MS-DESI Spectroscopic Extraction & Simulation

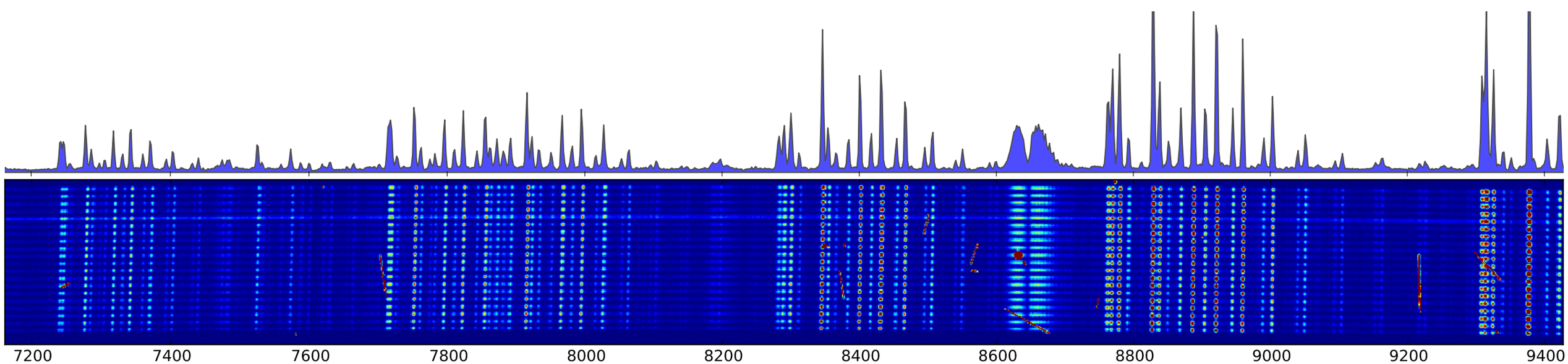
Stephen Bailey

with Adam Bolton, Ted Kisner, David Schlegel (extraction)

and Nick Mostek, Pat Jelinsky (simulation)

+ the BOSS spectroscopic pipeline team

5 March 2013



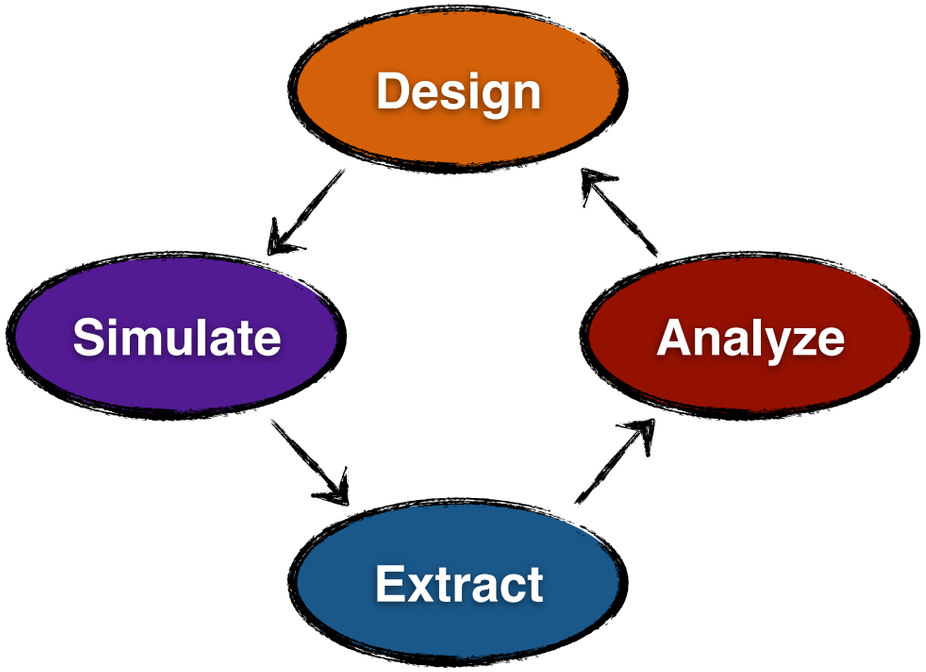
Extraction & Simulation

Spectroscopic Extraction

- CCD electrons –> Spectra
- Optimize the algorithms to get the most out of the data

Spectroscopic Simulation

- Spectra –> CCD electrons
- Verify instrument designs
- Extraction algorithm development
- Guides instrument, operations design

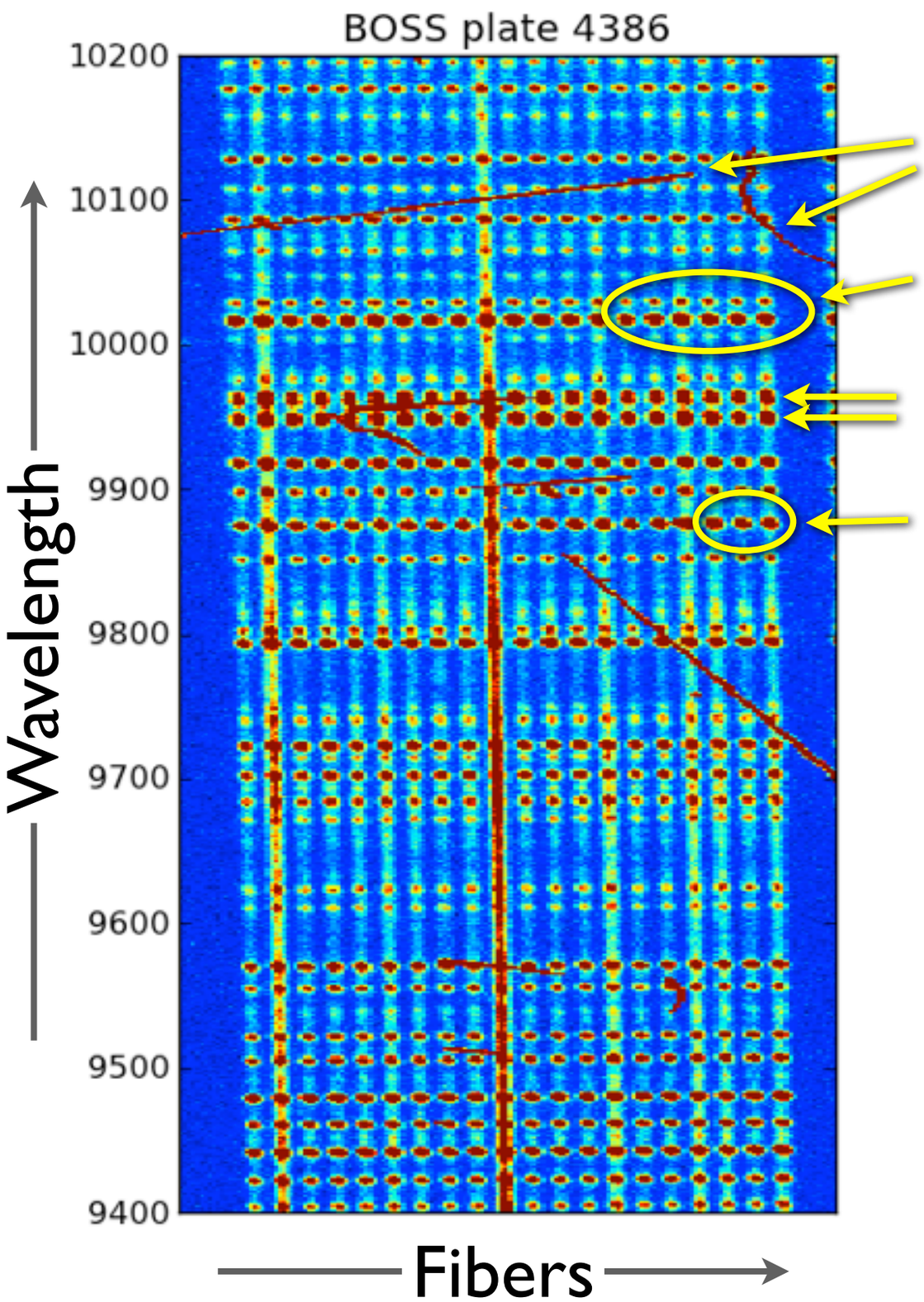


Beyond Spectra

- Spectra –> classifications, redshifts, papers...
  - Better quality spectra = better quality results

Real Data

Can you find the ELG?



Cosmic Rays, Noise

Overlapping spectra

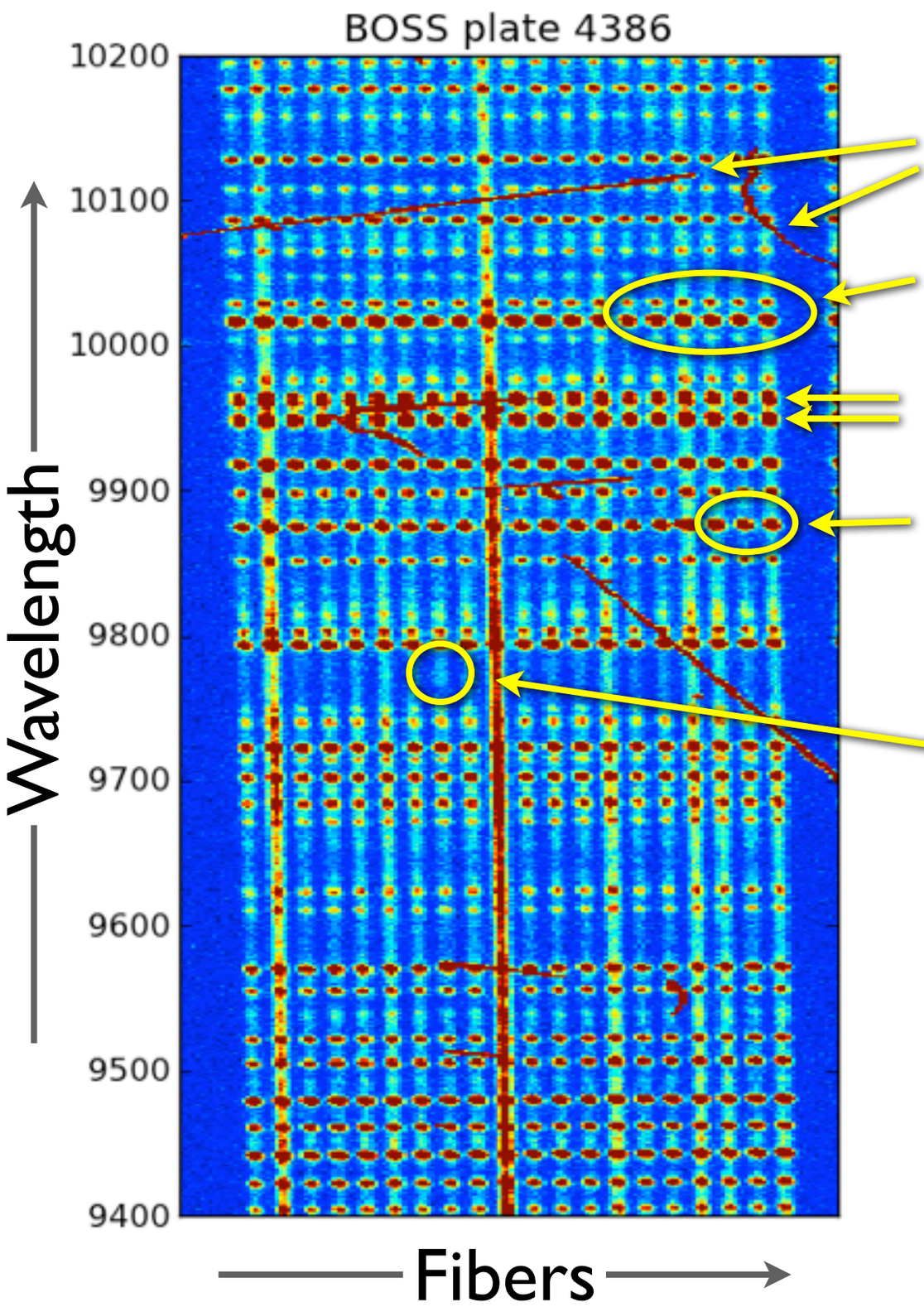
Mostly Sky

Lumpy PSFs

The task of the spectroscopic pipeline is to find the ELGs in this data

Real Data

Can you find the ELG?

