BigBOSS Data Reduction Software

Adam S. Bolton
The University of Utah
Department of Physics & Astronomy

NOAO BigBOSS Community Workshop 2011

The Premise

BigBOSS data reduction software is as important as BigBOSS data collection hardware to the scientific success of the project, both for the key cosmology survey and for the wider astronomical community.

The BigBOSS Perspective

- Significant experience in developing, operating, and maintaining survey-scale spectroscopic data-reduction pipelines (SDSS, BOSS)
- Full appreciation of scale and importance of software challenge
- Incremental reviews of software as well as hardware
- Close integration with instrument hardware teams
- Strong incentive and interest in “doing it right”, both for BigBOSS key cosmological survey requirements, and because it’s the right thing to do!

The NOAO Perspective

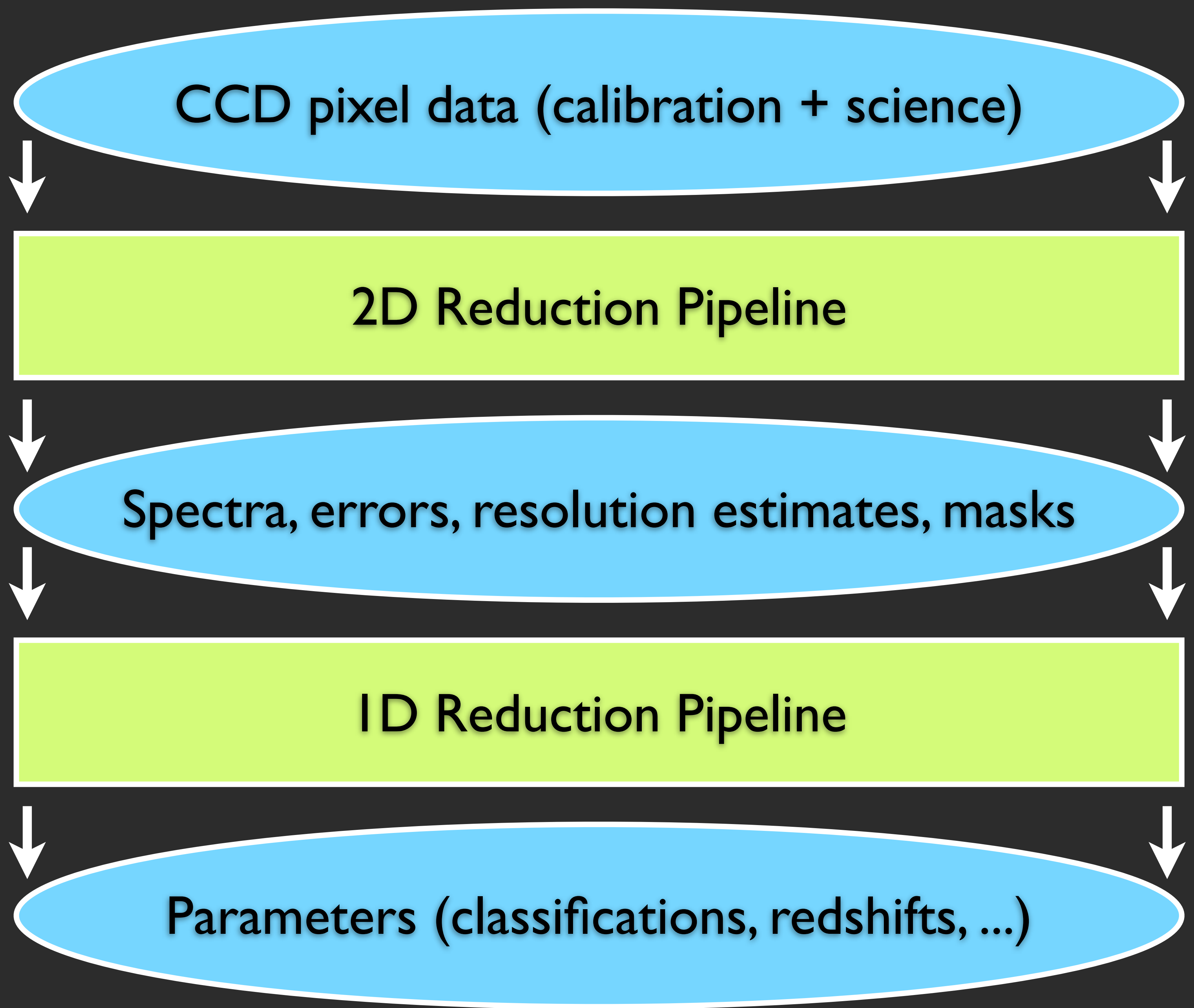
- BigBOSS data reduction pipeline to be developed collaboratively between BigBOSS team and NOAO
- “One pipeline” policy: not one for the cosmological survey team, one for the rest of the world
- Community software requirements input essential, both up front and ongoing
- Support for possibility of incorporating contributed software modules and derived data products (...?)
- Data archiving, reduction, and serving to be made possible on NOAO-scale computer hardware systems

