Information Content of Up-the-ramp Sampled IR Array Data

Bernard J. Rauscher
NASA Goddard Space Flight Center

Presented 3 December 2018
Image Sensors for Precision Astronomy (ISPA), CalTech, Pasadena, CA

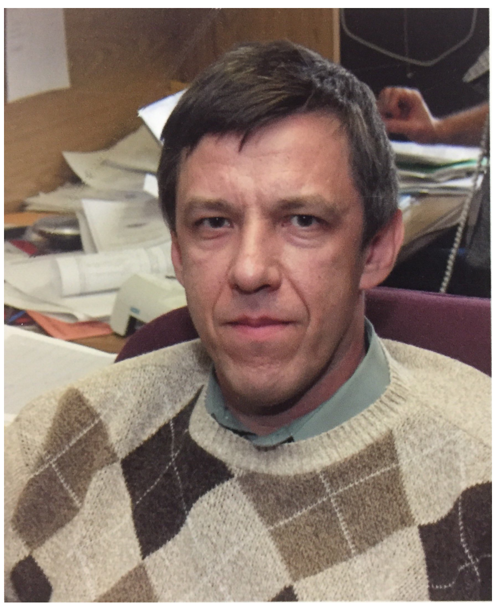
Bernard.J.Rauscher@nasa.gov

The Team

Dani Atkinson, GSFC/USRA

Rick Arendt, GSFC/UMBC

Dale Fixsen, GSFC/UMCP



Alexander Kutyrev, GSFC/UMCP

Greg Mosby, GSFC/USRA

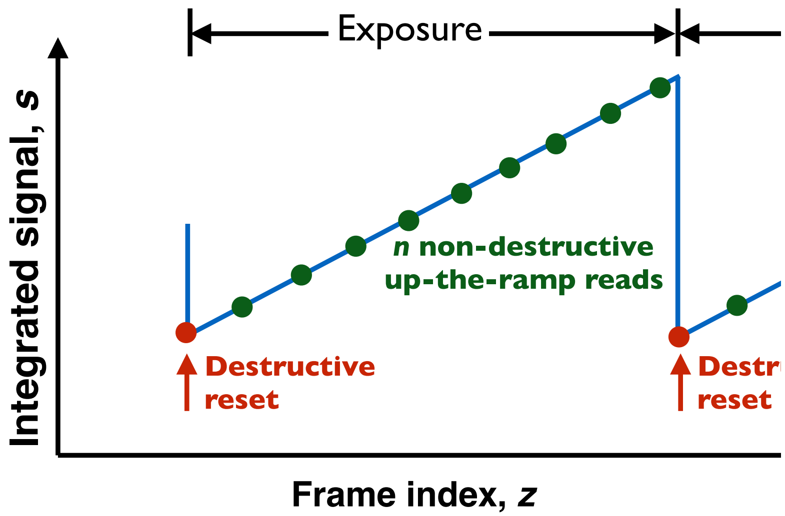
Harvey Moseley, GSFC

Overview

- The Wide Field Infrared Survey Telescope (*WFIRST*) requires excellent control of detector systematics
  - Microlensing bulge survey plans relative photometric residuals < 0.1%
  - High Galactic Latitude Weak Lensing Survey requires knowledge of PSF size and ellipticity to < 0.1%
- As a first step, in 2017 we began a study of prototype *WFIRST* H4RG-10 detectors with the goal of calibrating individual pixels to < 0.01% (no, we are not there yet!)
- Linearity correction quickly emerged as a challenge and motivated principal component analysis (PCA)
- PCA immediately led to new insights regarding information content of up-the-ramp sampled IR array data
- These charts are just the beginning of a study of linearity (and precision IR array calibration more generally), not the end of one…

The problem…

- For space, up-the-ramp sampling provides many benefits



- When characterizing linearity, up-the-ramp samples are generally fitted with a polynomial using a monomial expansion,

