 outputs.

3.1 Reference Pixels

We assume that the typical reader is an astronomy technologist or astronomer. When seen in science data using a common image visualizer such as SAOImage DS9,⁴ the outputs usually appear as thick vertical stripes running the full vertical “height” of the detector. For example, in an H4RG readout using 32 outputs, the pixels from one output would appear in a contiguous 128 pixel wide by 4096 pixel high rectangle. If one looks carefully at an exposure with some light in it, the reference pixels will appear as a four pixel wide dark frame surrounding the illuminated area unless the pipeline has cropped them off. The description given here is how the outputs most often appear in science data. Sometimes they appear transposed, but this is an unimportant detail of how the instrument builders have chosen to present the data.

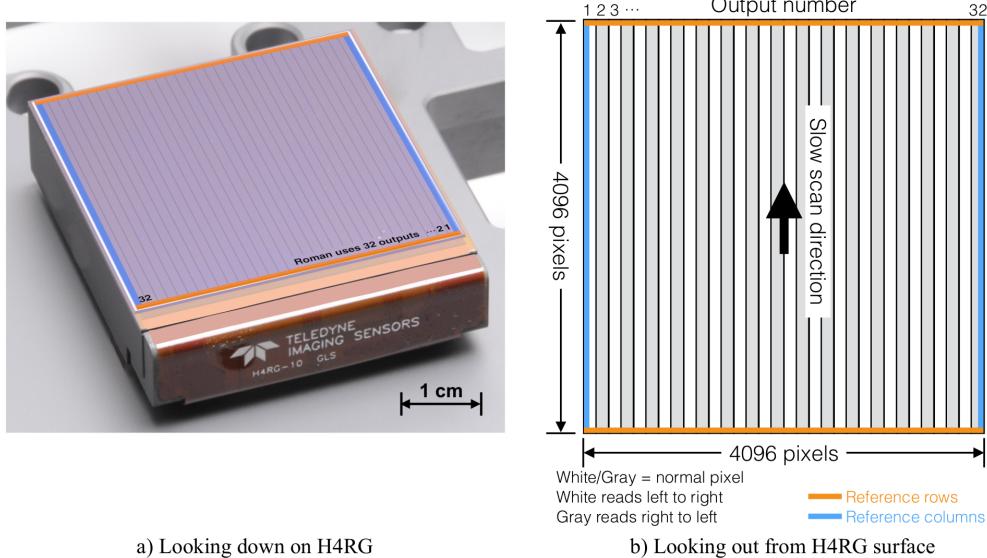


Fig 1 We developed SIRS for use with Roman ground test data. a) Roman’s Wide Field Instrument (WFI) will fly eighteen 4096×4096 pixel Teledyne H4RGs. b) The WFI configuration uses 32 outputs as shown here. All HxRGs embed a 4-pixel wide border of reference pixels in the data streams. The Roman fast scan directions use the default Teledyne configuration that alternates as shown here. The pixel readout rate is 2×10^5 pixels s^{-1} output $^{-1}$. There is a few pixel time overhead at the end of each row before clocking the first pixel of the next row. The new row time overhead will probably change before launch, but these details are easily accounted for within SIRS. The prototype H4RG-10 shown in panel a is included for illustrative purposes only. It is not the flight design. It should not be relied upon for engineering purposes.

Fig 1 shows how these features are physically laid out looking down on a Nancy Grace Roman Space Telescope (“Roman”) H4RG. The important point here is that there are *reference rows* and *reference columns*. There is also a reference output, but we defer detailed discussion to a future paper. Unlike the reference pixels, the reference output is not always available to observers.

The reference columns and rows do different but complementary things. Only the left and right-most outputs in Fig 1 contain reference columns, although in practice we find that there is often a useful degree of correlation with the other outputs. For every row that is read out, one gets eight reference samples from the reference columns. At Roman’s 200 kHz pixel rate, the corresponding Nyquist frequency is about, $f_c \approx 741$ Hz.

Although the reference rows are only sampled at the frame rate, $f_c \approx 0.18$ Hz, for Roman, they are present in every output.

In all HxRG family detectors, the ROIC provides a four pixel wide border of reference pixels that frames the regular pixels on all sides. Ever since the first engineered reference pixels became available in the Hubble WFC3’s detector, there has been lively debate about how to best use them.

3.2 SIRS Works with Common HxRG Clocking Patterns

SIRS places no special requirements on how the detector is operated. It works with readout schemes that are described in Teledyne’s user manuals. For current Roman testing, we are using Teledyne’s “Enhanced Clocking Mode” and “Pixel by Pixel Reset”. The WFI’s 32 H4RG outputs are read out using a Gen-III Leach Controller from Astronomical Research Cameras, Inc.,

at 2×10^5 pixels s^{-1} output $^{-1}$. There is a little time overhead at the end of each line to allow for clocking to and starting the next line. This overhead is currently about 7 pixel readout times –although this will probably change for flight.

In practice, there are many ways to implement Teledyne’s recommended clocking patterns. In our software, the implementation-dependent parameters are defined in the module `SIRS.jl`. We recommend that users check with their instrument support teams to adjust these settings as necessary.

4 Theory

SIRS has the same mathematical foundation as IRS².² The data are viewed as time series (also called “streams” here) rather than images. Building on the principal component analysis (PCA) that was done for JWST NIRSpec,¹ the readout system is assumed to be covariance stationary during the frame readout time.

With these assumptions, Fourier space approximates the eigenspace and thereby provides an uncorrelated representation of the read noise. The covariance matrix is diagonal in Fourier space. One practical consequence is that we can fit the Fourier components individually. In Section 5, we show by test that these are reasonable assumptions for Roman H4RGs.