Charge for today and tomorrow

Calibration, data-reduction and data-product requirements from the astronomical community will be core drivers of BigBOSS pipeline development

=> Outcome from this workshop is the most significant first step in establishing what these requirements are / will be.

Data-reduction Process Schematic



Department of Physics and Astronomy
ASTRONOMY AND ASTROPHYSICS
THE UNIVERSITY OF UTAH

Adam S. Bolton

Schematic of Requirements

Dark Energy
Requirements

Flat-Fielding
Spectrophotometry
Cosmetic Handling
Wavelength Accuracy
Sky Subtraction
LSF Accuracy
Noise Estimators
Pixel Covariance
Bitmasks
Etc. Etc.

Schematic of Requirements

Dark Energy Requirements

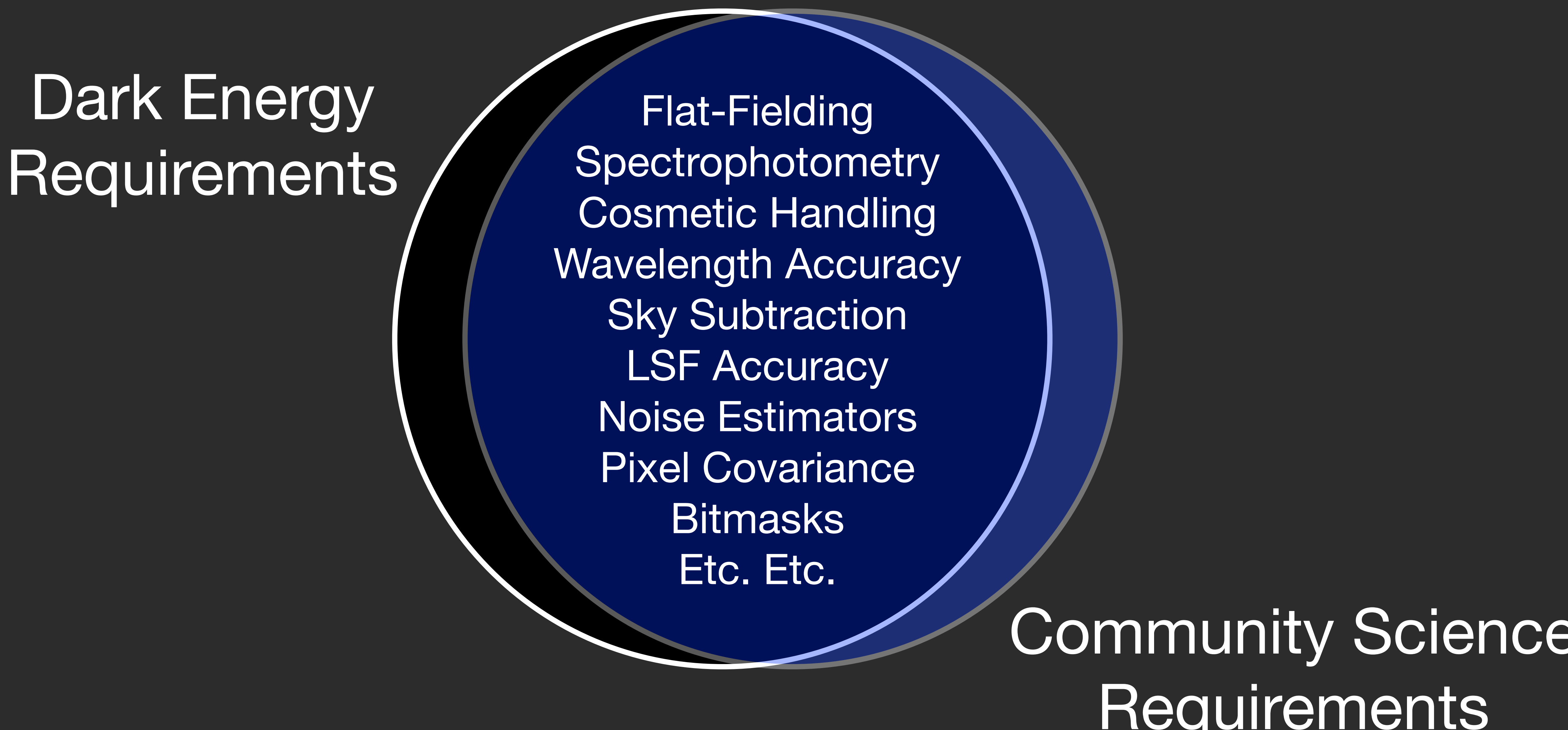
Flat-Fielding
Spectrophotometry
Cosmetic Handling
Wavelength Accuracy
Sky Subtraction
LSF Accuracy
Noise Estimators
Pixel Covariance
Bitmasks
Etc. Etc.