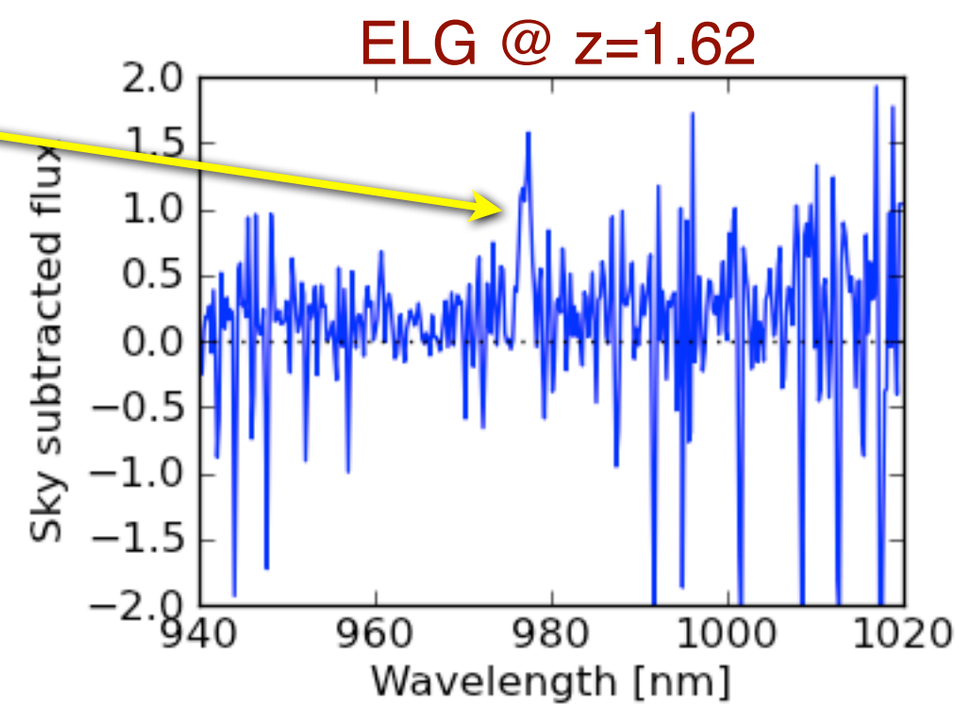


Cosmic Rays, Noise

Overlapping spectra

Mostly Sky

Lumpy PSFs



Ingredients of a Spectrum

Optimize each of these to maximize information extracted from the data.

Flux vs. wavelength

Error Model

- Off-diagonal covariance is often ignored
- Full covariance –> diagonalized covariance
- Mathematically useful (i.e. correct)

Instrument resolution

- Gaussian Line Spread Function (LSF) –> Full resolution matrix

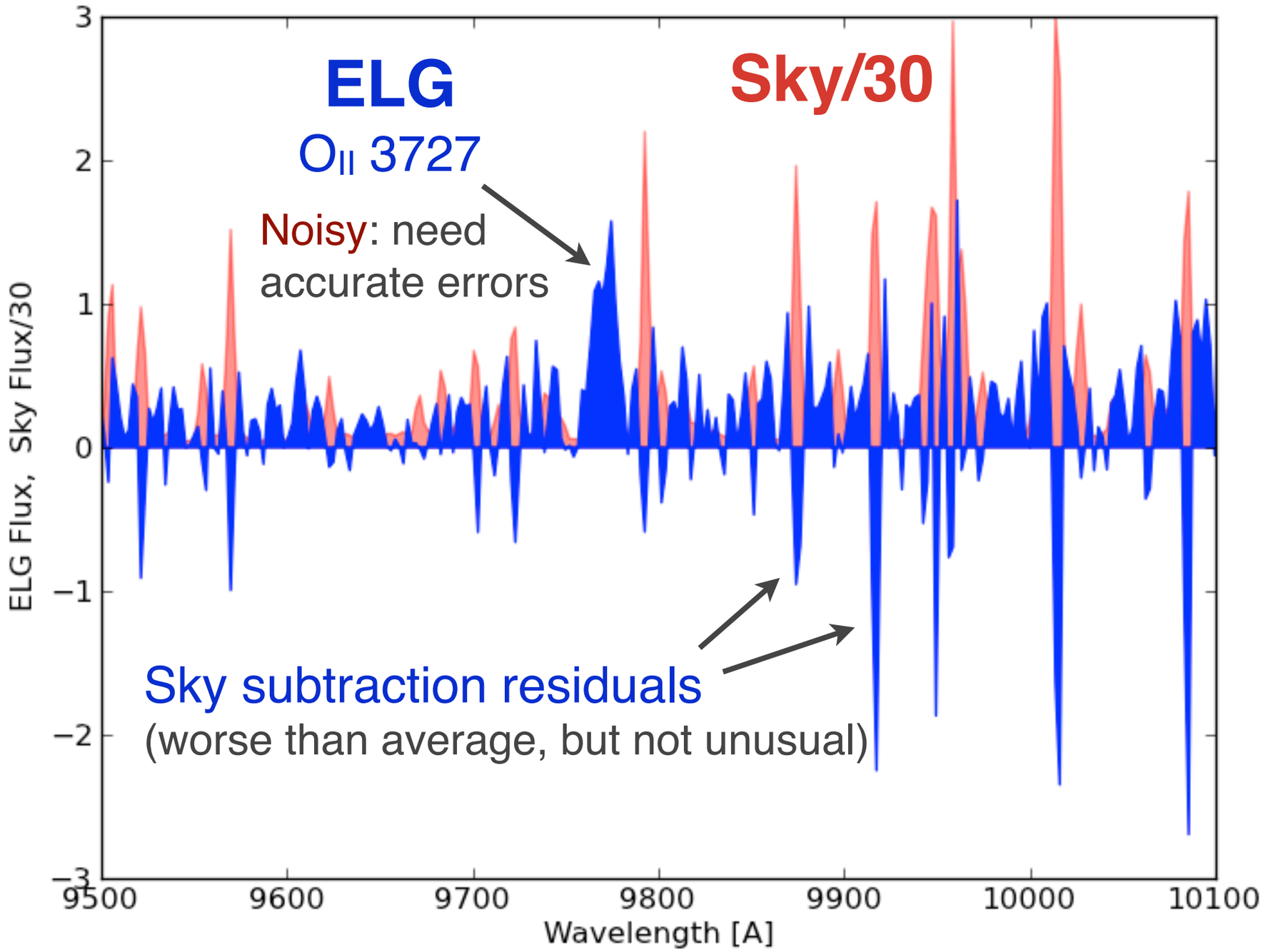
Sky subtraction (remember that real data?)

- Poisson limited
- Complication: sky varies in space and time

Coaddition: combining multiple exposures into one

- Full error propagation without introducing covariance

Improvements Needed

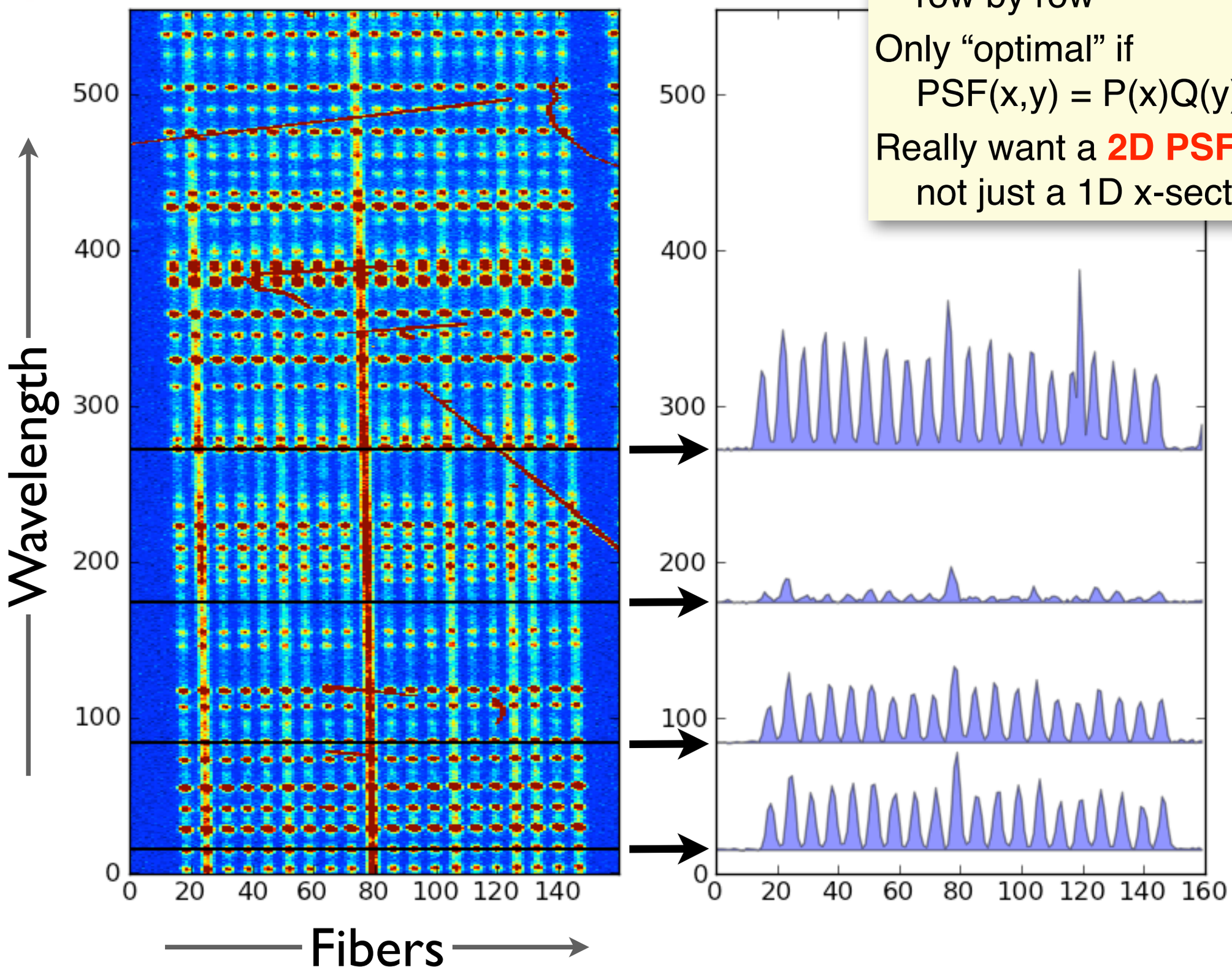


“Optimal” Extraction

Fit overlapping Gaussians row by row

Only “optimal” if PSF(x,y) = P(x)Q(y)