One might imagine that one could fit equally well in the time domain. This would, however, require fitting full matrices rather than just vectors on account of the off-diagonal covariance. Doing so requires $\approx \frac{n}{2} \times$ more data, where n is the length of the diagonal. For Roman’s H4RGs, $n = 4096$. The stationary condition not only saves some computing time, but more importantly it allows us to work with far less data. Although computer time is inexpensive and becoming less so every year, observing time on space observatories will always be precious.

4.1 Derivation of Key Equations

This derivation leads to almost the same equations as appear in Rauscher *et al.* (2017).² However, we eliminate a filter that is no longer needed and introduce a more compact notation that better reflects the references that we use here. This derivation can be straightforwardly extended to include more reference streams as they become available.

Each frame of data contains normal pixels (grouped by output), reference columns on the left and right sides, and reference rows on the top and bottom. The data from the normal pixels and reference columns can be formatted into time series vectors; (**n**) normal pixels, (**l**) left reference columns, and (**r**) right reference columns. Projected into Fourier space these become \mathbf{n} , ℓ , and \mathbf{r} . In this paper, boldface type indicates vectors and calligraphy fonts generally indicate Fourier space. Vectors that appear in calligraphy fonts are therefore complex.

We model the read noise in the normal pixels as a linear combination of the references,

$$\boldsymbol{\alpha} \circ \boldsymbol{\ell} + \boldsymbol{\beta} \circ \boldsymbol{r} = \mathbf{n}. \quad (2)$$

In this expression, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of frequency dependent weights. The symbol \circ represents the element-wise (Hadamard) product. SIRS uses linear algebra and machine learning to find the optimal values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. We describe the linear algebra here. The supervised machine learning happens when training data are used to infer optimal values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that are then used to calibrate other data.

The training data are a large set of continuously sampled darks. We typically use about 100 dark exposures, each containing 60 up-the-ramp reads. The training data therefore contain about $m \approx 6 \times 10^3$ frames of data. Because the Fourier components are uncorrelated, we can solve for them independently. Let n , ℓ , and τ represent one element each of their corresponding vectors. Using Equation 2, we can write a set of $m \gg 2$ equations in the two unknowns, α and β .

$$\begin{aligned} \alpha\ell^1 + \beta\tau^1 &= n^1 \\ \alpha\ell^2 + \beta\tau^2 &= n^2 \\ &\vdots \\ \alpha\ell^m + \beta\tau^m &= n^m \end{aligned} \tag{3}$$

In matrix notation, these become,

$$\begin{pmatrix} \ell^1 & \tau^1 \\ \ell^2 & \tau^2 \\ \vdots & \vdots \\ \ell^m & \tau^m \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} n^1 \\ n^2 \\ \vdots \\ n^m \end{pmatrix}. \tag{4}$$

The least squares solution of Equation 4 is,

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \ell^1 & \tau^1 \\ \ell^2 & \tau^2 \\ \vdots & \vdots \\ \ell^m & \tau^m \end{pmatrix}^+ \begin{pmatrix} n^1 \\ n^2 \\ \vdots \\ n^m \end{pmatrix}, \tag{5}$$

where the $^+$ sign denotes the Moore-Penrose inverse of the “ $\ell\tau$ matrix”. Because the $\ell\tau$ matrix’s columns are linearly independent (they are physically different references), the Moore-Penrose inverse has an analytic solution. Substituting it, Equation 5 simplifies to,

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \ell_i^* \ell^i & \ell_i^* \tau^i \\ \tau_i^* \ell^i & \tau_i^* \tau^i \end{pmatrix}^{-1} \begin{pmatrix} \ell_i^* n^i \\ \tau_i^* n^i \end{pmatrix}. \tag{6}$$

In these expressions, the $*$ symbol denotes complex conjugation. We have also used the Einstein summation convention to achieve a clean notation. For example, $\ell_i^* \ell^i$ stands for $\sum_{i=1}^m \ell_i^* \ell^i$.

To further simplify Equation 6, it is helpful to define some sums;

$$\mathbb{L} = \ell_i^* \ell^i, \tag{7}$$

$$\mathbb{R} = \tau_i^* \tau^i, \tag{8}$$

$$\mathbb{X} = \tau_i^* n^i, \tag{9}$$

$$\mathbb{Y} = \ell_i^* n^i, \text{ and} \tag{10}$$

$$\mathbb{Z} = \ell_i^* \tau^i. \tag{11}$$

The first two sums are real and the last three are complex. With these substitutions, Equation 6 simplifies to,

$$\alpha = \frac{\mathbb{R}\mathbb{Y} - \mathbb{X}\mathbb{Z}}{\mathbb{R}\mathbb{L} - \mathbb{Z}\mathbb{Z}^*} \text{ and} \tag{12}$$

$$\beta = \frac{\mathbb{X}\mathbb{L} - \mathbb{Y}\mathbb{Z}^*}{\mathbb{R}\mathbb{L} - \mathbb{Z}\mathbb{Z}^*}. \tag{13}$$

In Equation 2, we defined α and β as vectors in Fourier space, but Equations 12-13 are for single Fourier components. Equations 7-13 must therefore be evaluated for all frequencies that will be used to make the reference correction.

There are other ways to find the least squares solution to Equation 4. The least squares itself is just a generalization of the familiar solution for fitting a 2-parameter straight line to data. We could therefore have subtracted the right side of Equations 4 from the left side, summed the squares of the deviates (remembering that the deviates are complex numbers), and minimized this sum over α and β . The result is the same, although the geometric approach presented here scales more gracefully as additional reference streams become available.

To better understand how these relations are used in practice, we recommend examining the source code, and especially the `coadd.jl` module. For those using Jupyter notebooks, the notation is the same as that used here. The SIRS distribution also contains a PDF of `coadd.jl` (in the Documentation directory) to aid those not using Jupyter.

4.2 Examining α and β

We developed SIRS using data from 18 Roman flight candidate H4RGs. Some of these tests are described in more detail in Section 5. Here we briefly discuss the measured values of α and β for one of the Roman H4RG flight candidates: SCA 20663. These results are broadly representative.