Flat-Fielding
Spectrophotometry
Cosmetic Handling
Wavelength Accuracy
Sky Subtraction
LSF Accuracy
Noise Estimators
Pixel Covariance
Bitmasks
Etc. Etc.

Community Science Requirements

Schematic of Requirements

Overlap likely to be substantial (but not complete)

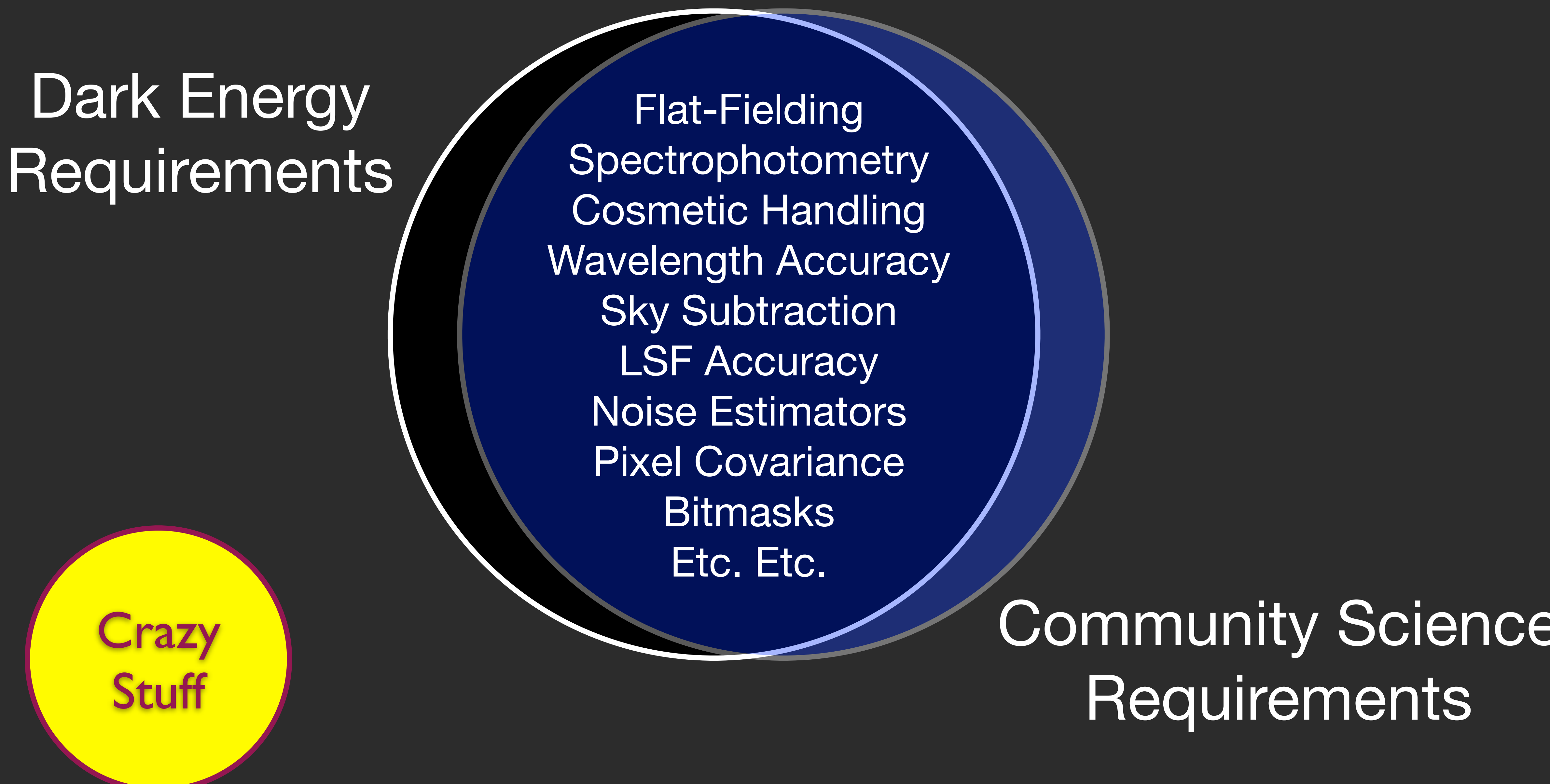


Department of Physics and Astronomy
ASTRONOMY AND ASTROPHYSICS
THE UNIVERSITY OF UTAH

Adam S. Bolton

Schematic of Requirements

Overlap likely to be substantial (but not complete)



Requirements to think about

Spectrophotometry:

- relative
- absolute
- PSF versus fiber versus...?

Wavelength / RV precision and accuracy:

- relative between spectra
- relative across time
- absolute