Really want a **2D PSF**, not just a 1D x-section



Comparison

What	Standard	Better
PSF	1D row-by-row	2D
Wavelength grid	Tied to CCD pixels	Arbitrary
Error propagation	Only valid if PSF(x,y) = P(x)Q(y)	Correct
Coaddition	Misaligned **λ**-grid –> covariance	Purely diagonal
Sky Subtraction	Extract; Model; Subtract –> Covariance	Can be integrated with extraction
Resolution	Typically Gaussian Line Spread Function	Full resolution matrix

Comparison

What	Standard	Better
PSF	1D row-by-row	2D
Wavelength grid	Tied to CCD pixels	Arbitrary
Error propagation	Only valid if PSF(x,y) = P(x)Q(y)	Correct
Coaddition	Misaligned λ-grid –> covariance	Purely diagonal
Sky Subtraction	Extract; Model; Subtract –> Covariance	Can be integrated with extraction
Resolution	Typically Gaussian Line Spread Function	Full resolution matrix

Bolton & Schlegel 2010
"Spectroperfectionism"

Spectroperfectionism in a Nutshell

2D PSF vs. spectrum and λ

$$p = Af + \text{noise}$$

Spectral flux –> CCD pixels

$$f = (A^T N^{-1} A)^{-1} A^T N^{-1} p$$

Solve for f

$$C_f = (A^T N^{-1} A)^{-1}$$

Covariance of f

$$\tilde{C} = R C_f R^T$$

Diagonalize C_f

$$\tilde{f} = R f$$

$\tilde{f}$ has diagonal error

–> *Resolution equivalent to 1D spectrograph*

Spectroperfectionism in a Nutshell

2D PSF vs. spectrum and λ

$$p = Af + \text{noise}$$

Spectral flux –> CCD pixels

$$f = (A^T N^{-1} A)^{-1} A^T N^{-1} p$$

Solve for f

$$C_f = (A^T N^{-1} A)^{-1}$$

Covariance of f

$$\tilde{C} = R C_f R^T$$

Diagonalize C_f

$$\tilde{f} = R f$$

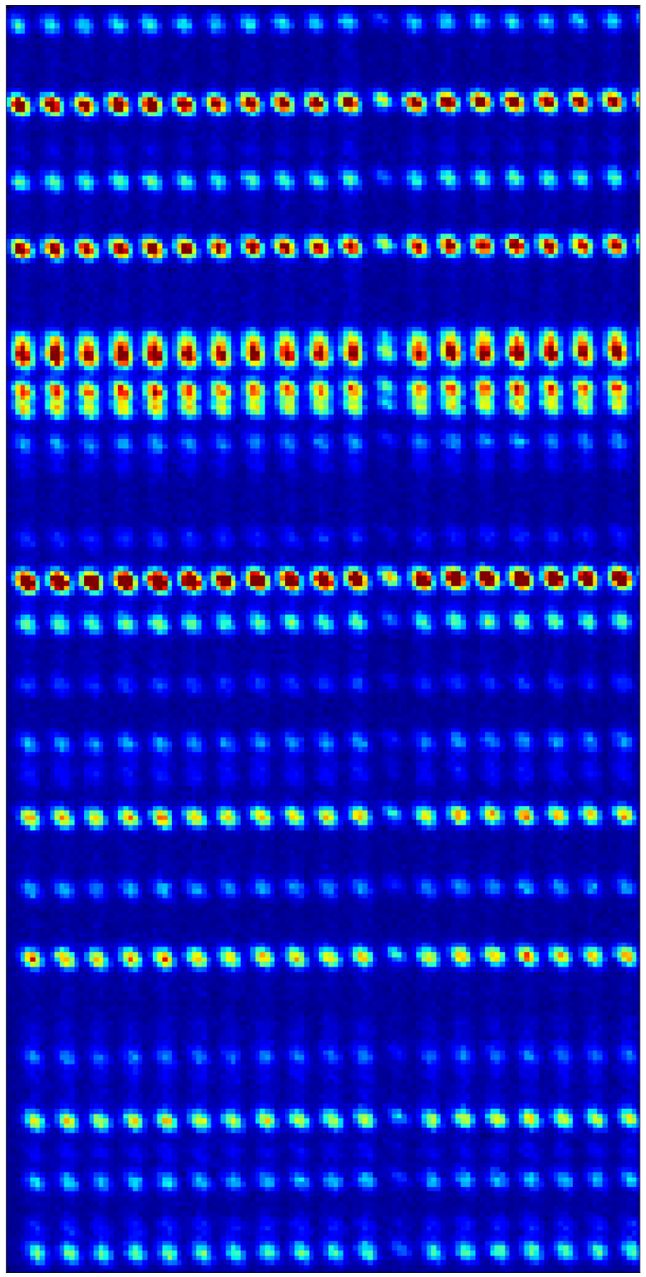
$\tilde{f}$ has diagonal error

–> *Resolution equivalent to 1D spectrograph*

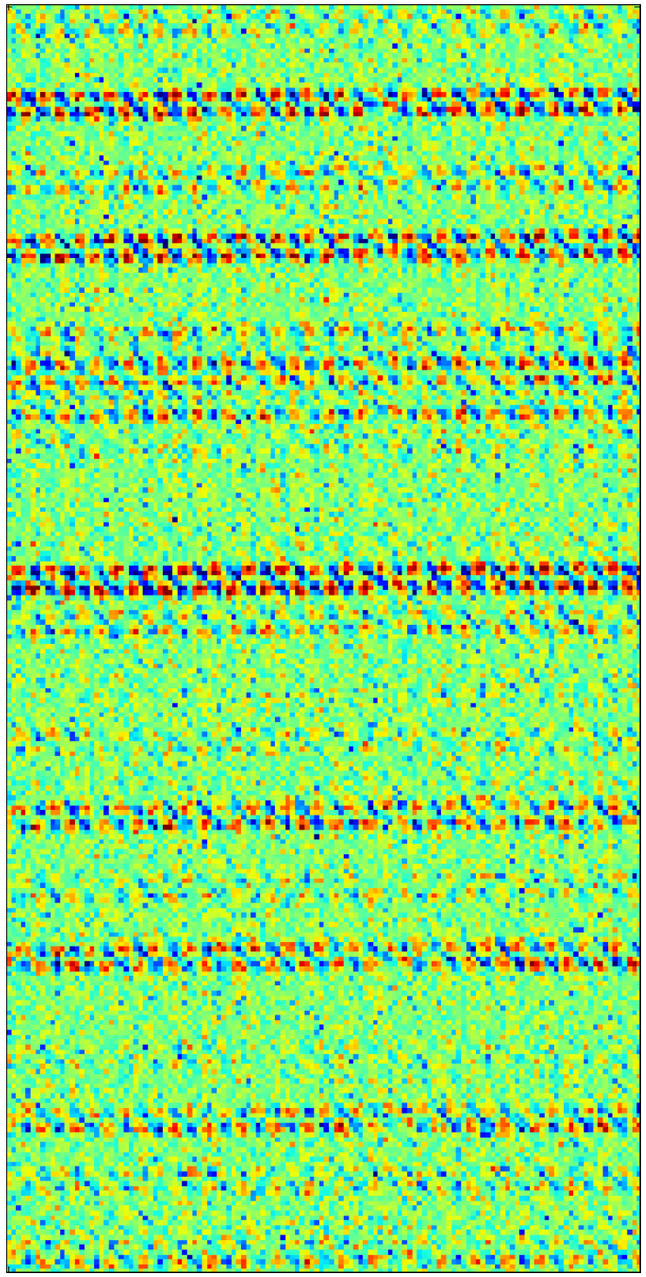
Three extraction products:

1. Spectrum (resolution convolved)
2. Resolution Matrix
3. Error model (*diagonalized* covariance)

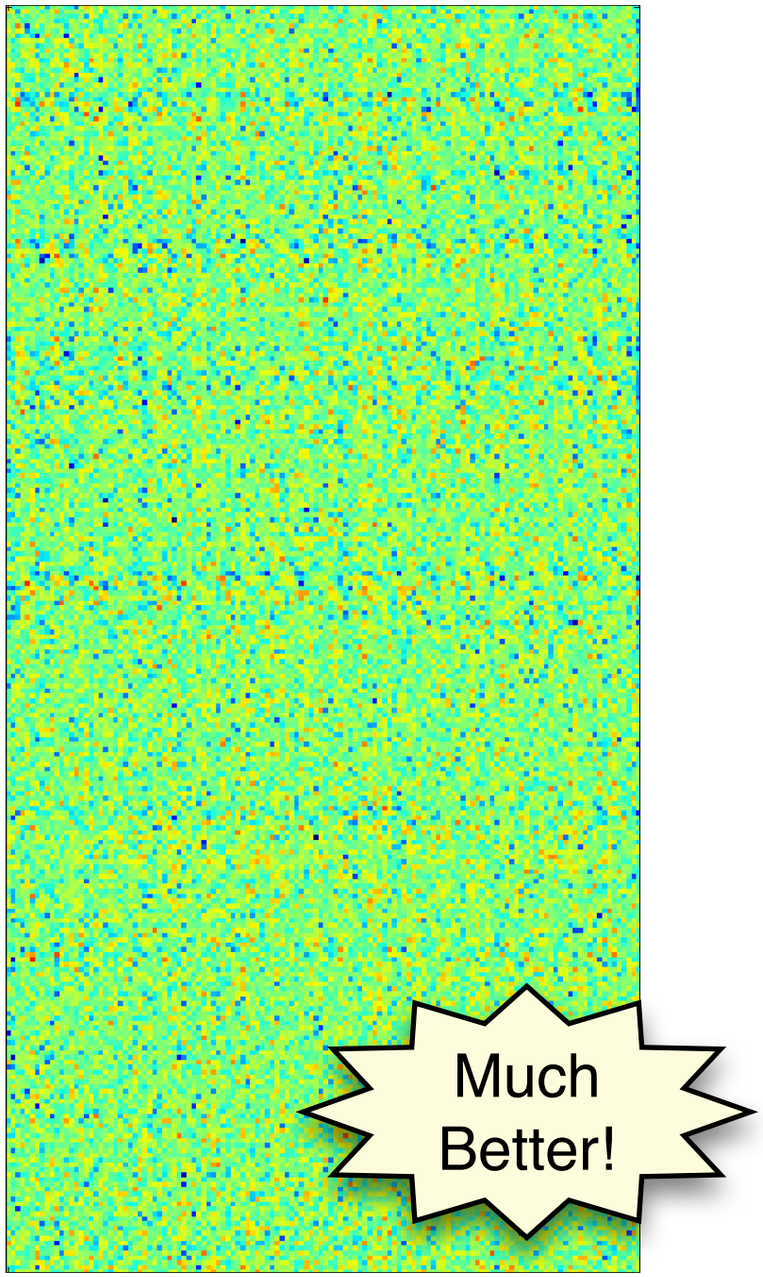
Extraction Residuals



Original Data

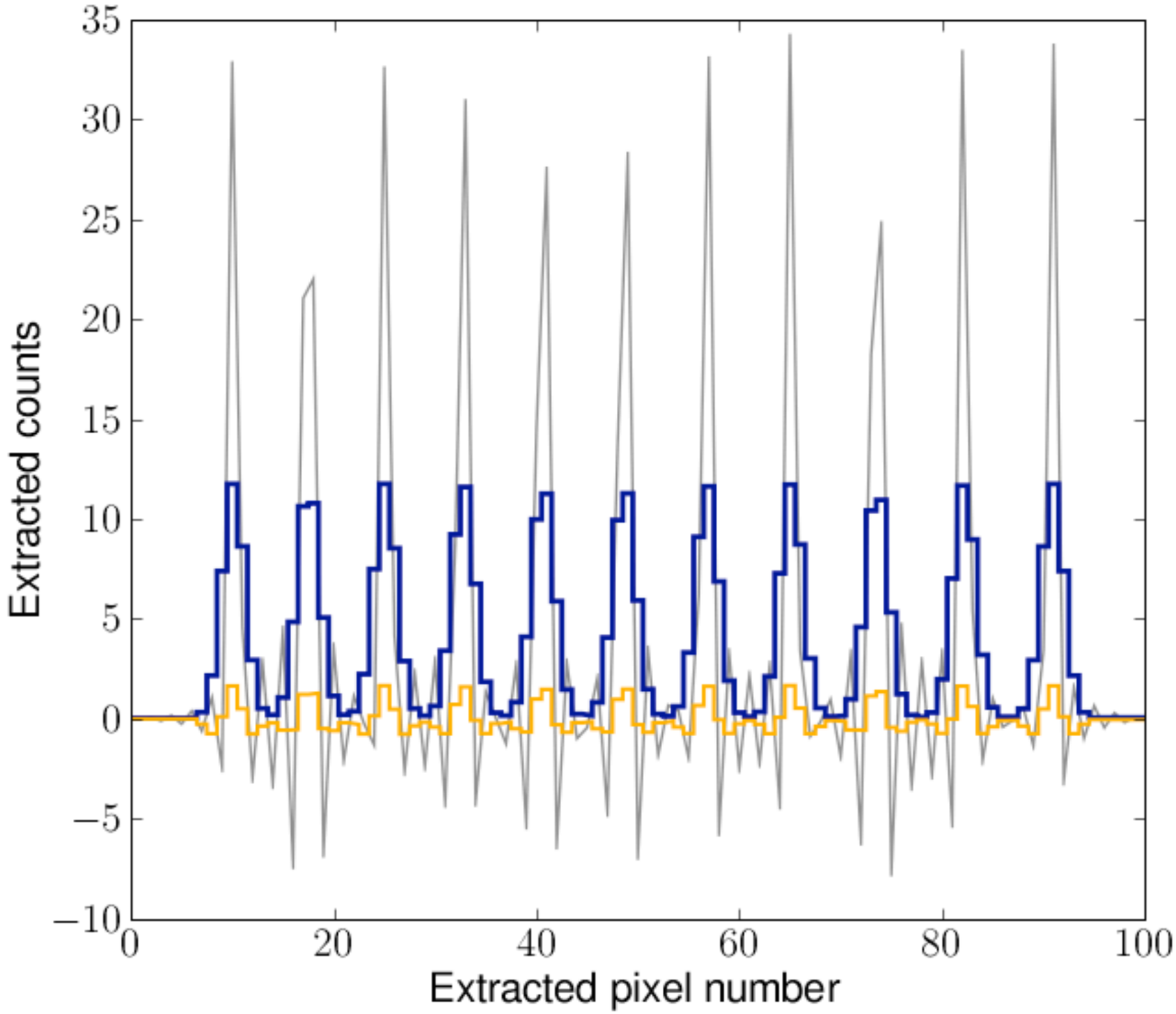


"Optimal" Residuals



Spectroperfectionism Residuals

Extracted Spectra



Practical Perfectionism

Matrices too big in naive implementation

- Must develop optimizations

Several R&D implementations in the works

- Python "divide-and-conquer"
  - Split problem into small overlapping regions
  - Extract with edge effects
  - Stitch together central regions away from edge effects
- C++ MPI parallelized
  - Can extract much larger chunks
  - Expands algorithmic options
- HETDEX has also ported this algorithm for their extraction pipeline

Single-node multi-core parallelization

Multi-node parallelization: can scale to supercomputers

Testing on real data

- BOSS data keeps us focused on real world issues
- Simulations too (for BOSS, MaNGA, MS-DESI)

Integrated Steps

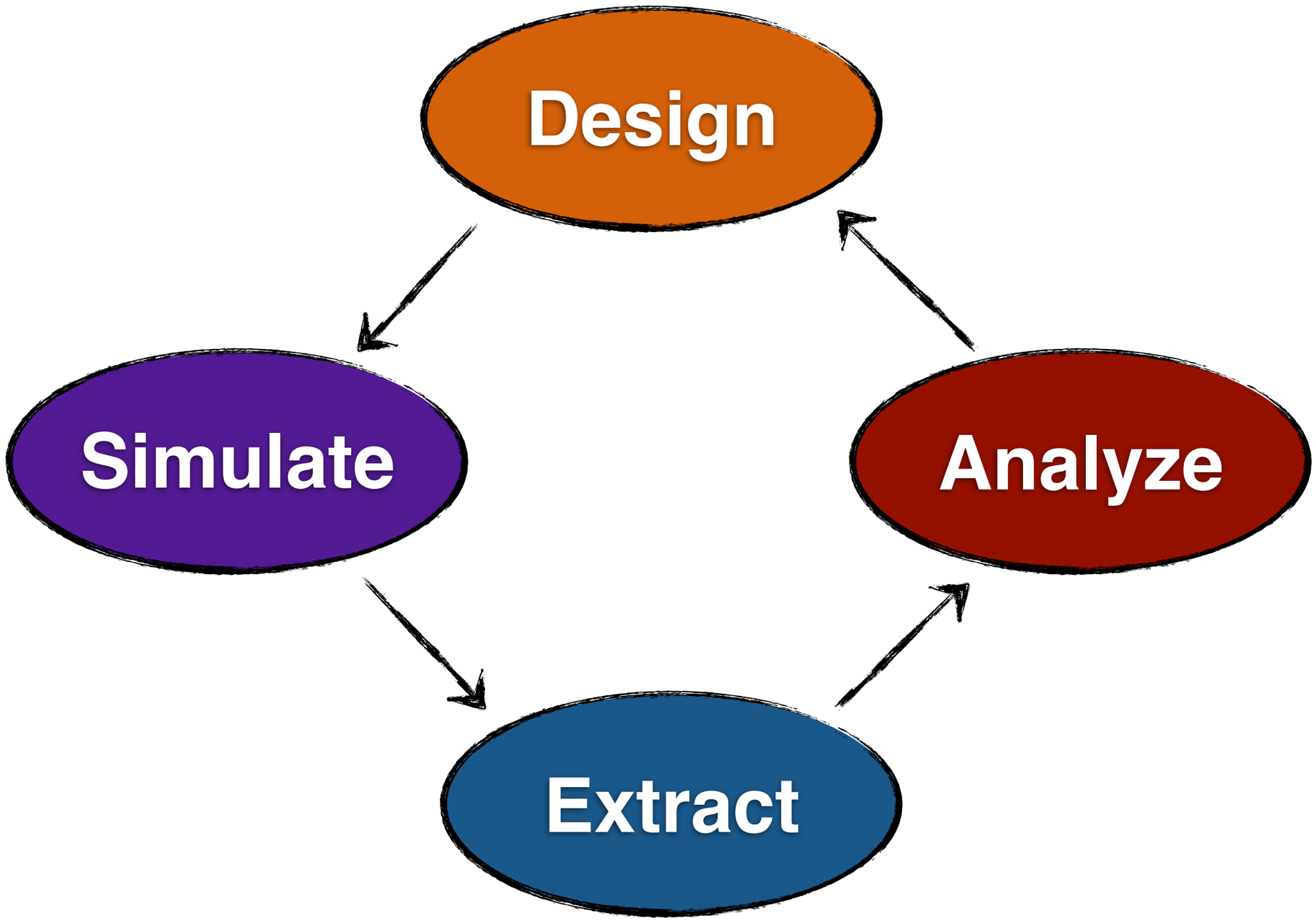
$$p = A(f_{\mathrm{obj}} + f_{\mathrm{sky}}) + \mathrm{noise}$$

Pixels from multiple exps

All fibers:
– some obj+sky
– some just sky

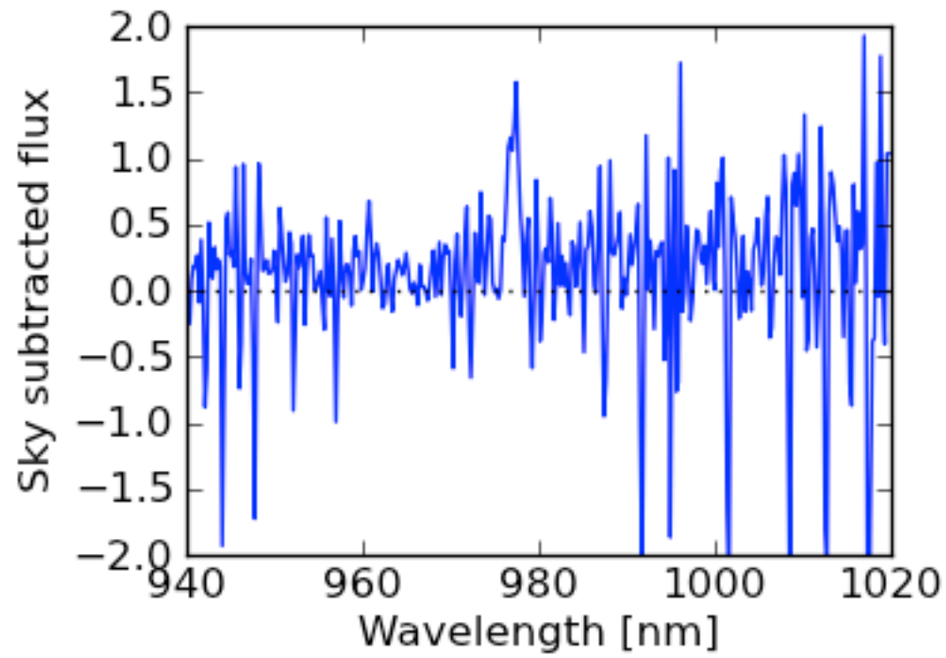
SP model enables simultaneous:
- Extraction
- Sky subtraction
- Coaddition

All while maintaining pure diagonal errors
- It “just” involves big matrices

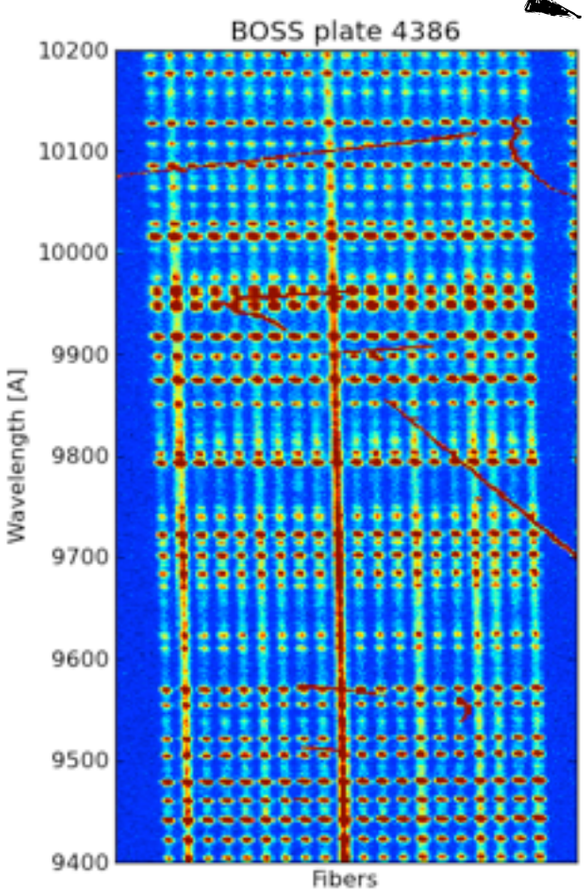

Design
Simulate
Analyze
Extract

Simulation = (Extraction)$^{-1}$

Both require
- 2D PSF Model
- (i,λ) <–> (x,y)