Fig 2 shows the low frequency behavior on a semi-log scale as a function of output number. The H4RG's outputs are numbered 1–32, with output #1 being closest to the left reference columns that α operates on and output #32 being closest to the right reference columns that β operates on. Output #16 is just to the left of the array's center line.

As expected, the relative importance of α and β shifts depending on which reference columns are closest. Although the noise is dominated by $1/f$ at low frequencies, many spectral bands and lines are present. We have not tried to track down all of their causes, but we find that cryocoolers often appear at about 8 Hz, as seems to be the case here.

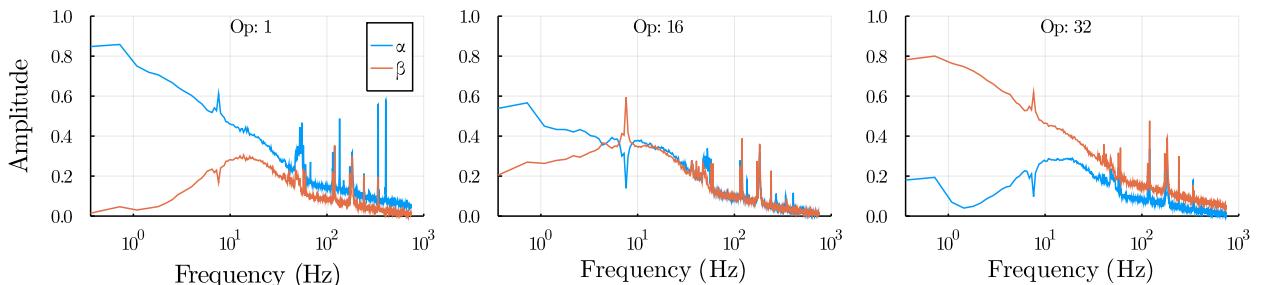


Fig 2 This figure shows α and β for 3 of H4RG 20663's 32 outputs as a function of position. Output #1 is the left-most output, output #16 is just left of center, and output #32 is the right-most output. As expected, the relative weights taken by α and β shift depending on which reference columns are closest. Broadly speaking, outputs 2–15 look intermediate between outputs 1 & 16 and similarly outputs 18–31 look intermediate between outputs 17 & 32.

Fig 3 shows the same data on a linear scale and includes the highest frequencies near $f_{Ny} = 200$ kHz. Although we only show one output, these plots are typical if allowance is made for the shifting relative importance of α and β .

Fig 3a again shows the strong low frequency correlation. It also highlights some features that one often sees with US power: lines at 60 Hz (and harmonics) with a strong band at 180 Hz on

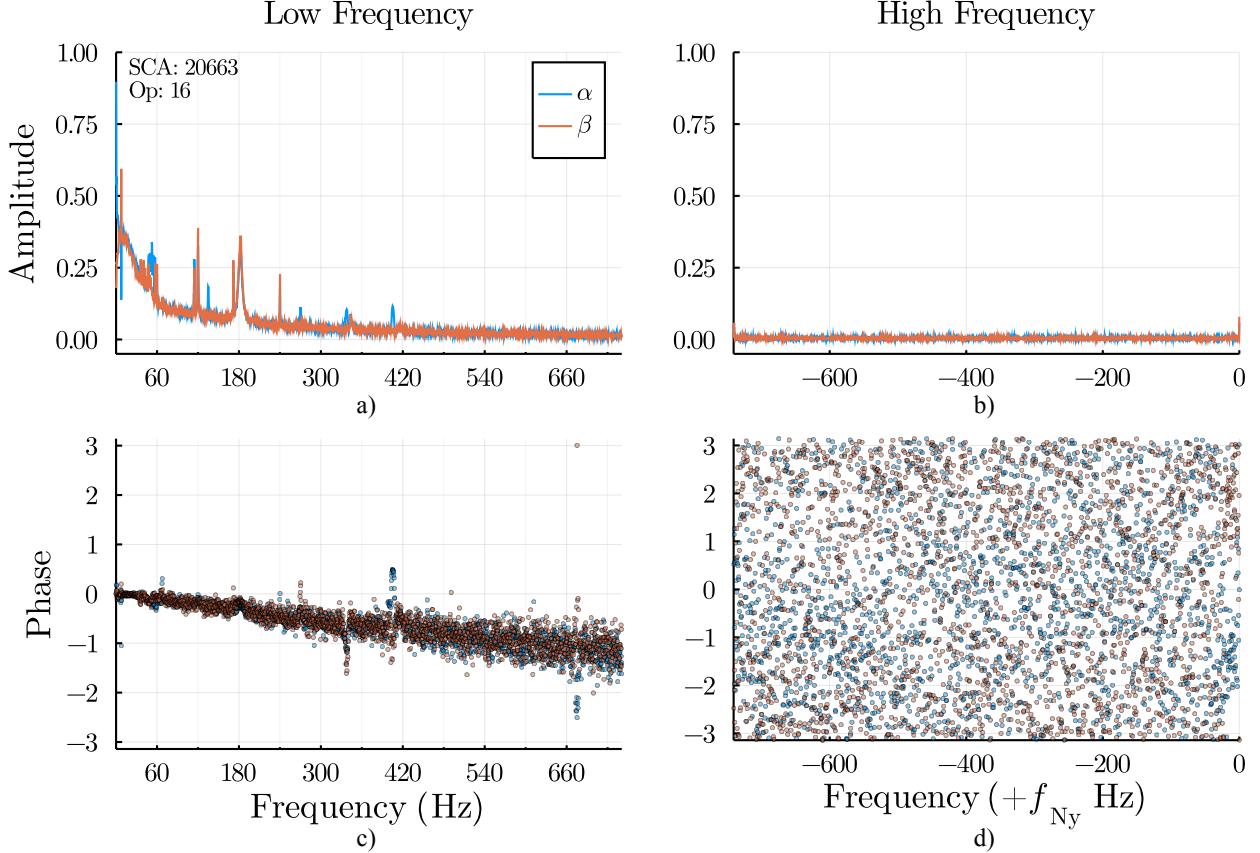


Fig 3 a) This figure re-plots Fig 2 (center) on a linear scale. Visible features include $1/f$ noise and lines and bands associated with the US power grid at 60, 120, 180, and 240 Hz. c) Because the reference pixels are the first pixels read in every line, the absolute value of the phase increases from zero at 0 Hz to $\frac{\pi}{2}$ at $f_{Ny, line}$. b) There is very little amplitude near the Nyquist frequency, $f_{Ny} = 200$ kHz, and d) phase is almost entirely noise at these frequencies. We therefore zero-out frequencies $f > f_{Ny, line}$ when working with the existing Roman data.