Accuracy of error estimates

Accuracy of line-spread function estimates

Sky subtraction quality

Informational mask bits

Performance in high-SNR regime

Requirements to think about

Access to raw data (science, calibration)

Reduction software
- as service (for archive and synchronous users)
- as product (for PI project users)
- as project (new capability development)

Support for flexibility of observation modes

Calibration analysis tasks/functions/modules

Requirements to think about

Access to raw data (science, calibration)

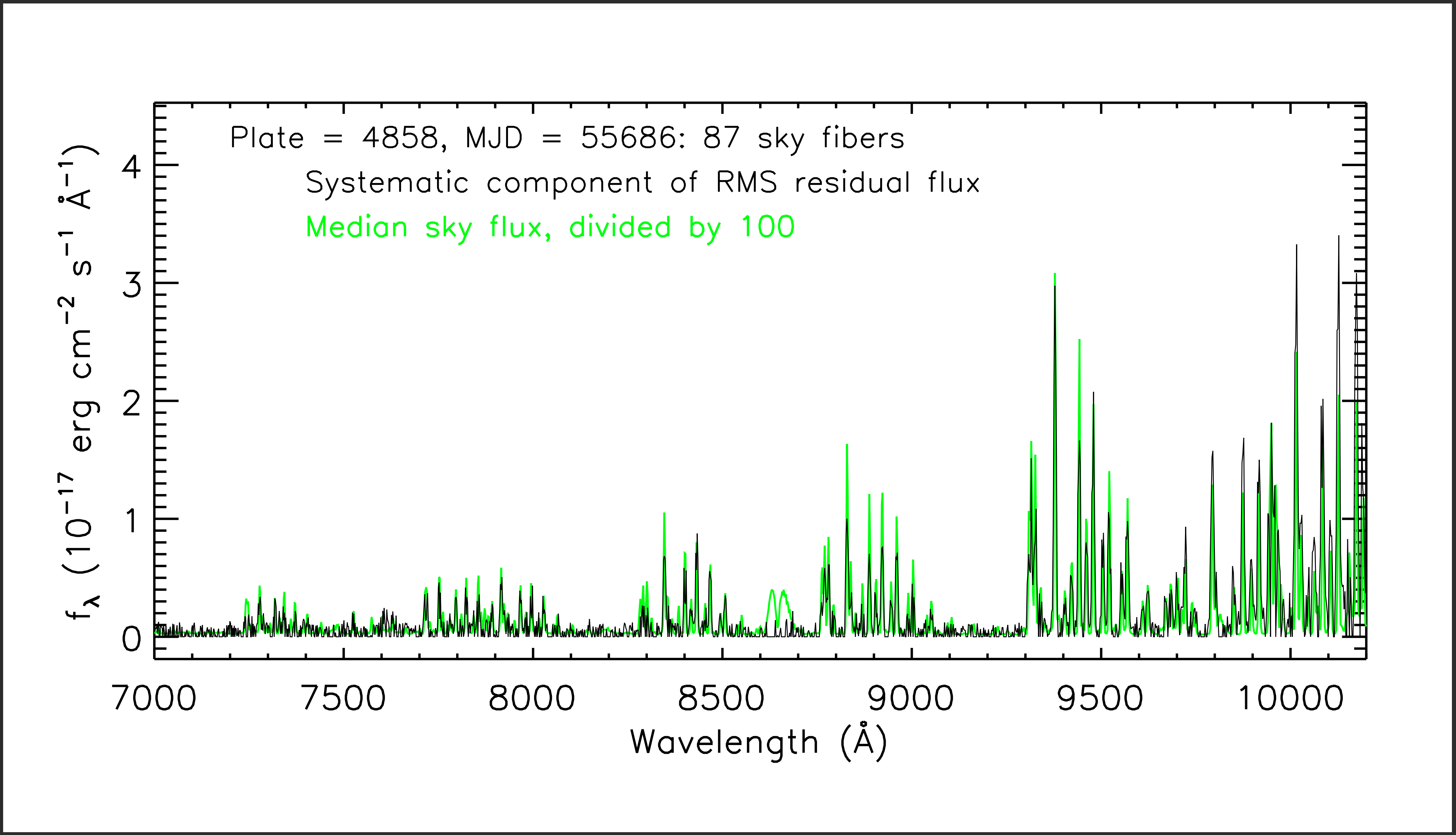
Reduction software
- as service (for archive and synchronous users)
- as product (for PI project users)
- as project (new capability development)