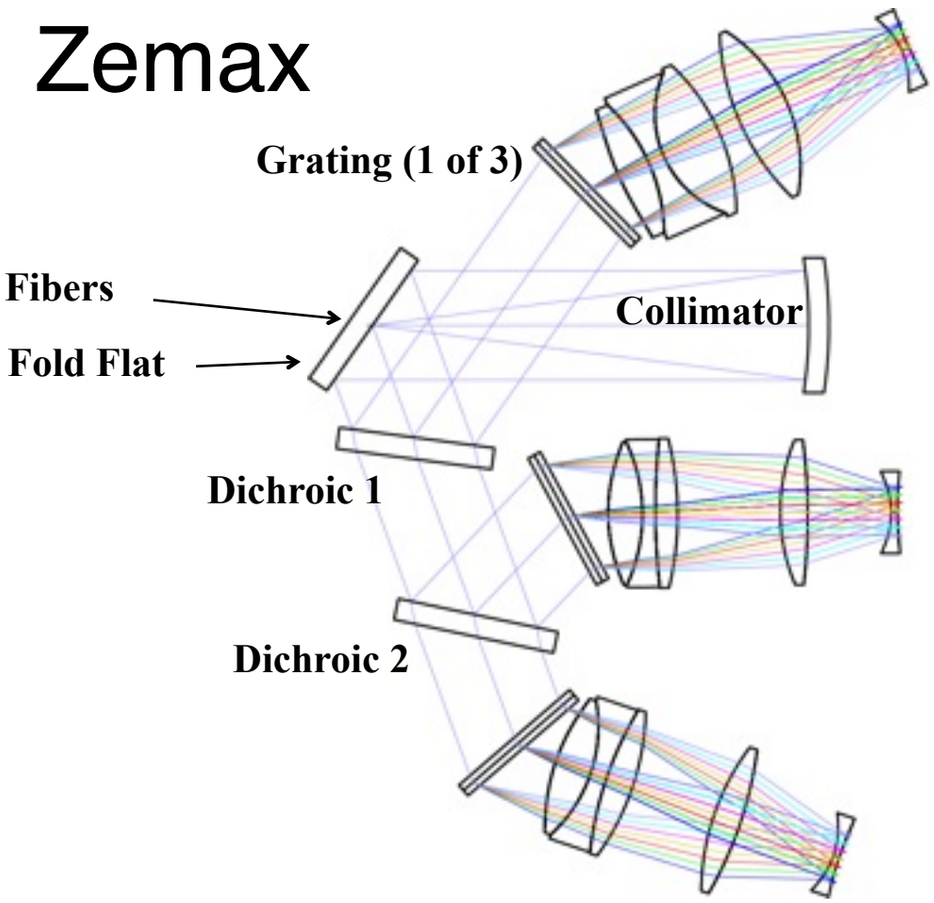
Simulate

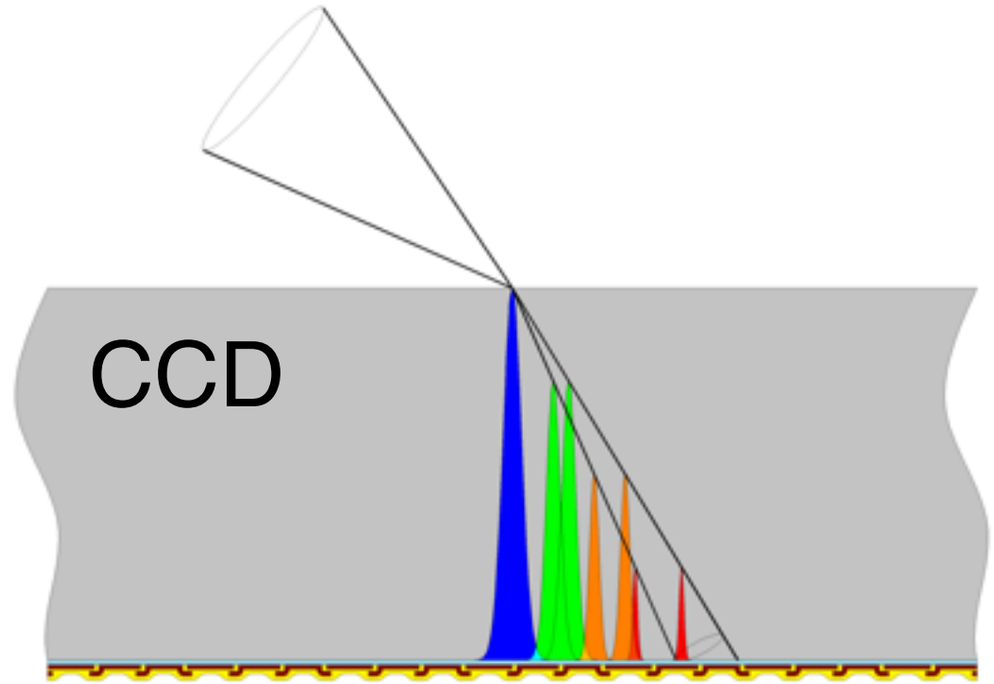


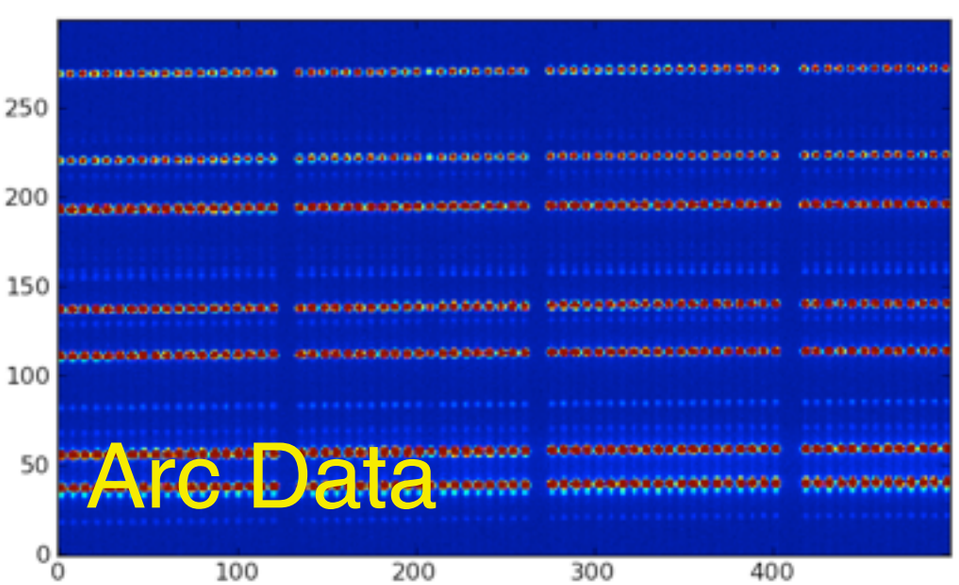
Extract

Spectrograph Simulations

Zemax





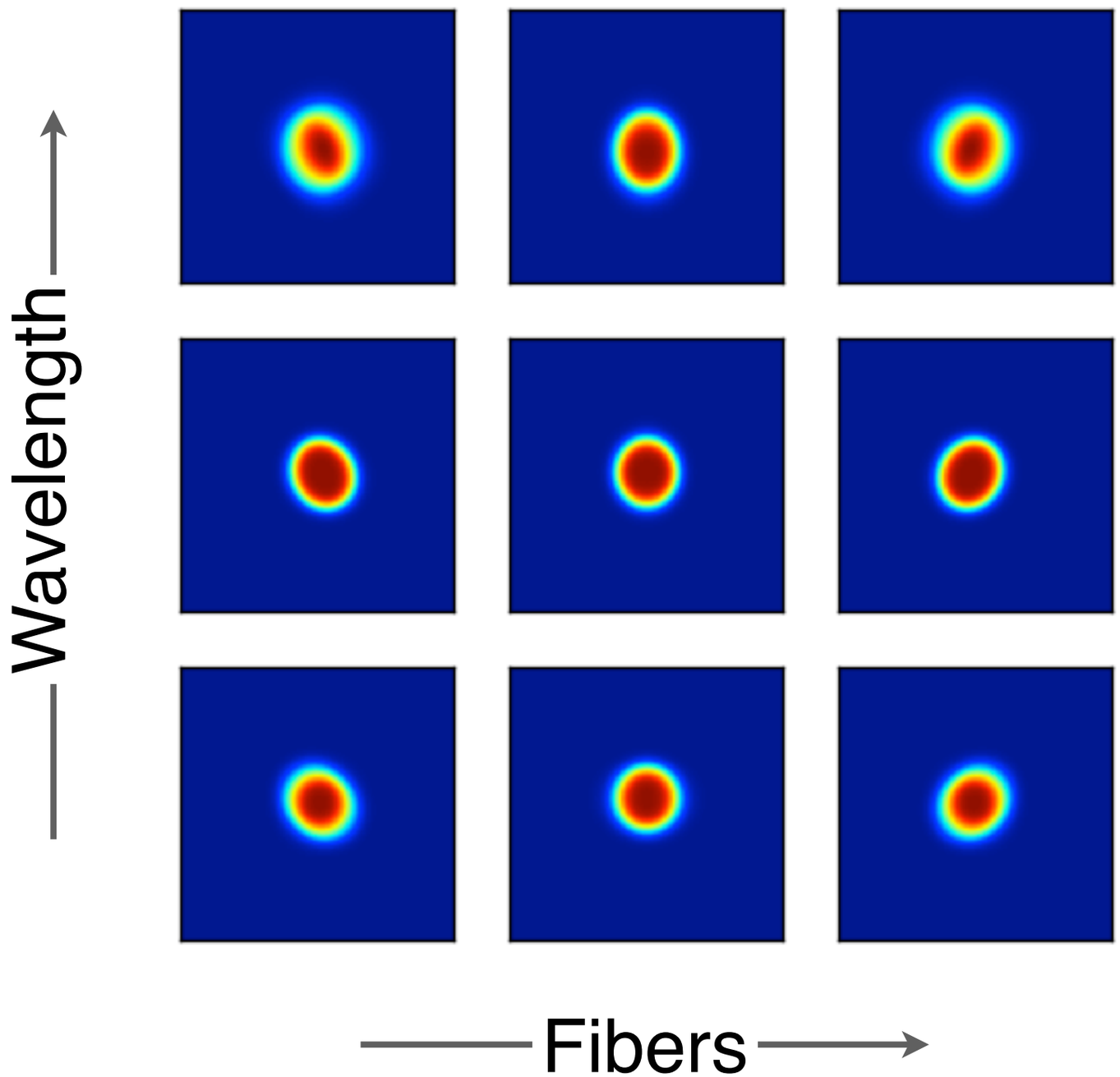


Pixels[i,j] =
PSF(fiber#, wavelength)