account of the three phase power. Panel c shows that at 0 Hz, everything is in phase. But, because the reference pixels are the first pixels read on every line, the absolute value of the phase increases to $\frac{\pi}{2}$ at $f_{Ny, line}$. The phase noise increases as the amplitude decreases.

Panels b and d show the high frequency behavior near $f_{Ny} = 200$ kHz. Our conclusion is that there is no useful reference information at these frequencies for these detectors. We are therefore zeroing-out α and β for $f > f_{Ny, line}$ in the current Roman test data.

This was a pleasant surprise. In JWST NIRSpec, there was significant alternating column noise (ACN) at f_{Ny} . This mixed with low frequency $1/f$ to produce a high frequency peak (See Fig 8a of Ref. 2). This was caused by a Teledyne-proprietary design choice in the H2RG column buses.^{2,5} Roman's H4RG-10s do not have this spectral feature.

Because we thought that similar ACN might still exist in the Roman H4RGs, SIRS fits for both low and high frequencies as shown here. If we had known in advance that only the low frequencies were important, it would have simplified the implementation somewhat. For now, we have left this feature in as things may change when the detectors are mated to the real flight electronics.

5 Validation using Roman Test Data

To validate the reduction of correlated noise using the SIRS technique described in Section 4 and verify no changes in measured response, we compare the noise properties of a set of SIRS corrected data and a set of data corrected with a standard choice of reference pixel correction. For the standard choice of reference pixel correction, we choose to correct each output by the median of the last 4 rows of reference pixels. This is a common choice of reference pixel correction that is used in the analysis of the Roman mission’s sensor acceptance testing.

The set of data used in these validation tests are from the Roman flight candidate detector 20663 taken by Goddard’s Detector Characterization Lab (DCL) as part of the sensor’s acceptance test. All data was taken using Astronomical Research Camera Inc.’s Gen-III Leach controllers with a ARC-22 timing card and ARC-46 eight channel video processing boards. The H4RG is read out using line-by-line reset clocking patterns for 32 output mode described in the H4RG-10 user manual implemented with custom DCL software. The H4RG is cooled to 95 K with a cryocooler and operated at a 1 V reverse bias for these tests. The DCL also tested this detector at 90 K and 0.5 V reverse bias. SIRS worked comparably well. The results shown here are representative.

The data for the noise comparison are a set of 105 dark exposures, consisting of 60 consecutive reads each. This set of data is typically used to estimate the total slope noise of a detector. Dark current is known to be very low ($\lesssim 0.001$ e-/s/pixel) for these devices from qualification testing and acceptance tests results. The second data set to look for any changes in measured response is a set of 20 exposures consisting of 100 consecutive reads with $1.4\ \mu\text{m}$ illumination using a tungsten lamp directed through a monochromator to produce a flatfield onto the detector.

All the following analysis also excludes pixels ($\sim 1.2\%$) that have been flagged as inoperable according to the results of Goddard Detector Characterization Lab Acceptance Testing.

5.1 Noise comparison

We can qualitatively compare the amount of correlated noise in the SIRS corrected set of data and the standard corrected set of data by inspecting the slope images from a representative exposure. Fig 4 shows the slope image from fitting the 60 samples up-the-ramp of a single exposure for SIRS corrected up-the-ramp data and up-the-ramp data using the standard correction. In Fig 4, the horizontal banding from correlated noise in the slow readout direction is evident and stronger in the data corrected using only the last 4 rows of reference pixels.

To quantify the excess noise in data using only the standard correction, we can examine the residual variation in each output along the slow readout direction after subtracting off the output mean slope. Fig 5 shows the residuals of the central output (#15) for both sets of data for a representative exposure in the top panel. On the lower panel of Fig 5, the ratio of the standard deviation of the SIRS corrected data residuals to data residuals using the standard correction is plotted for each output of the same exposure. For the exposure shown in Fig 5, the SIRS corrected slope image data has approximately 33% less variation. Overall, when looking at a total of 105 exposures, the residual variation in the SIRS corrected data is about 25% less than the residual variation in data corrected using the standard reference pixel correction.

The histograms of the residuals per output channel in Fig 5 appear coarsely normal, but to assess this quantitatively, we can look at statistical properties for both sets of residuals to test for normality. Normally distributed residuals would indicate random variation with pixel row and that the optimal model for the slope image data in an output is indeed a constant value independent