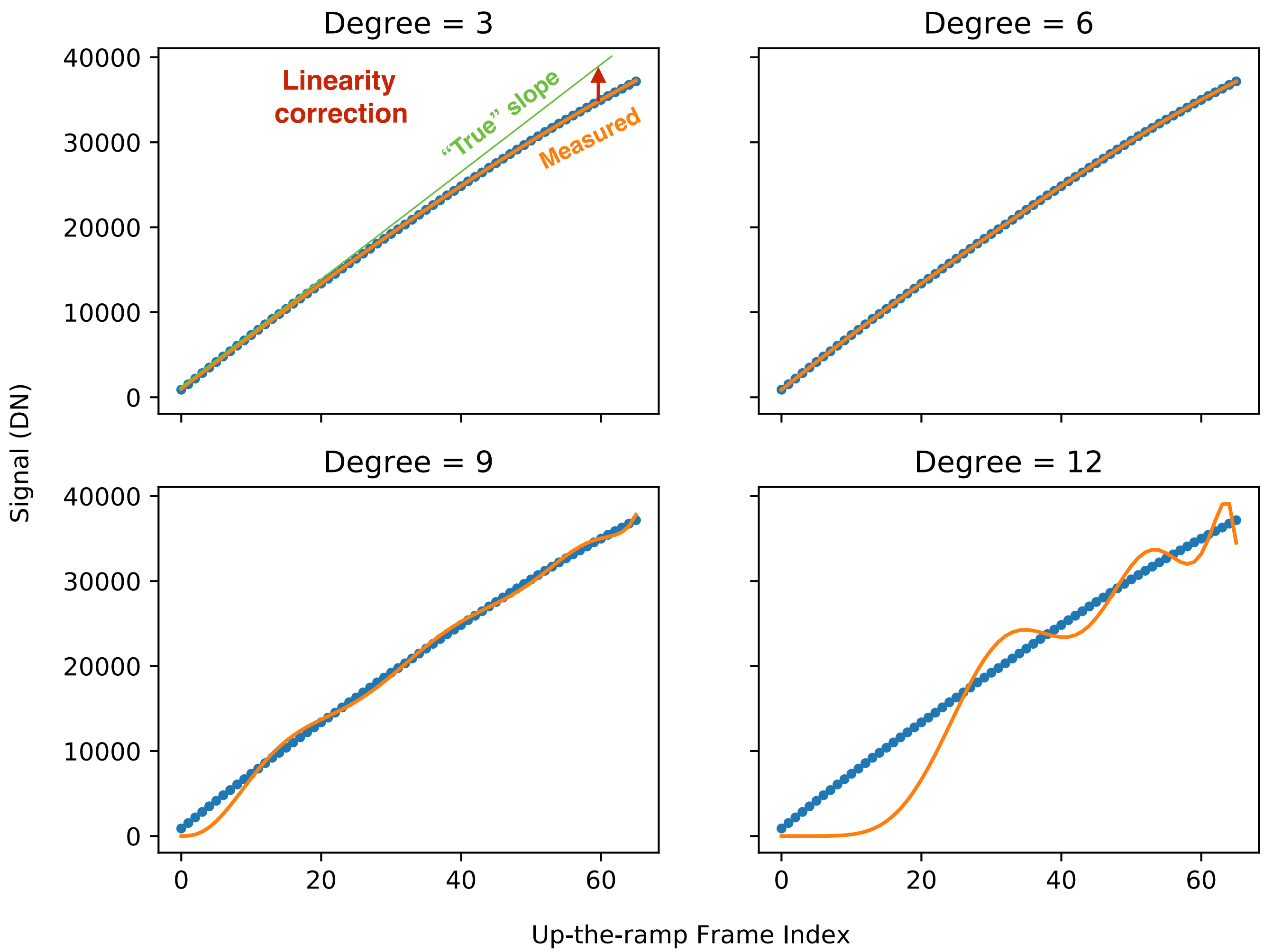
$$s(z) = \sum_{i=0}^{\deg} a_i z^i$$

- Basis vectors are, B ∊ $\{z^0, z^1, z^2, \ldots z^{\deg}\}$. Fit degree (deg) is typically, deg ~ 3
- Resulting a_i coefficients used to "linearize" data before fitting straight lines (2 free parameters) to make bias and slope images
- In the *WFIRST* data, there was no clear way to tell the right value of deg for characterizing linearity… Fits became computationally unstable for deg = 6 or so…
- Although computational instability could have been fixed with higher precision arithmetic, the data were not providing sufficient insight into the correct value of deg…

Bernard.J.Rauscher@nasa.gov

Some examples

- In low degree fits, one can see systematic residuals even by eye
- But, as higher degree is used, the fits become unstable (can be fixed with higher precision arithmetic)
- But, onset of computational instability is not a well motivated way to choose deg…



Bernard.J.Rauscher@nasa.gov

Principal Component Analysis (PCA)

- Input data were a 65 frame up-the-ramp sampled *WFIRST* flatfield exposure to the onset of soft saturation ($\sim 10^5\ e^-$; PCA has since been done for many other data sets and different instruments)

- 98% of the pixels passed a quality check. The data from each pixel were represented by a column vector, **d**

- We put the n resulting column vectors into a $65 \times n$ matrix **D**

- We computed the covariance matrix,

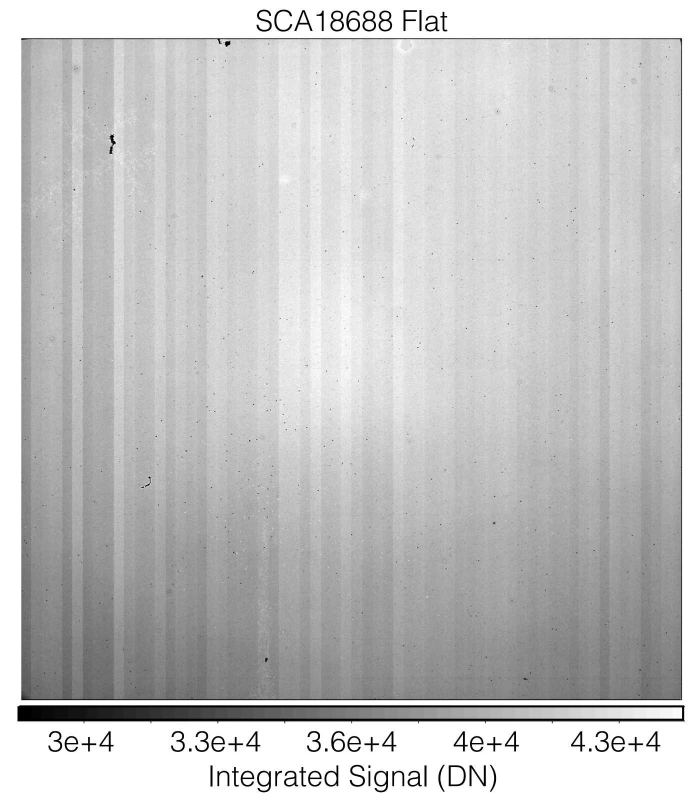
$$\Omega = \frac{(D - \langle D \rangle)\,(D - \langle D \rangle)^T}{n}$$