Sims by Pat Jelinsky

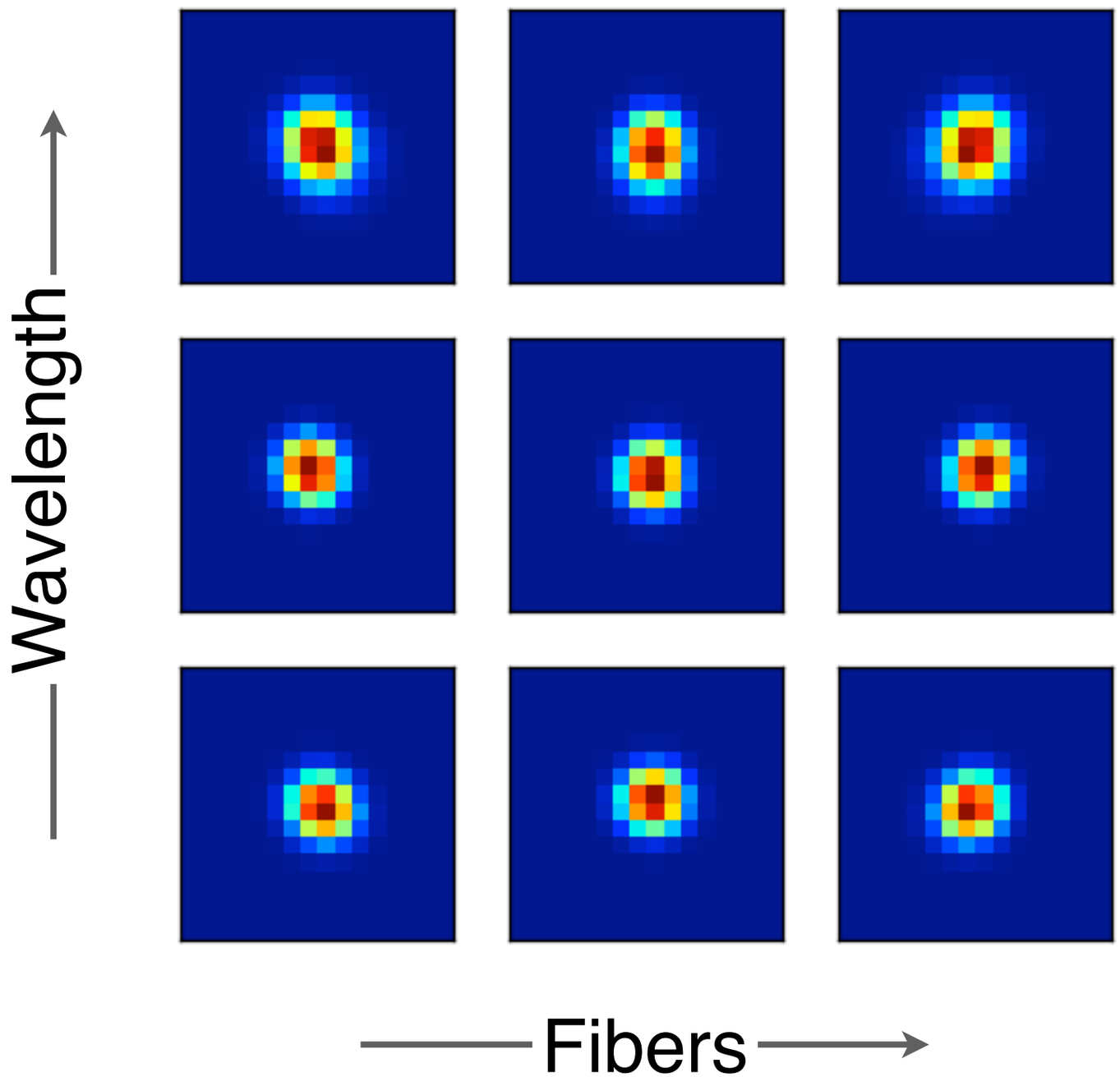
Instrument Model

2D PSF from Zemax ray traces + diffraction + CCD diffusion



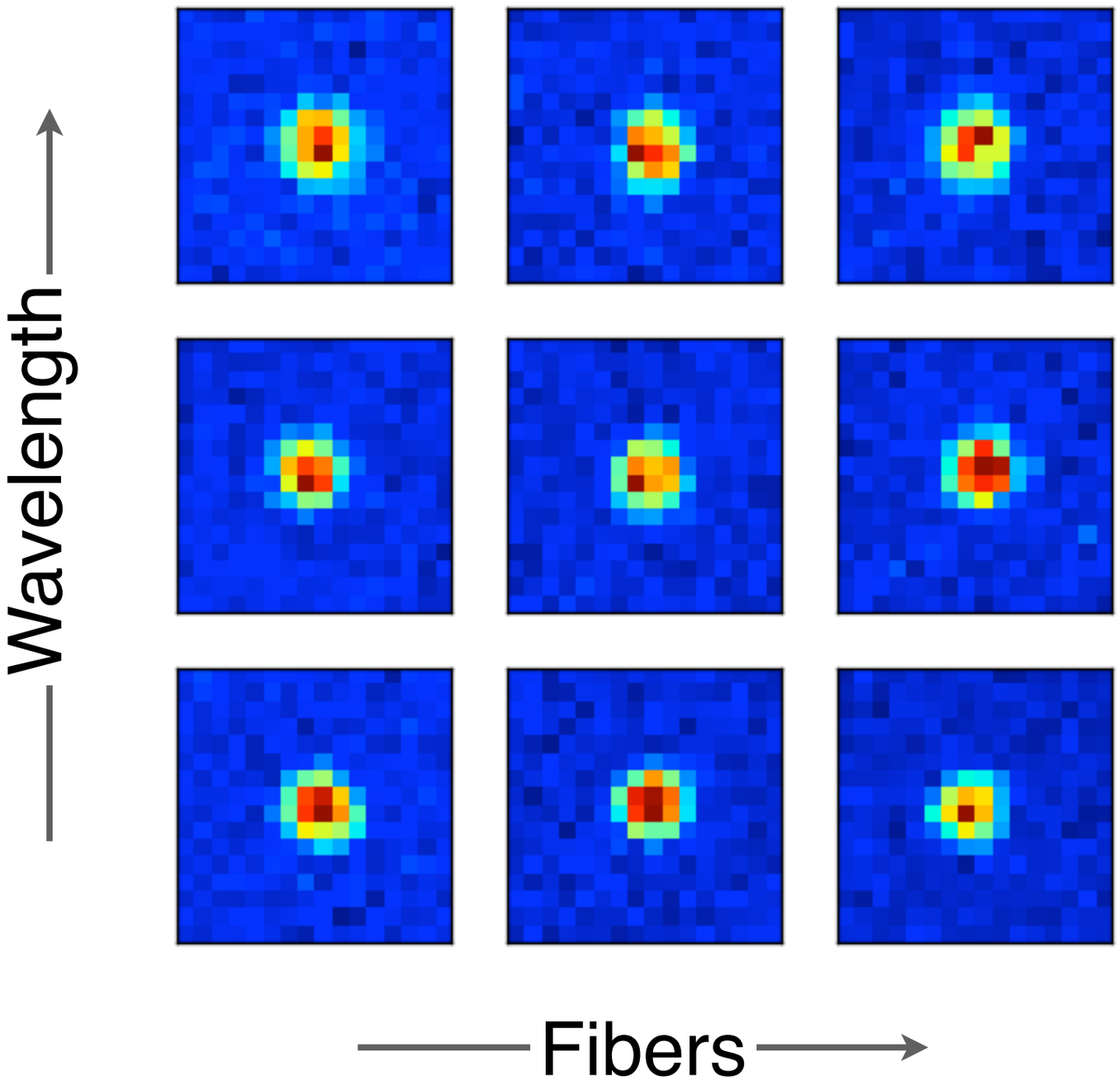
Instrument Model

Pixelize on CCD

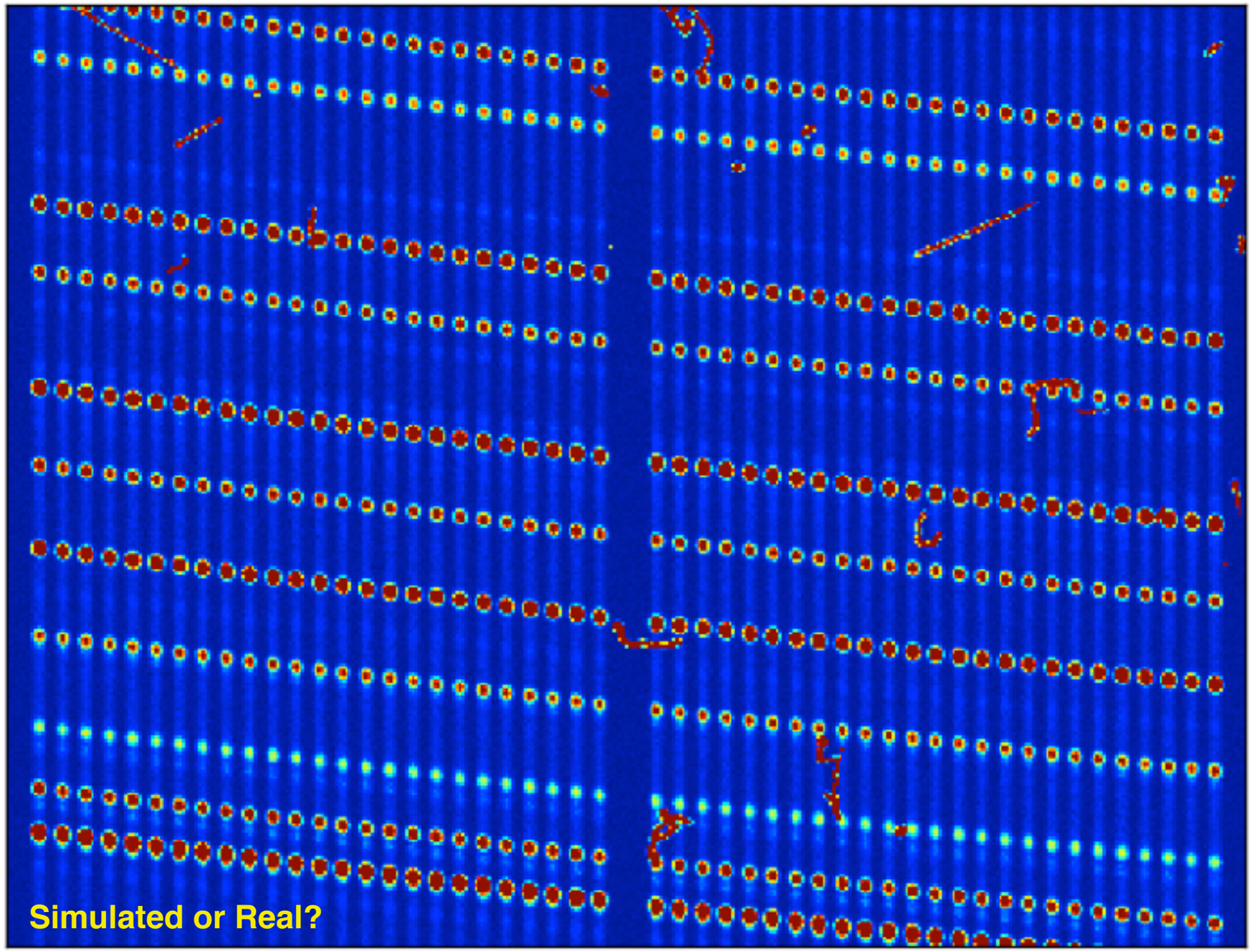


Instrument Model

Add noise



Many fibers, many λ



Two Spectro Sim Codes

BBspecsim (N. Mostek)

- 2D pixel-level simulation of single fiber
- Sky + ELG input spectra
- Applies throughput, 2D PSF
- Optimal extractions (i.e. row-by-row 1D)
- Sky subtraction using 25 simulated sky fibers
- Used for most BigBOSS design studies