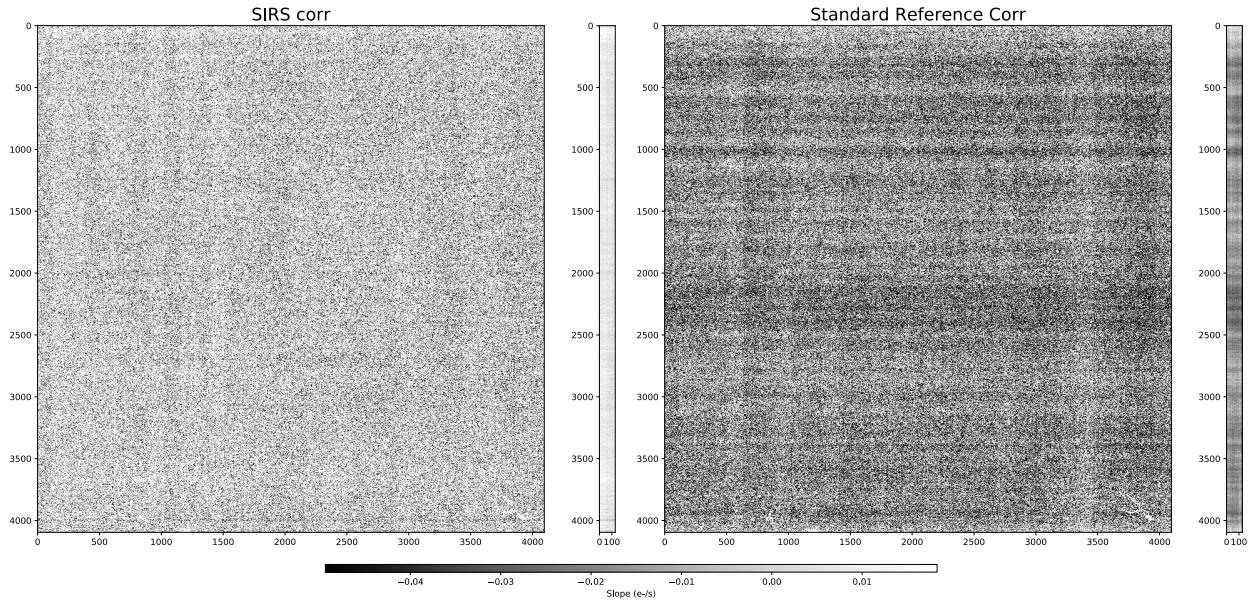


Fig 4 Representative comparison of slope images derived from (left) SIRS corrected ramp and a standard reference subtraction corrected ramp (right). Beside each image the output median for the 32 outputs is plotted to highlight the horizontal banding. The slope image derived from the SIRS corrected data show substantial reduction in the horizontal banding. The greyscale in the images is up-the-ramp group slope in units $e^- s^{-1} \text{ pixel}^{-1}$.

of pixel row. Non-normality would suggest additional parameters are needed to describe the data. Fig 6 shows the calculated skewness, kurtosis, and the Durbin-Watson statistic for the residuals in all 32 outputs as 3 histograms. Skewness and kurtosis are the commonly defined standardized third and fourth moments of a probability distribution, respectively. The Durbin-Watson statistic is a measure of the serial auto-correlation in the residuals.⁶ For a normal distribution, skewness would be zero, kurtosis would be 3, and the Durbin-Watson statistic would be 2 assuming no auto-correlation. For the representative exposure in Fig 6, the SIRS corrected data residuals have less skewness, data residuals from both corrections have excess kurtosis, and the SIRS corrected data have less auto-correlation. These results in conjunction with the results of Fig 5 suggest that the residuals from both types of data are not completely normally distributed. But they suggest that the residuals from the SIRS corrected data could be represented by a Gaussian distribution with narrower standard deviation with enhanced tails.

Normality is a key feature to look for in this data because it implies the pixel noise properties could permit averaging over multiple pixels to increase precision and accuracy by the reduction of uncorrelated noise. We compare how the pixel noise averages down for both types of reference pixel correction using square apertures of increasing size in the central output (#15) of their respective slope images. The square aperture sizes are 3×3 , 5×5 , 9×9 , 21×21 , 51×51 , and 101×101 pixels. The mean signal in an aperture is calculated for each exposure. The standard deviation of the mean as a function of aperture size is then calculated. The variation in the mean should decrease with increasing aperture size for data containing few correlations. Fig 7 shows the results for computing aperture means for 105 exposures and estimating the standard deviation of that mean for each aperture across those exposures. The middle panel of Fig 7 shows that the variation (noise) of the aperture mean in the SIRS corrected slope data decreases more rapidly with aperture

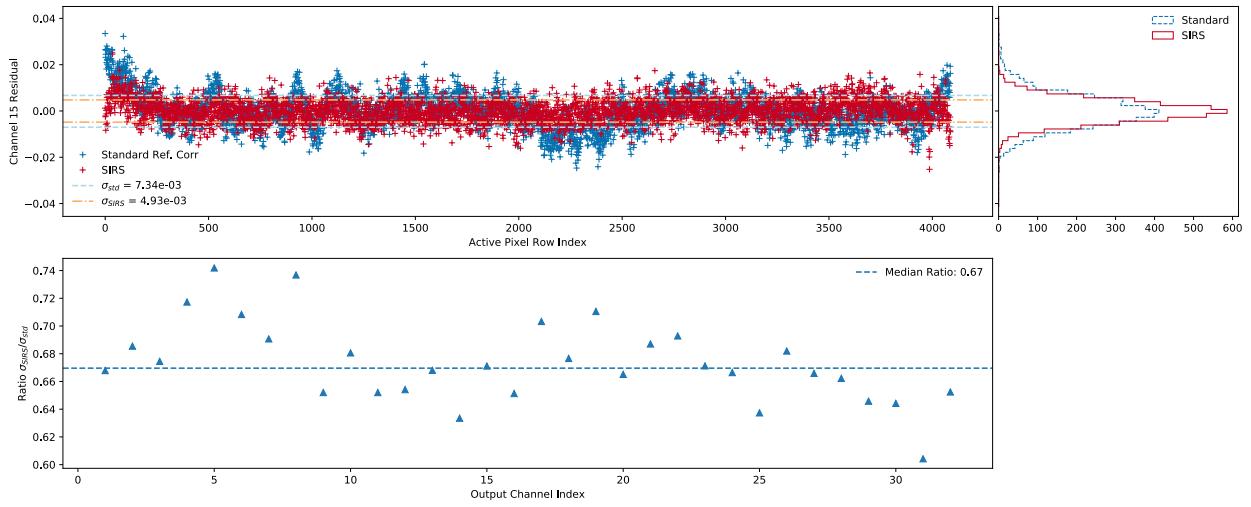


Fig 5 Representative comparison of residuals in the median slope in output #15 as a function of pixel row. Top panel shows the residual between the output median and the average value of the output median over all active pixel rows for both SIRS and standard reference pixel corrected data. Bottom panel shows the ratio of the standard deviation of the residuals in the top panel (SIRS divided by the standard correction) for all output channels. The residuals of the SIRS corrected slope image data are more strongly peaked at zero and show less variation than the data using the standard correction.