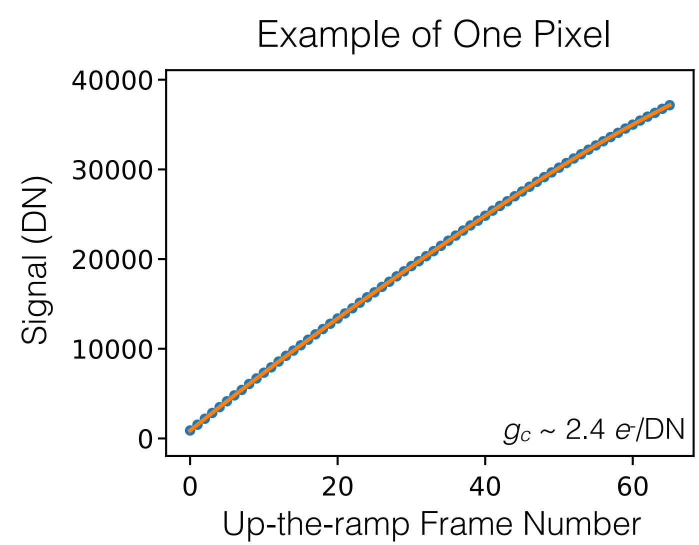
  where <**D**> is averaged over columns and broadcast to the shape of **D**

- The eigenvectors of $\mathbf{\Omega}$ provide an orthogonal basis for representing the data

- The eigenvalues provide a quantitative measure of the information content by component

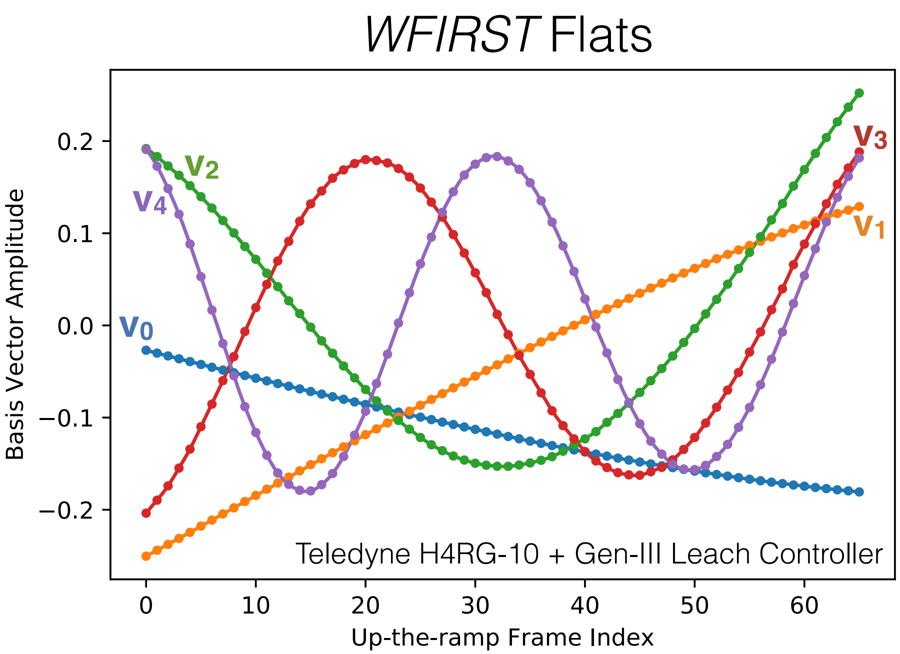
- We have since repeated the PCA on a variety of systems and data sets (including flats, darks, and astronomical observations) and gotten very consistent results