Focused tool for specific design studies

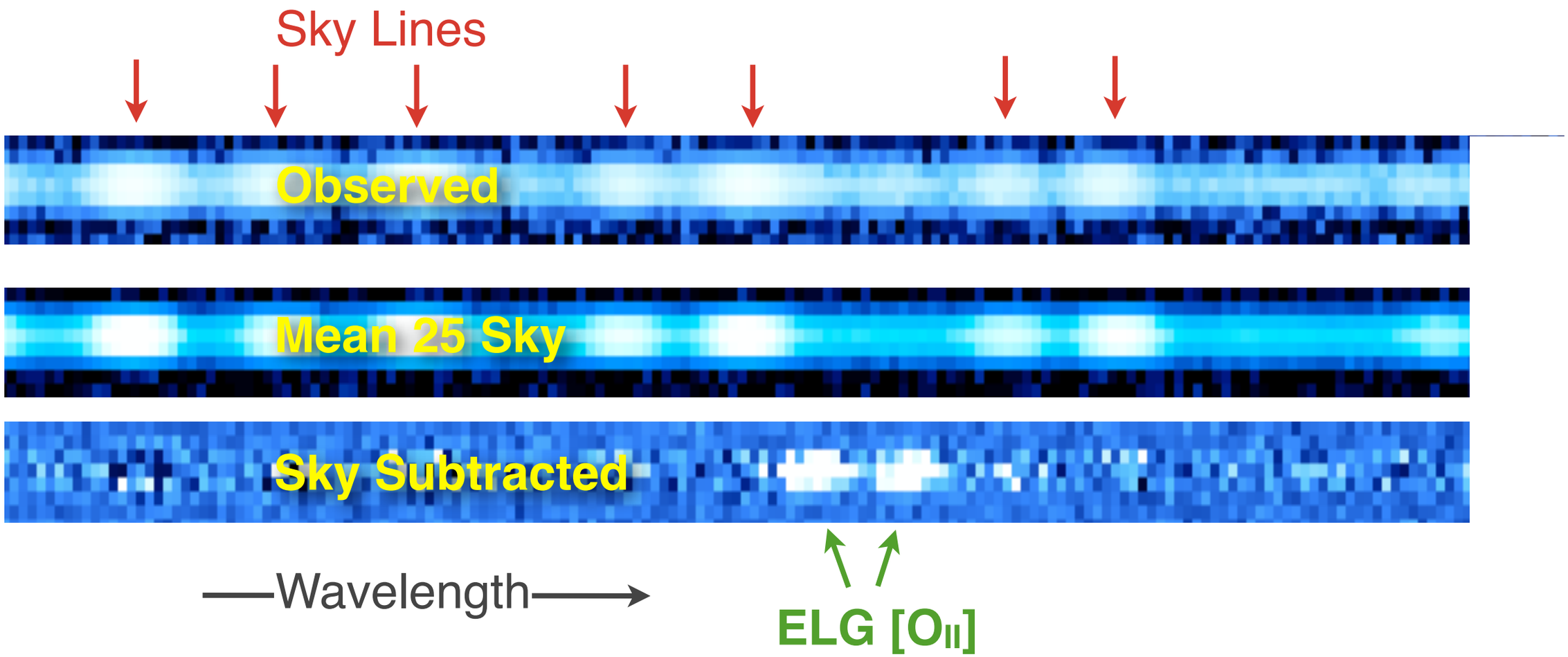
Specter (S. Bailey)

- Multi fiber full CCD 2D pixel-level simulator
- Test data for spectroperfectionism development
- Fiber:fiber cross talk issues
- PSF stability, predictability, calibration methods
  - Extraction code requirements on hardware design

Realistic “data” for final validation and prep for real data

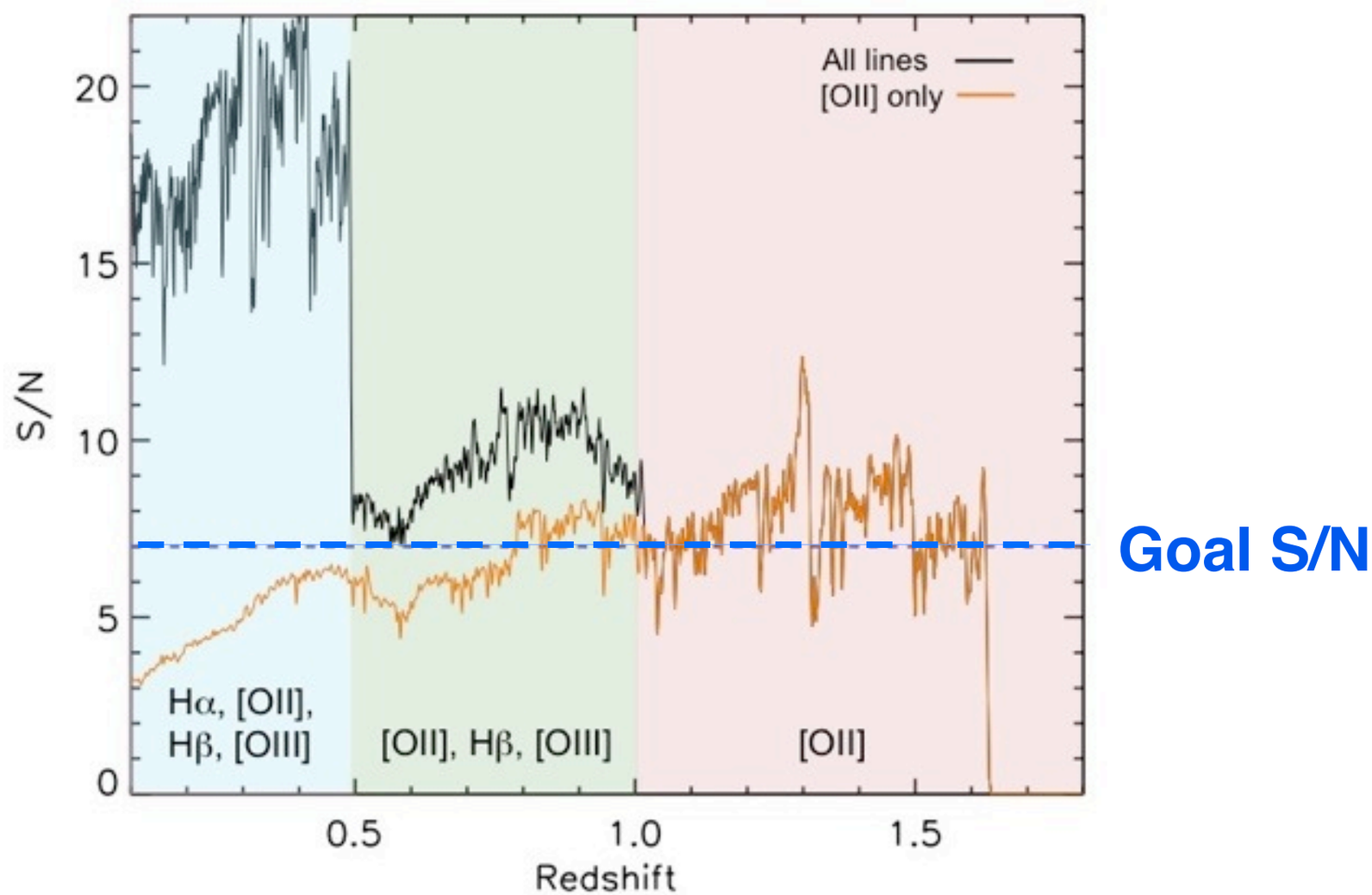
Nick Mostek

BBspecsim



Nick Mostek

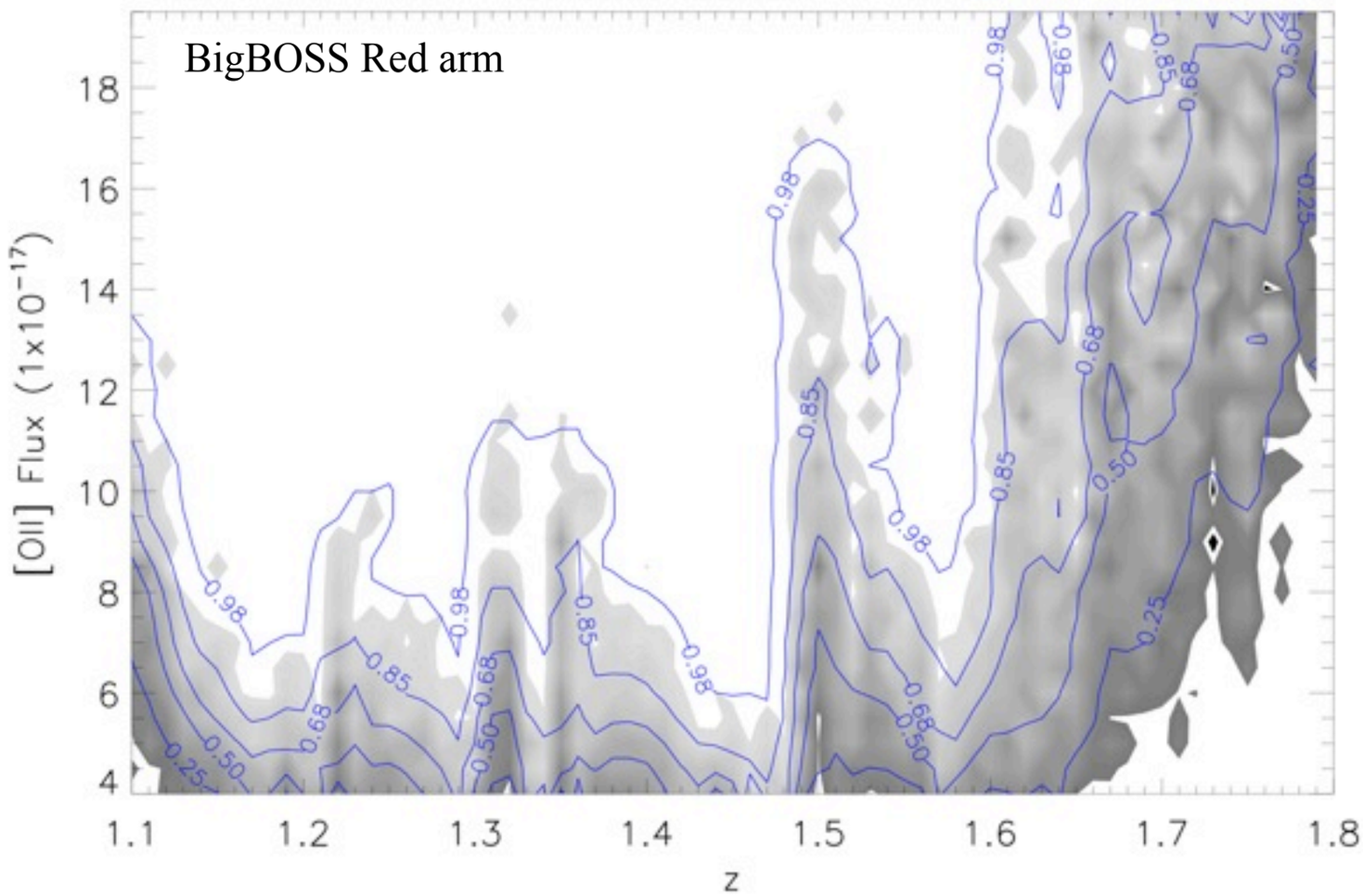
BBspecsim: S/N



- Simulated emission line S/N assuming constant flux ratios relative to F([OII])=8x10^{-17} ergs/s/cm^2
- Goal of S/N>7 reached in 20 minutes in median observing conditions
- Better S/N and redshift success when multiple lines are measured