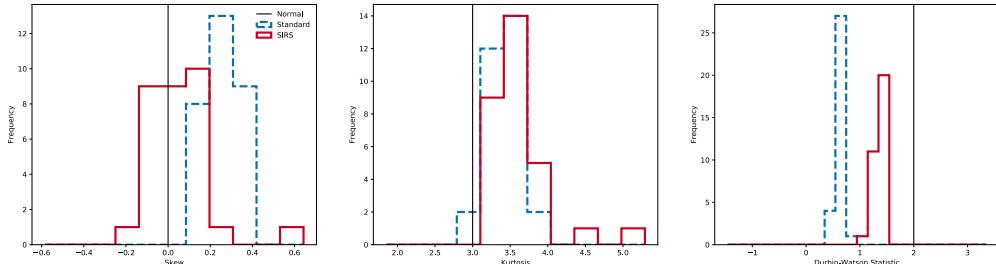


Fig 6 Statistical properties for output residuals from Fig 5. Left: distribution of skewness calculated for each output's residuals. Middle: distribution of kurtosis calculated for each output's residuals. Right: distribution of Durbin-Watson statistic calculated for each output's residuals. For a normal distribution, skewness would be zero, kurtosis would be 3, and the Durbin-Watson statistic would be 2. The SIRS corrected data residuals have less skewness, data residuals from both corrections have excess kurtosis, and the SIRS corrected data have less auto-correlation.

size and to nearly half the value of the variation of the aperture mean in the slope data corrected with the standard correction. The right panel of Fig 7 shows the relative standard deviation (standard deviation divided by absolute value of mean). We can see that the relative standard deviation is smaller for the SIRS corrected data, and that it decreases below the mean value in contrast to the relative standard deviation for the data corrected using the standard correction. This is consistent with the larger amount of correlation in the data using the standard correction.

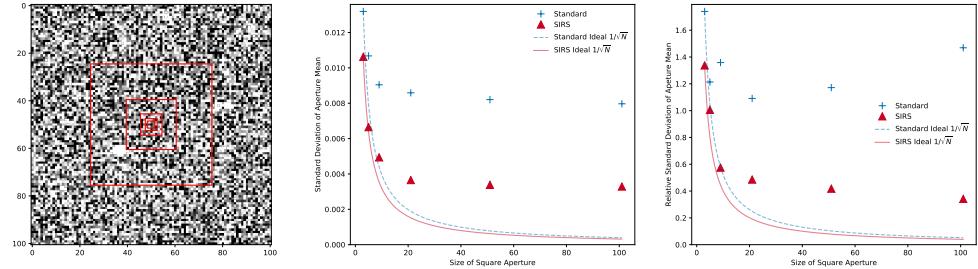


Fig 7 Left: Image of the region (101×101) where the aperture averaging is done with red rectangles overlaying the smaller apertures. Middle: Standard deviation of the mean as a function of aperture, calculated using 105 exposures. Right: Relative standard deviation (standard deviation divided by absolute value of mean) as a function of aperture, calculated using 105 exposures. The variation (noise) of the aperture mean in the SIRS corrected slope data decreases more rapidly with aperture size and to nearly half the value of the variation of the aperture mean in the slope data. However, both corrections bottom out after the aperture is larger than 9×9 indicating there is still room for improvement. The relative standard deviation is smaller for the SIRS corrected data, and it decreases below the mean value in contrast to the relative standard deviation for the data corrected using the standard correction.

5.2 Response comparison

To look for any potential differences in the measured response when using the SIRS correction versus the standard correction, we can examine the mean signal as a function of frames up-the-ramp using the second data set with flatfield illumination. The array mean is estimated in each frame after subtracting the baseline signal in the first frame and plotted as a function of time in Fig 8. No significant differences are apparent in the response implied from the two distinct reference pixel correction schemes.

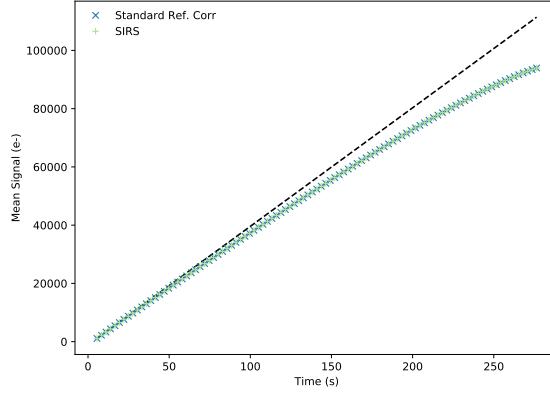


Fig 8 Mean signal in illuminated frames up-the-ramp for a representative illuminated exposure. The array mean is estimated in each frame after subtracting the baseline signal in the first frame and plotted as a function of time. The means for the SIRS corrected data lie on top of the means for the data corrected using the standard reference subtraction. There are no noticeable differences in implied response due to the SIRS correction.

6 Future Work

The DCL acquired the data that underlie this paper using flight candidate H4RGs and non-flight Leach controllers. For flight, most of the same detectors will be built into “triplets”. A triplet consists of: (1) H4RG detector, (2) interconnect cable, and (3) ACADIA application specific integrated circuit (ASIC). The differences include the ACADIA replacing the Leach controller, a different clocking pattern, and additional references are expected to be available.