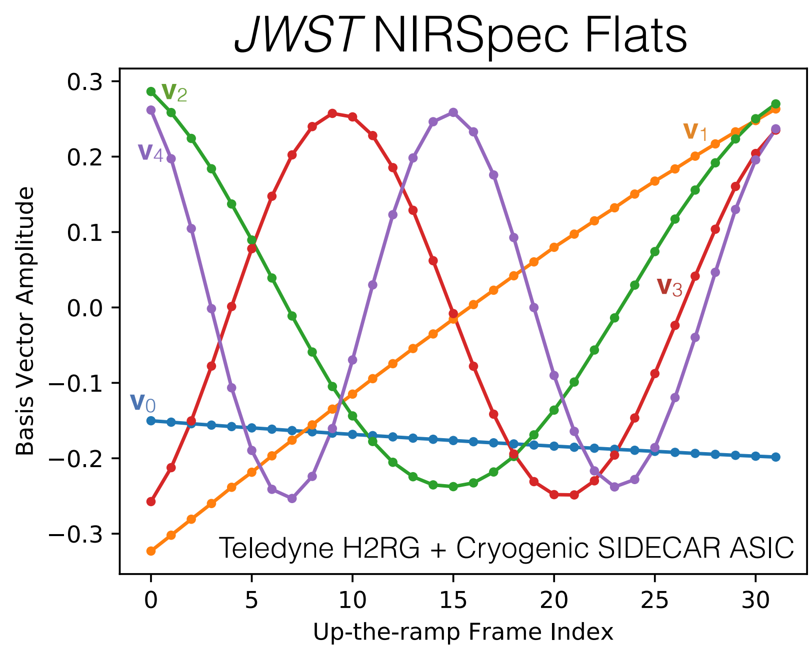


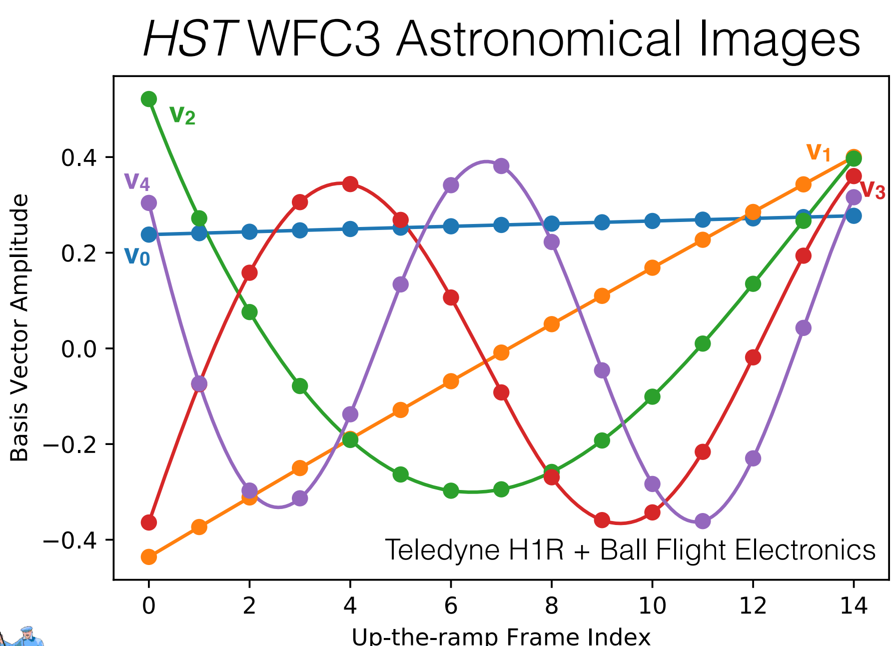


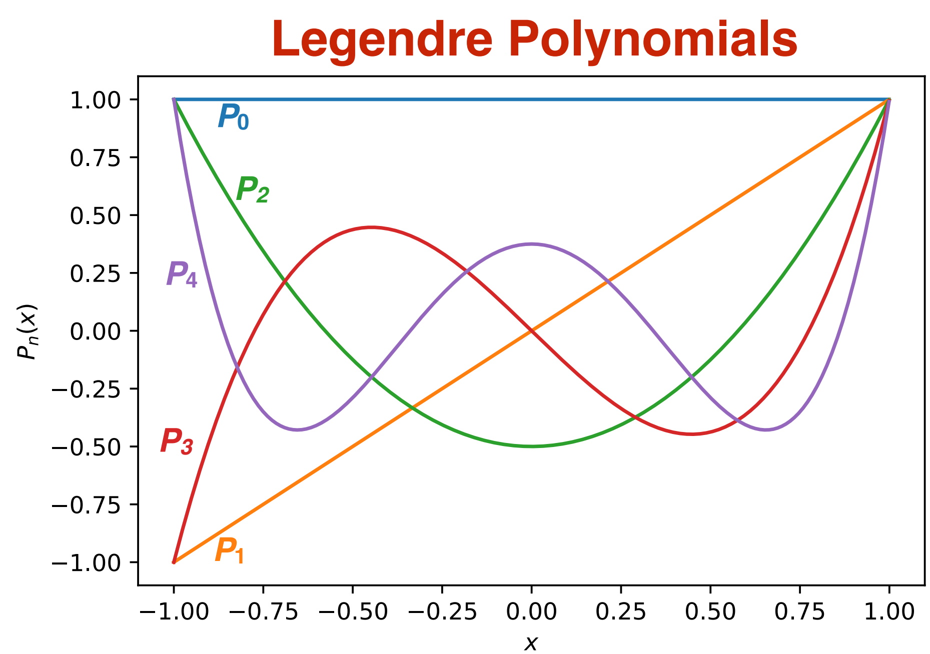
Bernard.J.Rauscher@nasa.gov

Eigenvectors





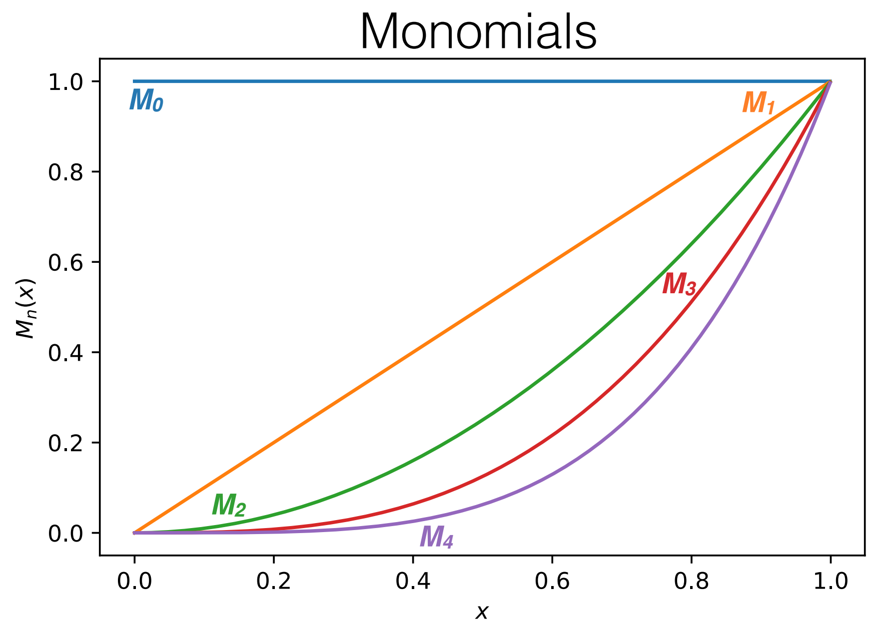




Bernard.J.Rauscher@nasa.gov

What does this tell us?

- The eigenvectors $\mathbf{v}_2$ and higher don't look like monomials. Fitting a polynomial (monomial basis) of order > 1 not ideal