Support for flexibility of observation modes

Calibration analysis tasks/functions/modules

+ other items!

Current BOSS Sky Subtraction Quality



(algorithm based on Horne 1986 style row-by-row extraction)

Next-Generation Extraction

Full 2D model to raw data via "calibration matrix"

A

Generalizes and incorporates:

- Trace solution
- Wavelength solution
- 2D spectrograph PSF and its variation
- Relative and absolute throughput variation
- CCD pixel sensitivity variations
- Etc.

Determination of A poses significant challenges to hardware stability, calibration data, and algorithms

Bolton & Schlegel 2010

Next-Generation Extraction

Full 2D model to raw data via "calibration matrix" A

(CCD pixel counts) = ***A*** (input spectrum counts) + (noise)

$$\chi^2 (\boldsymbol{m} \mid \text{raw data}) = (\boldsymbol{p} - \boldsymbol{A}\,\boldsymbol{m})^T \boldsymbol{N}^{-1} (\boldsymbol{p} - \boldsymbol{A}\,\boldsymbol{m})$$

↓

$$\chi^2 (\boldsymbol{m} \mid \text{extracted spectra}) = (\boldsymbol{f} - \boldsymbol{R}\,\boldsymbol{m})^T \boldsymbol{C}^{-1} (\boldsymbol{f} - \boldsymbol{R}\,\boldsymbol{m})$$

f = extracted spectrum
R = band-diagonal line-spread function matrix
C = diagonal spectrum covariance matrix

Same χ^2 in either case!

Bolton & Schlegel 2010

Next-Generation Extraction

Advantages:

- Extraction as lossless compression
- Incorporates explicit model of 2D data
- Poisson-limited sky subtraction
- Data products “look & feel like spectra”

Major Concern:

- Increased computing requirements

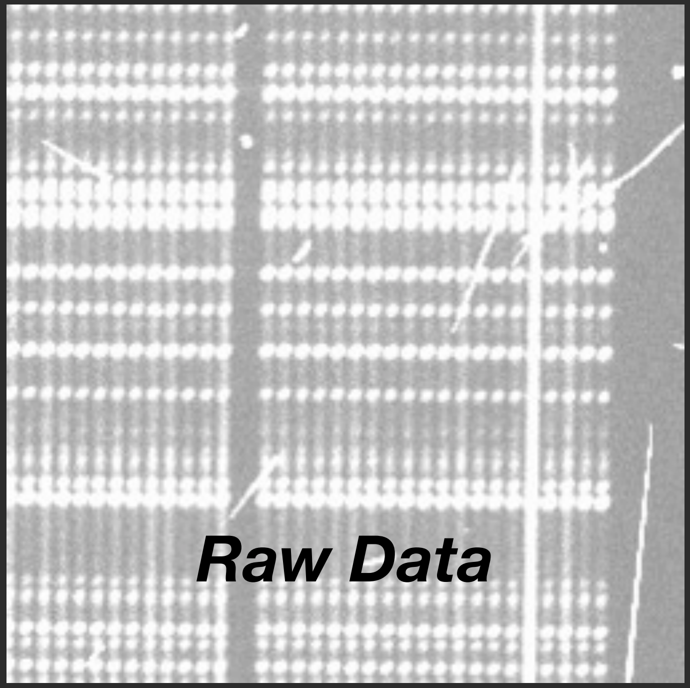
Bolton & Schlegel 2010

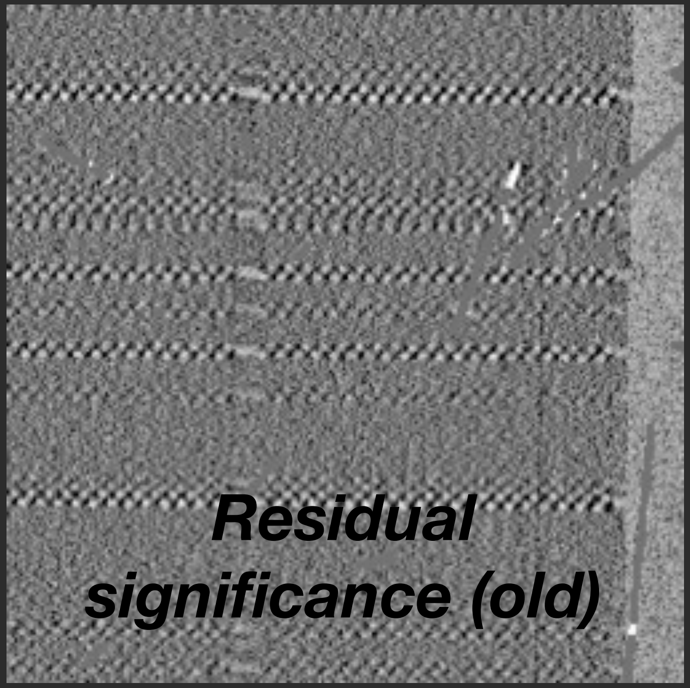
Next-Generation Extraction

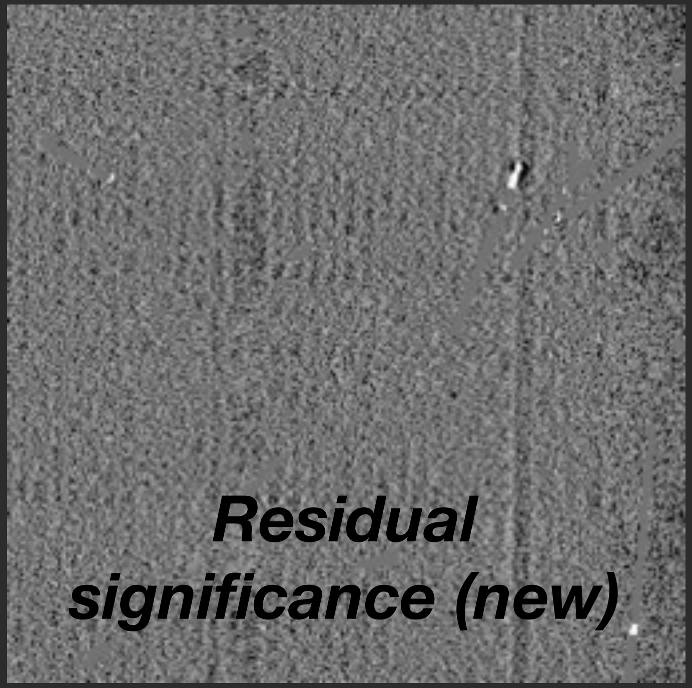
Bolton & Schlegel 2010

Path forward (demonstrated summer 2011):

- Decompose among bundles, exposures, spectrographs, and wavelength ranges







=> 2016 computer system requirements:

~ $1.2M for BigBOSS survey team

~ $100k for NOAO production use

Department of Physics and Astronomy
ASTRONOMY AND ASTROPHYSICS
THE UNIVERSITY OF UTAH

Adam S. Bolton

Charge for today and tomorrow

Calibration, data-reduction and data-product requirements from the astronomical community will be core drivers of BigBOSS pipeline development

=> Outcome from this workshop is the most significant first step in establishing what these requirements are / will be.

Thanks!

Questions?

Discussion?