The ACADIA is a new IR array controller ASIC that aims to improve on the Teledyne SIDE-CAR⁷ ASICs that were used by JWST, Euclid, and many other observatories. Because it is new, we cannot say anything about the performance of flight ACADIAs at this time other than we understand that they are meeting their requirements. For more information about ACADIA, we refer the interested reader to Loose *et al.* (2018).⁸

Roman plans to use the H4RG’s “guide window” for observatory guiding. Although the Roman Project has tested guide windows extensively, they were not used for the data that are the basis of this paper. Using the guide window requires a more complex clocking pattern that interleaves controlling the guide window with reading out normal pixels. The new timing pattern will have different gaps and overheads than the one that was used for this paper. This will in turn require modifications to the SIRS software.

In May, 2021, we expect the first triplet data to become available. To prepare for this, we have already solved Equation 2 for the “3-stream” case; where the reference streams are the left reference columns, the right reference columns, and the H4RG’s reference output (all HxRGs have a reference output; see Ref. 2 for a description). The solution is only a little more complicated than Equations 12-13, although we have not had a chance to validate it with data yet.

When triplet data become available, we plan to upgrade our software and study optimal reference correction for triplets. This will include validating the 3-stream solution mentioned above. Depending on the outcome, it is likely that we will publish a follow-on paper that builds on this one to provide an optimal reference correction scheme for flight.

SIRS is one element of a linear algebra and machine learning (LAML) approach to IR array calibration that we are experimenting with at Goddard. Others include JWST NIRSpec’s IRS²,² and new concepts for end-to-end detector calibration using Roman’s Relative Calibration System (RCS). The RCS is essentially a sophisticated flatfield projector that provides accurate and precise knowledge of relative brightness over a wide dynamic range. We look forward to writing about these concepts as they mature.

7 Summary

We have described Simple Improved Reference Subtraction (SIRS). SIRS uses training data with linear algebra and machine learning to reference correct Teledyne HxRG detector systems. SIRS is optimal using least squares as the figure of merit. For systems showing significant banding aligned with the fast scan directions, SIRS may reduce correlated noise including $1/f$ banding and some types of electromagnetic interference.

SIRS works with some of the most common HxRG readout patterns and makes no special demands on the hardware or how the detector is read out. When applicable, SIRS can be used to reduce the correlated noise of even archival data.

Disclosures

We have no relevant financial interests in the manuscript and no other potential conflicts of interest to disclose.

Acknowledgments

This work was supported by NASA as part of the Nancy Grace Roman Space Telescope and James Webb Space Telescope Projects. Resources supporting this work were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center. We are grateful for the invaluable support of the NASA Goddard Space Flight Center Detector Characterization Laboratory (DCL) team. The DCL produced all of the data used for this study.

Software: Julia,⁹ Python3,¹⁰ Matplotlib,¹¹ NumPy,¹² SciPy,¹³ statsmodels,¹⁴ ray,¹⁵ Cython¹⁶, SAOImage DS9⁴

Data, Materials, and Code Availability

The Julia language SIRS package is freely available for download from the NASA GitHub. It includes several jupyter notebook examples and an example showing how to apply SIRS reference correction using python-3 given the SIRS calibration file. To assist Roman affiliates, we have pre-computed SIRS calibration files for all flight candidate H4RGs.

References

- 1 S. H. Moseley, R. G. Arendt, D. J. Fixsen, *et al.*, “Reducing the read noise of H2RG detector arrays: eliminating correlated noise with efficient use of reference signals,” in *Proc SPIE*, A. D. Holland and D. A. Dorn, Eds., **7742**, 404–412, International Society for Optics and Photonics, SPIE (2010).
- 2 B. J. Rauscher, R. G. Arendt, D. Fixsen, *et al.*, “Improved Reference Sampling and Subtraction: A Technique for Reducing the Read Noise of Near-infrared Detector Systems,” *PASP* **129**, 105003 (2017).
- 3 G. Mosby, B. J. Rauscher, C. Bennett, *et al.*, “Properties and characteristics of the WFIRST H4RG-10 detectors,” *JATIS Submitted*, 1–47 (2020).
- 4 W. A. Joye and E. Mandel, “New Features of SAOImage DS9,” in *Astronomical Data Analysis Software and Systems XII*, H. E. Payne, R. I. Jedrzejewski, and R. N. Hook, Eds., *Astronomical Society of the Pacific Conference Series* **295**, 489 (2003).
- 5 B. J. Rauscher, “Teledyne H1RG, H2RG, and H4RG Noise Generator,” *Publ. Astron. Soc. Pacific* **127**(957), 1144–1151 (2015).
- 6 J. Durbin and G. S. Watson, “Testing for serial correlation in least squares regression: I,” *Biometrika* **37**(3/4), 409–428 (1950).
- 7 M. Loose, J. Beletic, J. Garnett, *et al.*, “High-performance focal plane arrays based on the HAWAII-2RG/4G and the SIDECAR ASIC,” in *Focal Plane Arrays for Space Telescopes III*, T. J. Grycewicz, C. J. Marshall, and P. G. Warren, Eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **6690**, 66900C (2007).

- 8 M. Loose, B. Smith, G. Alkire, *et al.*, “The ACADIA ASIC: detector control and digitization for the Wide-Field Infrared Survey Telescope (WFIRST),” in *Proc SPIE*, A. D. Holland and J. Beletic, Eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **10709**, 107090T (2018).
- 9 J. Bezanson, A. Edelman, S. Karpinski, *et al.*, “Julia: A fresh approach to numerical computing,” *SIAM review* **59**(1), 65–98 (2017).
- 10 G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA (2009).
- 11 J. D. Hunter, “Matplotlib: A 2d graphics environment,” *Computing in science & engineering* **9**(3), 90–95 (2007).
- 12 C. R. Harris, K. J. Millman, S. J. van der Walt, *et al.*, “Array programming with NumPy,” *Nature* **585**, 357–362 (2020).
- 13 P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python,” *Nature Methods* **17**, 261–272 (2020).
- 14 S. Seabold and J. Perktold, “statsmodels: Econometric and statistical modeling with python,” in *9th Python in Science Conference*, (2010).
- 15 P. Moritz, R. Nishihara, S. Wang, *et al.*, “Ray: A distributed framework for emerging AI applications,” *CoRR* **abs/1712.05889** (2017).
- 16 S. Behnel, R. Bradshaw, C. Citro, *et al.*, “Cython: The best of both worlds,” *Computing in Science Engineering* **13**(2), 31–39 (2011).