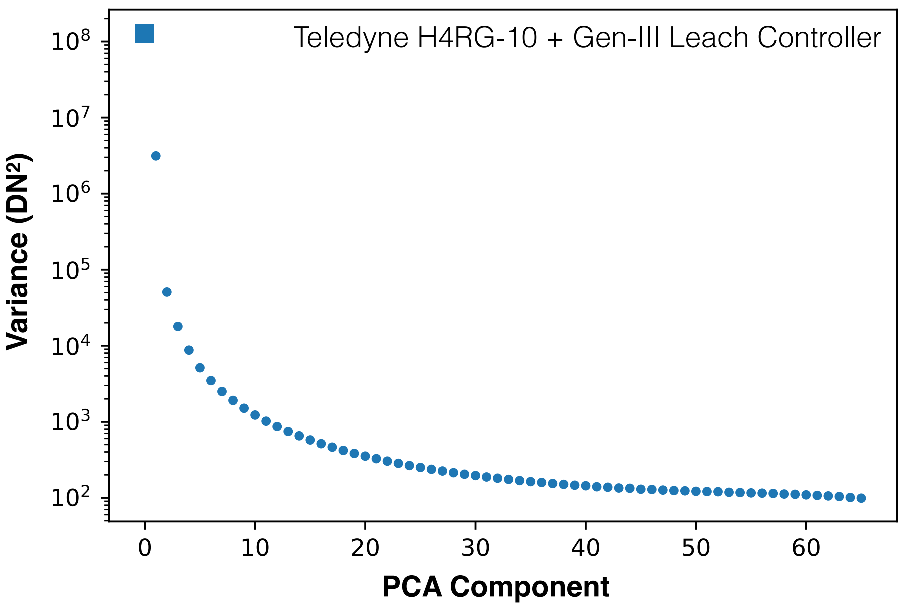
- If one happens to know the eigenvectors for the data set, then one should probably be using them as basis vectors instead of the monomials[1]
  - Eigenvalues directly quantify information content by fit degree
  - Eigenspace provides a linearly uncorrelated representation of the information
- If one doesn't know the eigenvectors, then the **Legendre polynomials** may be a good approximation
  - Coordinates in Legendre space (fit coefficients) approximately quantify information content by fit degree
  - Legendre space provides a much less correlated representation of the information than monomials

[1]May be possible in some cases. *E.g.* might be possible for transiting exoplanets.

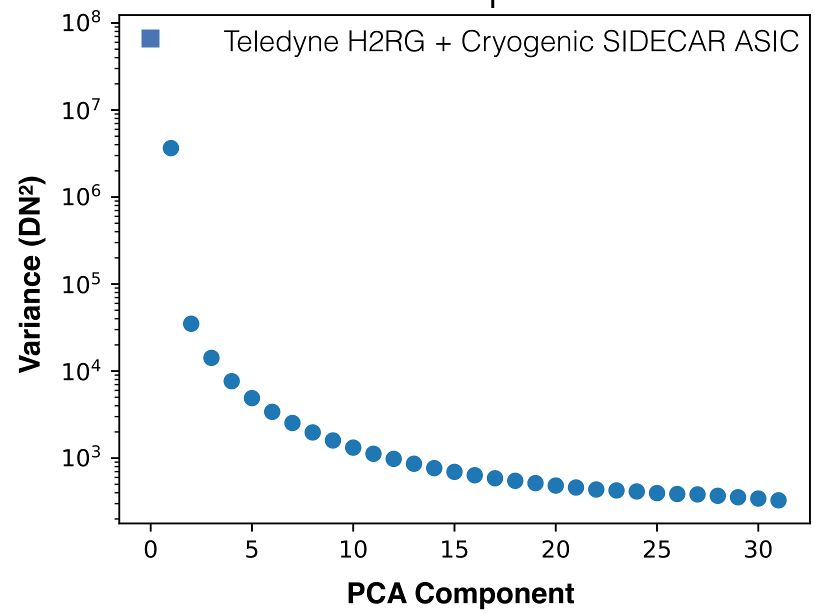
Bernard.J.Rauscher@nasa.gov

Information Content by Component

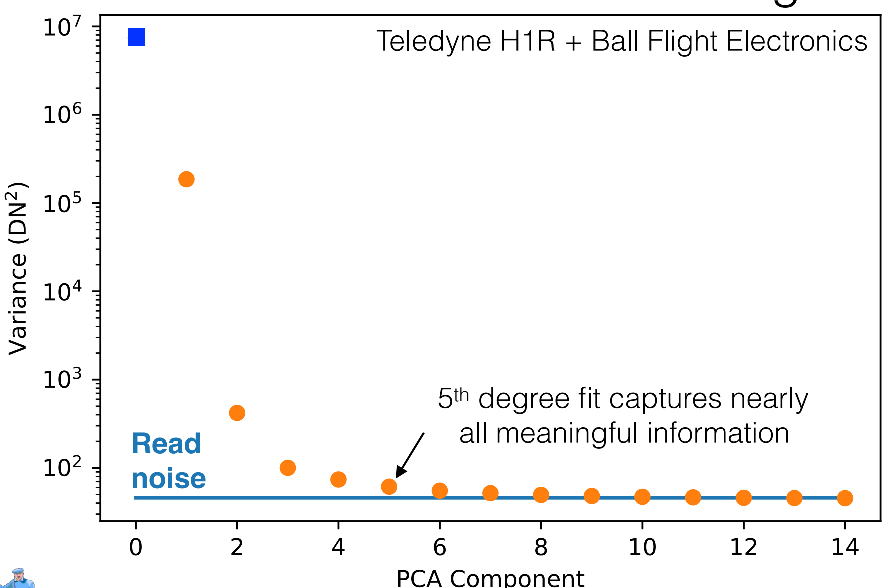
WFIRST Flats



JWST NIRSpec Flats



HST WFC3 Astronomical Images



- PCA0 has a different symbol because it is dominated by the detector's bias pattern
- If one equates variance with information content, then there was always significant information contained in PCA2 and higher for these data
  - Some is just noise
  - But, some is scientifically meaningful and left behind by current practice
- This information is not fully utilized in today
  - Standard practice sequentially "calibrates out" this information (using cal files) before fitting a straight line
  - Linearization followed by line fitting not mathematically equivalent to inferring brightness from all information simultaneously

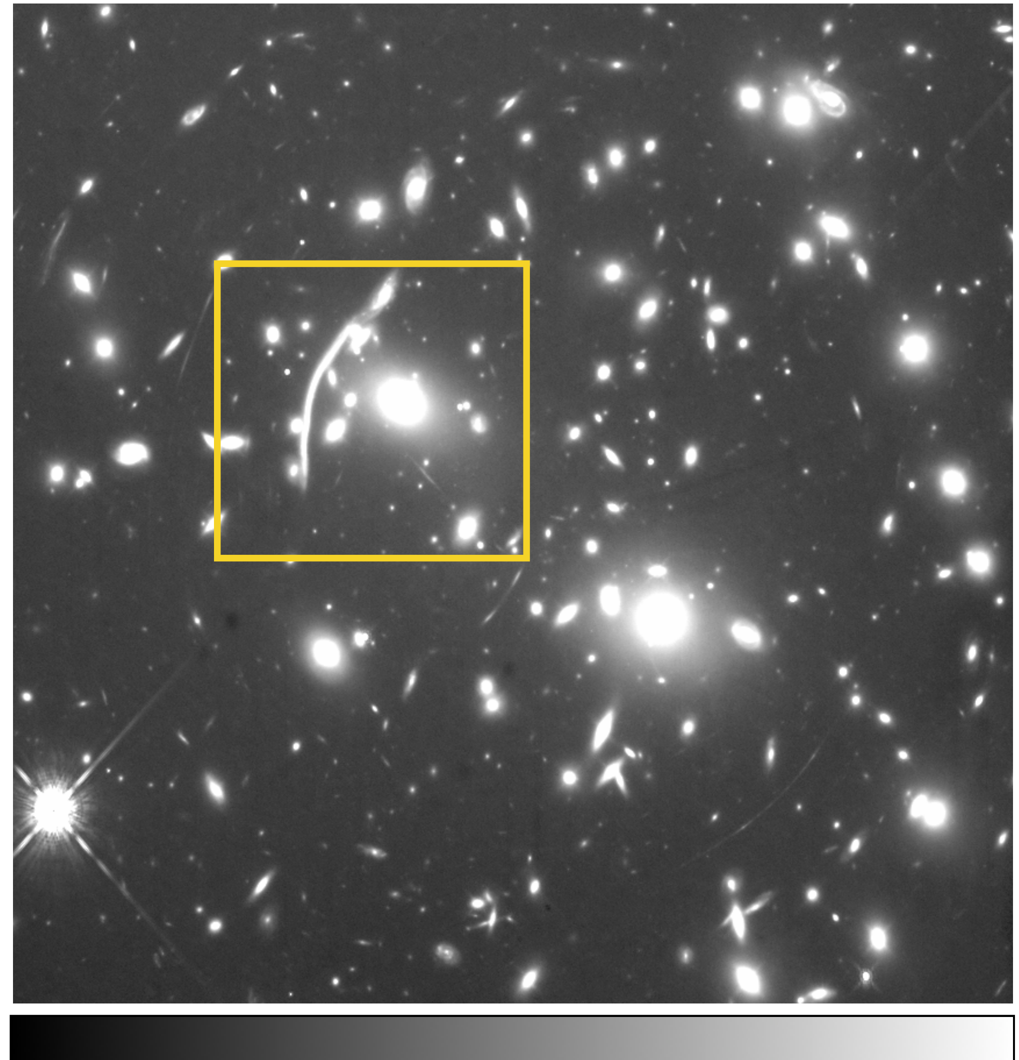
Bernard.J.Rauscher@nasa.gov

Information Content of Astronomical Data

- *HST* WFC3 Abell 370 “Frontier Field” selected as a test case
  - Information rich
  - Pixels values range from sky background to saturation
- 36 dithered exposures downloaded from MAST archive at STScI
  - Filter F160W (λ_c =1.545 μm, FWHM = 0.29 μm)
  - SPARS100 clocking pattern, 15 up-the-ramp samples
  - EXPTIME = 1403 *s*
  - About 14 hours total exposure time
- Uniformly mapped frame index, *z*, to interval $-1 \leq x \leq +1$ and fitted each pixel with Legendre polynomials

$$s(x) = \sum_{i=0}^{5} \lambda_i P_i(x)$$

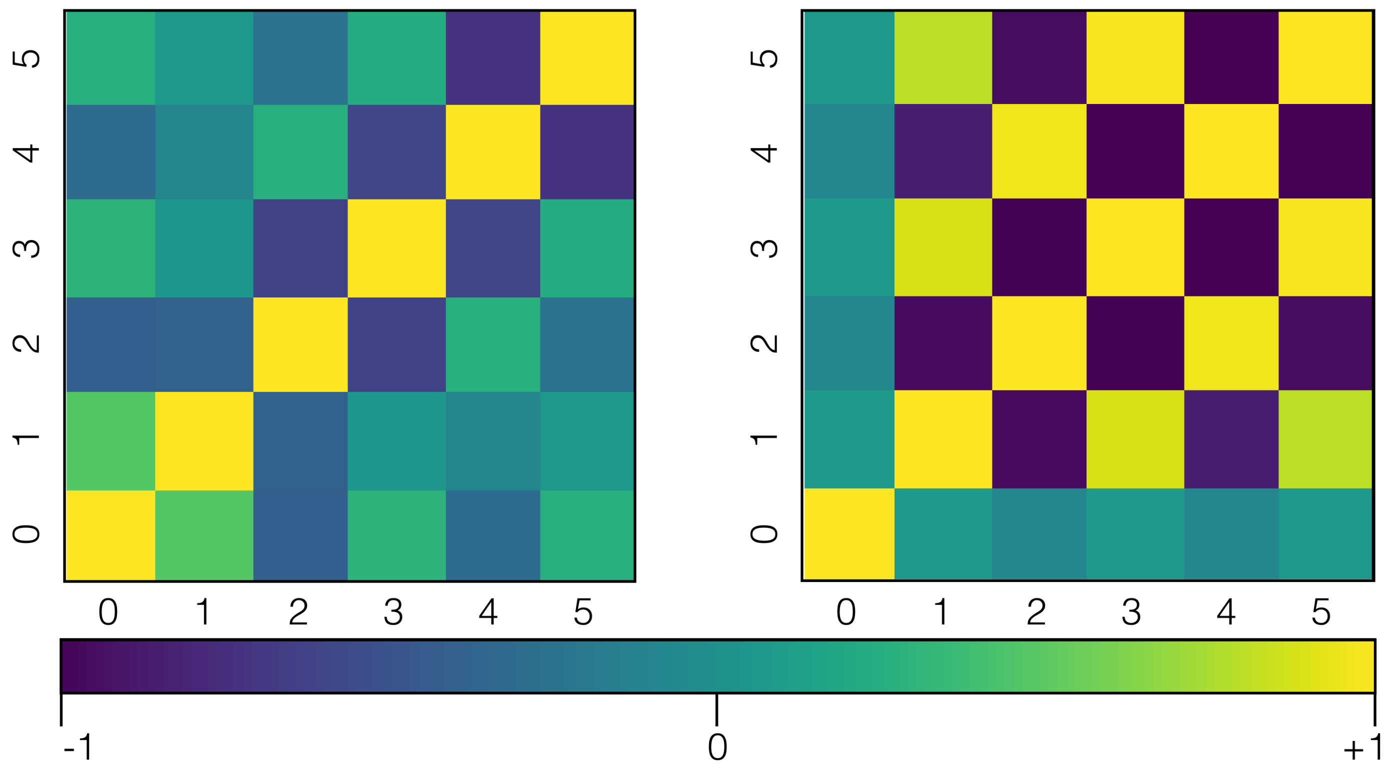
- Image at right is the median λ_1 (slope)
- Yellow box is Region of Interest (ROI) for following charts