Bernard J. Rauscher is an experimental astrophysicist at NASA’s Goddard Space Flight Center. His research interests include astronomy instrumentation, extragalactic astronomy and cosmology, and the search for life on other worlds. Rauscher’s work developing detector systems for JWST was recognized by a (shared) Congressional Space Act award and NASA’s Exceptional Achievement Medal.

Dale J. Fixsen is an experimental astrophysicist at the University of Maryland and NASA’s Goddard Space Flight Center. He received his Ph.D. in physics from Princeton University in 1982. Fixsen’s work has included developing mathematical models for COBE FIRAS, far infrared instrumentation, and building NIR astronomy instruments using H1RGs.

Gregory Mosby, Jr. is a Research Astrophysicist and detector scientist at NASA Goddard Space Flight Center. He received his BS degree in astronomy and physics from Yale University in 2009, and his MS and PhD degrees in astronomy from the University of Wisconsin - Madison in 2011 and 2016, respectively. His current research interests include near infrared detectors, astronomical instrumentation, and applications of machine learning to observational astronomy. He is a member of SPIE.

List of Figures

- 1 We developed SIRS for use with Roman ground test data. a) Roman’s Wide Field Instrument (WFI) will fly eighteen 4096×4096 pixel Teledyne H4RGs. b) The WFI configuration uses 32 outputs as shown here. All HxRGs embed a 4-pixel wide border of reference pixels in the data streams. The Roman fast scan directions use the default Teledyne configuration that alternates as shown here. The pixel readout rate is 2×10^5 pixels s^{-1} output $^{-1}$. There is a few pixel time overhead at the end of each row before clocking the first pixel of the next row. The new row time overhead will probably change before launch, but these details are easily accounted for within SIRS. The prototype H4RG-10 shown in panel a is included for illustrative purposes only. It is not the flight design. It should not be relied upon for engineering purposes.
- 2 This figure shows α and β for 3 of H4RG 20663’s 32 outputs as a function of position. Output #1 is the left-most output, output #16 is just left of center, and output #32 is the right-most output. As expected, the relative weights taken by α and β shift depending on which reference columns are closest. Broadly speaking, outputs 2-15 look intermediate between outputs 1 & 16 and similarly outputs 18-31 look intermediate between outputs 17 & 32.
- 3 a) This figure re-plots Fig 2 (center) on a linear scale. Visible features include $1/f$ noise and lines and bands associated with the US power grid at 60, 120, 180, and 240 Hz. c) Because the reference pixels are the first pixels read in every line, the absolute value of the phase increases from zero at 0 Hz to $\frac{\pi}{2}$ at $f_{Ny, line}$. b) There is very little amplitude near the Nyquist frequency, $f_{Ny} = 200$ kHz, and d) phase is almost entirely noise at these frequencies. We therefore zero-out frequencies $f > f_{Ny, line}$ when working with the existing Roman data.
- 4 Representative comparison of slope images derived from (left) SIRS corrected ramp and a standard reference subtraction corrected ramp (right). Beside each image the output median for the 32 outputs is plotted to highlight the horizontal banding. The slope image derived from the SIRS corrected data show substantial reduction in the horizontal banding. The greyscale in the images is up-the-ramp group slope in units $e^- s^{-1} pixel^{-1}$.
- 5 Representative comparison of residuals in the median slope in output #15 as a function of pixel row. Top panel shows the residual between the output median and the average value of the output median over all active pixel rows for both SIRS and standard reference pixel corrected data. Bottom panel shows the ratio of the standard deviation of the residuals in the top panel (SIRS divided by the standard correction) for all output channels. The residuals of the SIRS corrected slope image data are more strongly peaked at zero and show less variation than the data using the standard correction.

- 6 Statistical properties for output residuals from Fig 5. Left: distribution of skewness calculated for each output's residuals. Middle: distribution of kurtosis calculated for each output's residuals. Right: distribution of Durbin-Watson statistic calculated for each output's residuals. For a normal distribution, skewness would be zero, kurtosis would be 3, and the Durbin-Watson statistic would be 2. The SIRS corrected data residuals have less skewness, data residuals from both corrections have excess kurtosis, and the SIRS corrected data have less auto-correlation.
- 7 Left: Image of the region (101×101) where the aperture averaging is done with red rectangles overlaying the smaller apertures. Middle: Standard deviation of the mean as a function of aperture, calculated using 105 exposures. Right: Relative standard deviation (standard deviation divided by absolute value of mean) as a function of aperture, calculated using 105 exposures. The variation (noise) of the aperture mean in the SIRS corrected slope data decreases more rapidly with aperture size and to nearly half the value of the variation of the aperture mean in the slope data. However, both corrections bottom out after the aperture is larger than 9×9 indicating there is still room for improvement. The relative standard deviation is smaller for the SIRS corrected data, and it decreases below the mean value in contrast to the relative standard deviation for the data corrected using the standard correction.
- 8 Mean signal in illuminated frames up-the-ramp for a representative illuminated exposure. The array mean is estimated in each frame after subtracting the baseline signal in the first frame and plotted as a function of time. The means for the SIRS corrected data lie on top of the means for the data corrected using the standard reference subtraction. There are no noticeable differences in implied response due to the SIRS correction.

List of Tables

- 1 HxRG Formats