Bernard.J.Rauscher@nasa.gov

Legendre Polynomials Provide a Less Correlated Representation

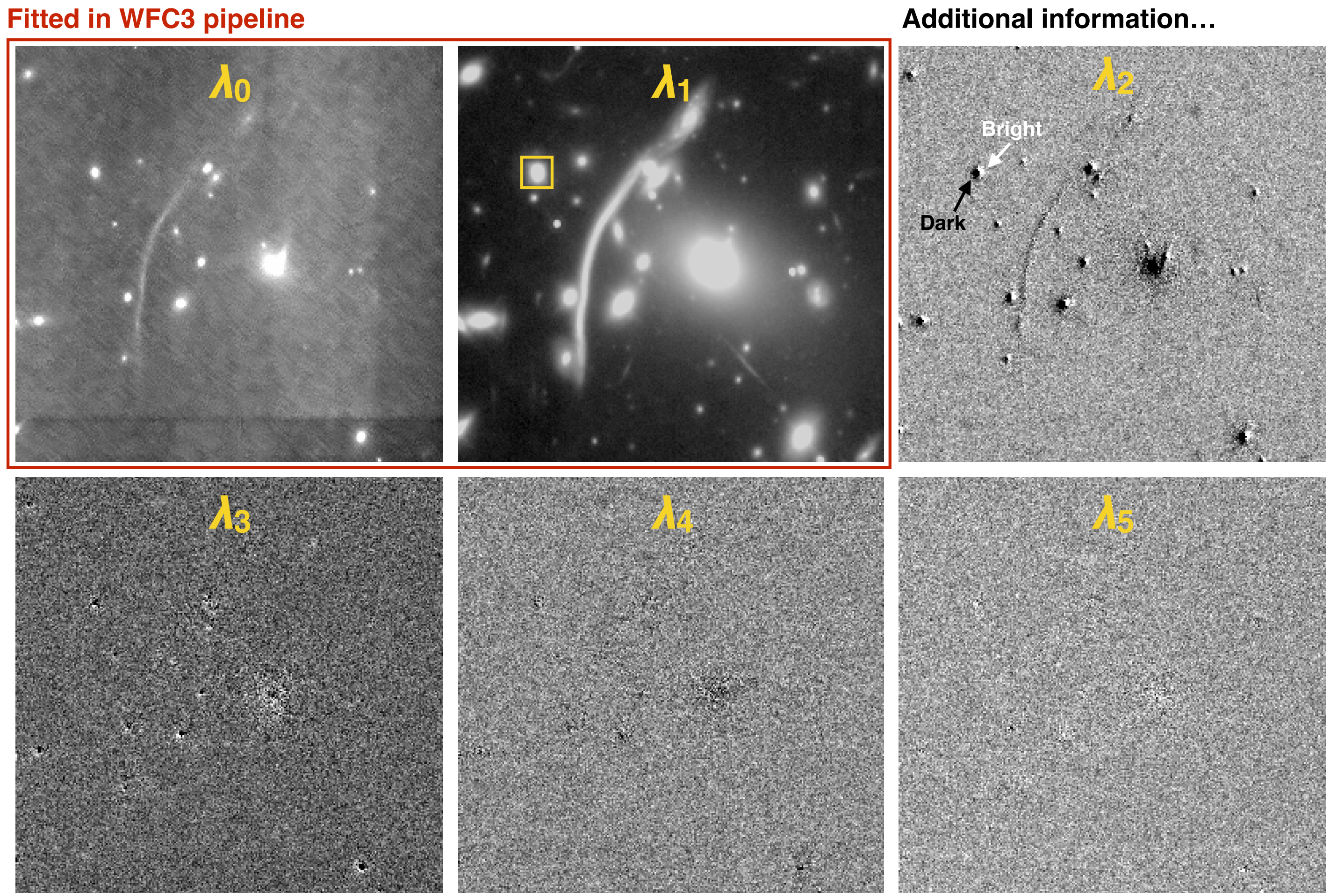
- We fitted the same Abell 370 to 5th degree using Legendre Polynomials and monomials and computed the Pearson correlation matrices
- Legendre polynomials have much less linear correlation
- When doing computations, one could arguably ignore off diagonal terms in Legendre space, whereas one would need to know most of the covariance matrix in monomial space



a) Legendre fit b) Monomial fit

Bernard.J.Rauscher@nasa.gov

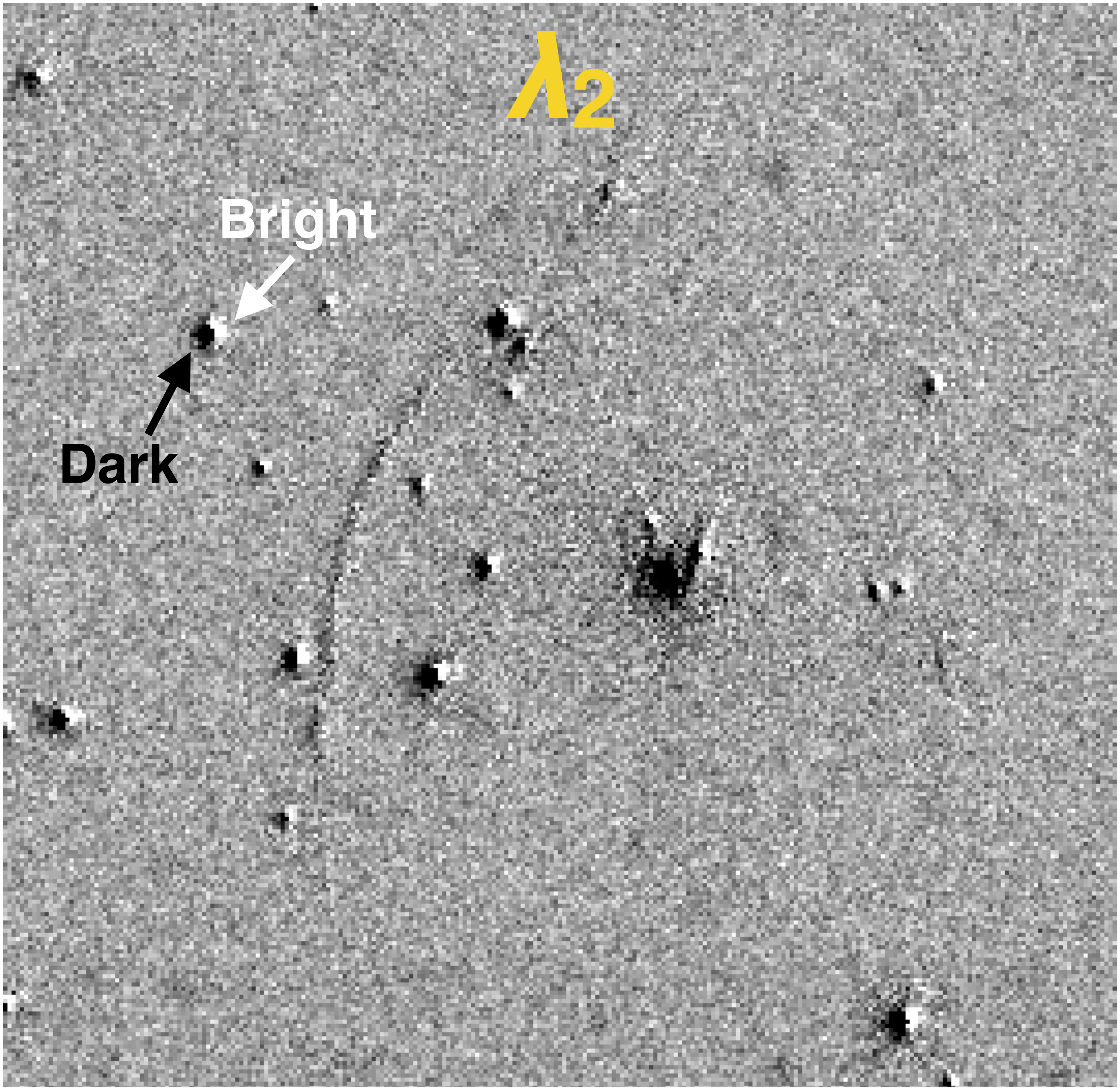
Information Content by Legendre λ



- λ_0 = Familiar bias image
- λ_1 = Familiar slope image
- $\lambda_{3\text{-}5}$ = Additional information, some of which is left behind in current pipeline
- "Bright-Dark" artifacts are interesting. Morphology and appearance are consistent with known 1% of pixel size pointing jitter. Jitter direction does not match FITS headers, but members of WFC3 team at Goddard have told us that jitter directions in FITS headers may not be reliable for this application. In any case, for any real data, there will always be some jitter and this is how it should manifest in λ_2 = slope image.

Bernard.J.Rauscher@nasa.gov

Zooming in on “Bright-Dark” Artifacts



Bernard.J.Rauscher@nasa.gov

Summary

- For a variety of IR detector systems and input data, the Legendre polynomials provided a better basis for modeling up-the-ramp sampled IR pixels than the conventional monomials
  - Legendre polynomials are an orthogonal basis whereas the monomials are not
  - Legendre polynomials approximately diagonalize the covariance matrix whereas the monomials do not
  - Legendre polynomials (approximately) quantify information content by fit degree
  - For situations were the actual eigenvectors are known, it would be even better to use the eigenvectors
- Up-the-ramp sampled IR data contain more information than is captured in today's slope images
  - Pipelines that fit only bias and slope (even with linearization before fitting) are leaving information behind
  - For archiving, if not possible to downlink/save all up-the-ramp samples, then we recommend downlinking/archiving a few Legendre coefficients as a compromise to capture more information. Downlinking/archiving only slope images leaves information behind
- Ongoing work includes studying new calibration approaches in Legendre space. We look forward to saying more about this as we learn more.

Bernard.J.Rauscher@nasa.gov