Nick Mostek

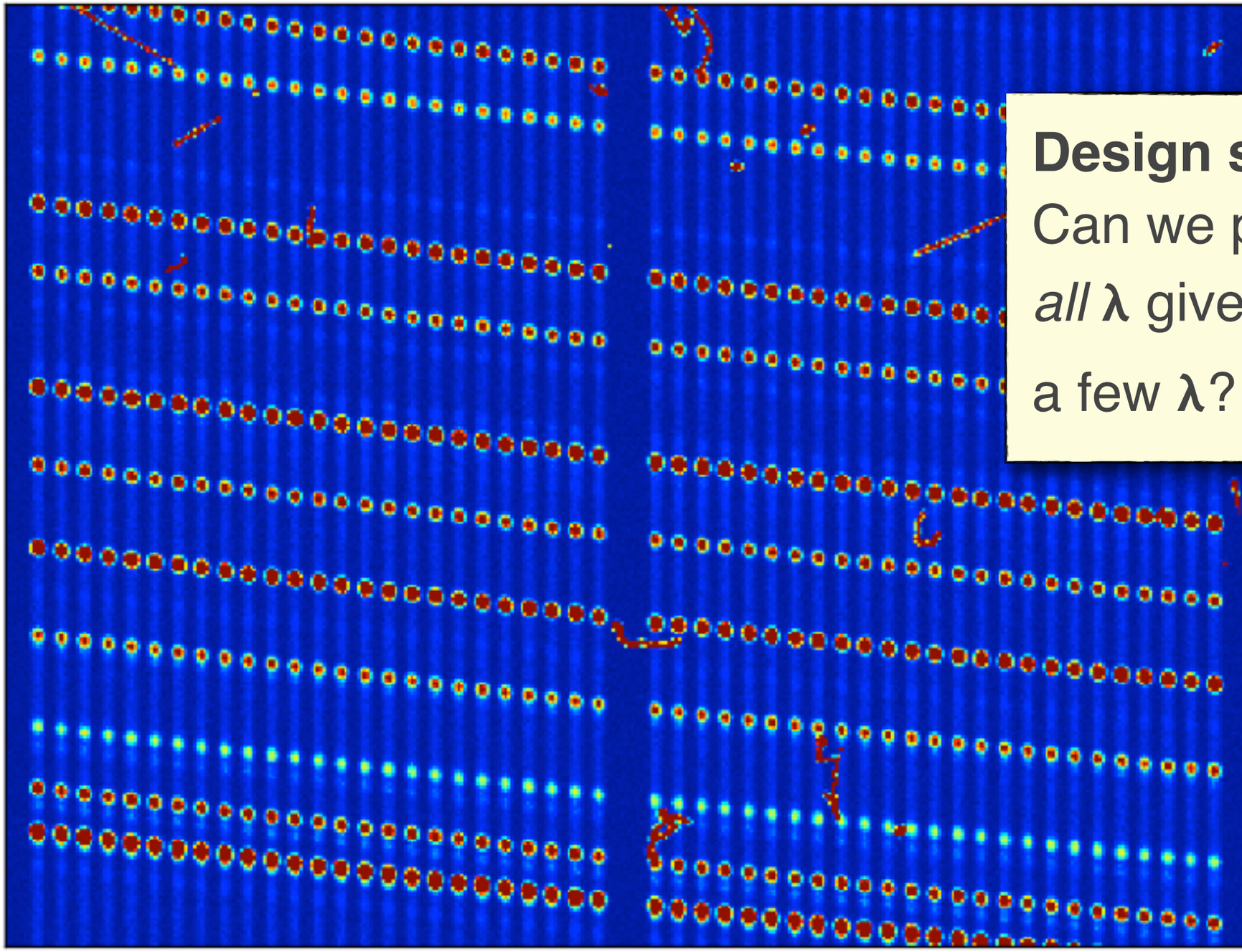
BBspecsim: Redshift Success



- Have performed a Monte Carlo simulation with random ELG redshifts and line fluxes
- Data fit with emission line templates (*zfind* SDSS legacy code)
- Redshift success rates depend on S/N and redshift, averaged in 0.1Δz bins

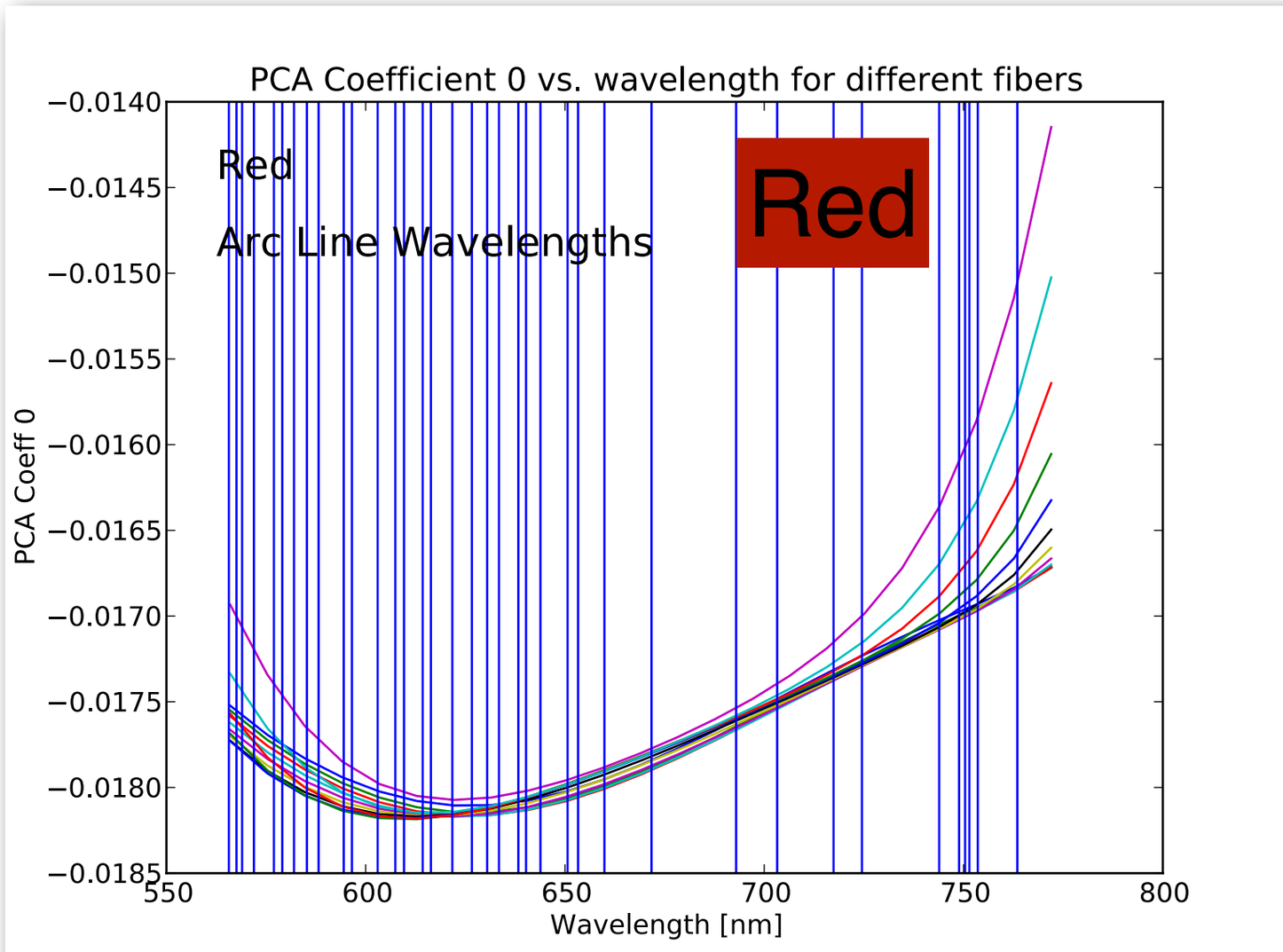
[OII] emission lines at S/N>7 are well detected
AND have a high redshift success rate (~95%)

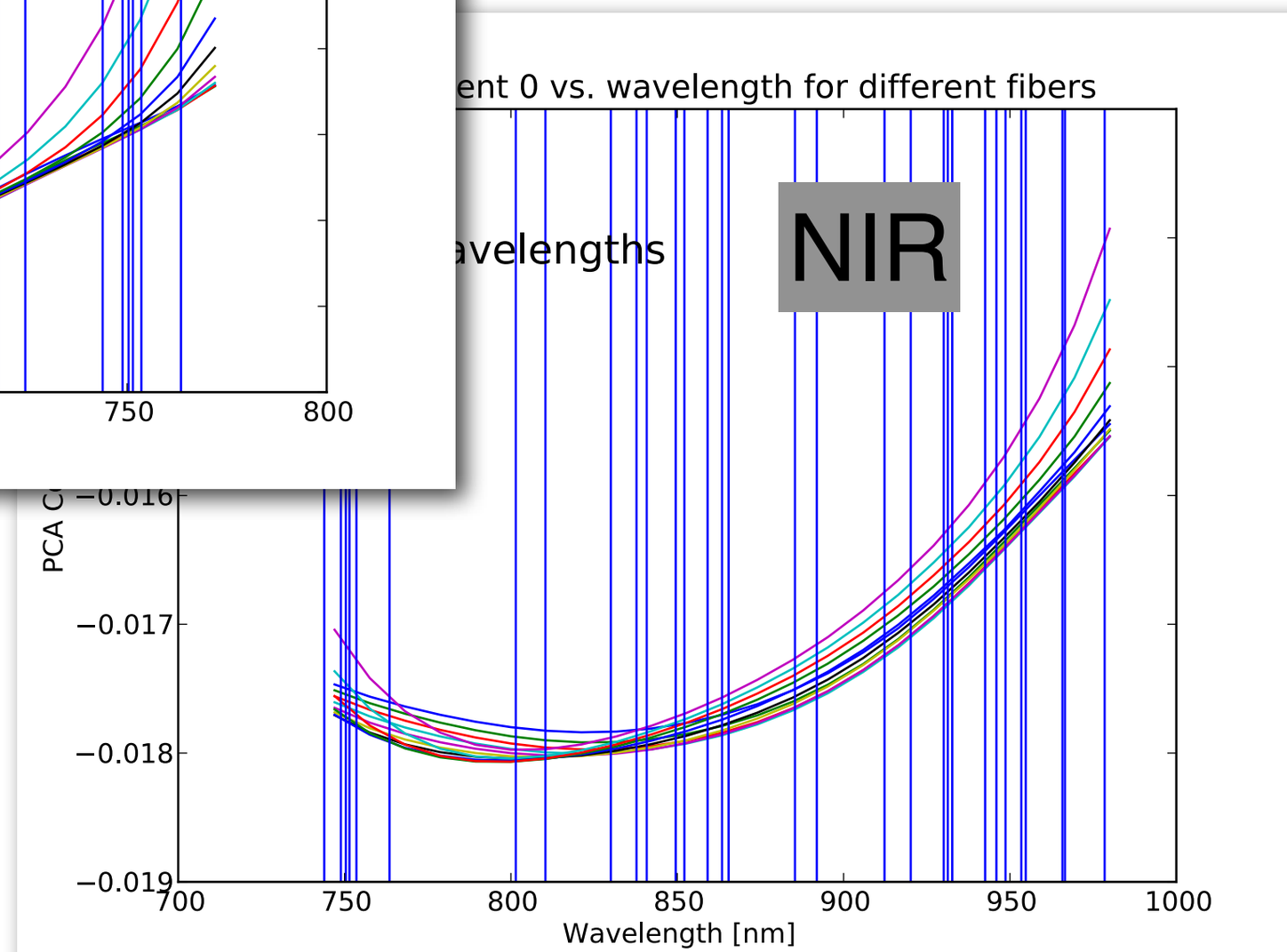
Specter: Multi-fiber Simulator



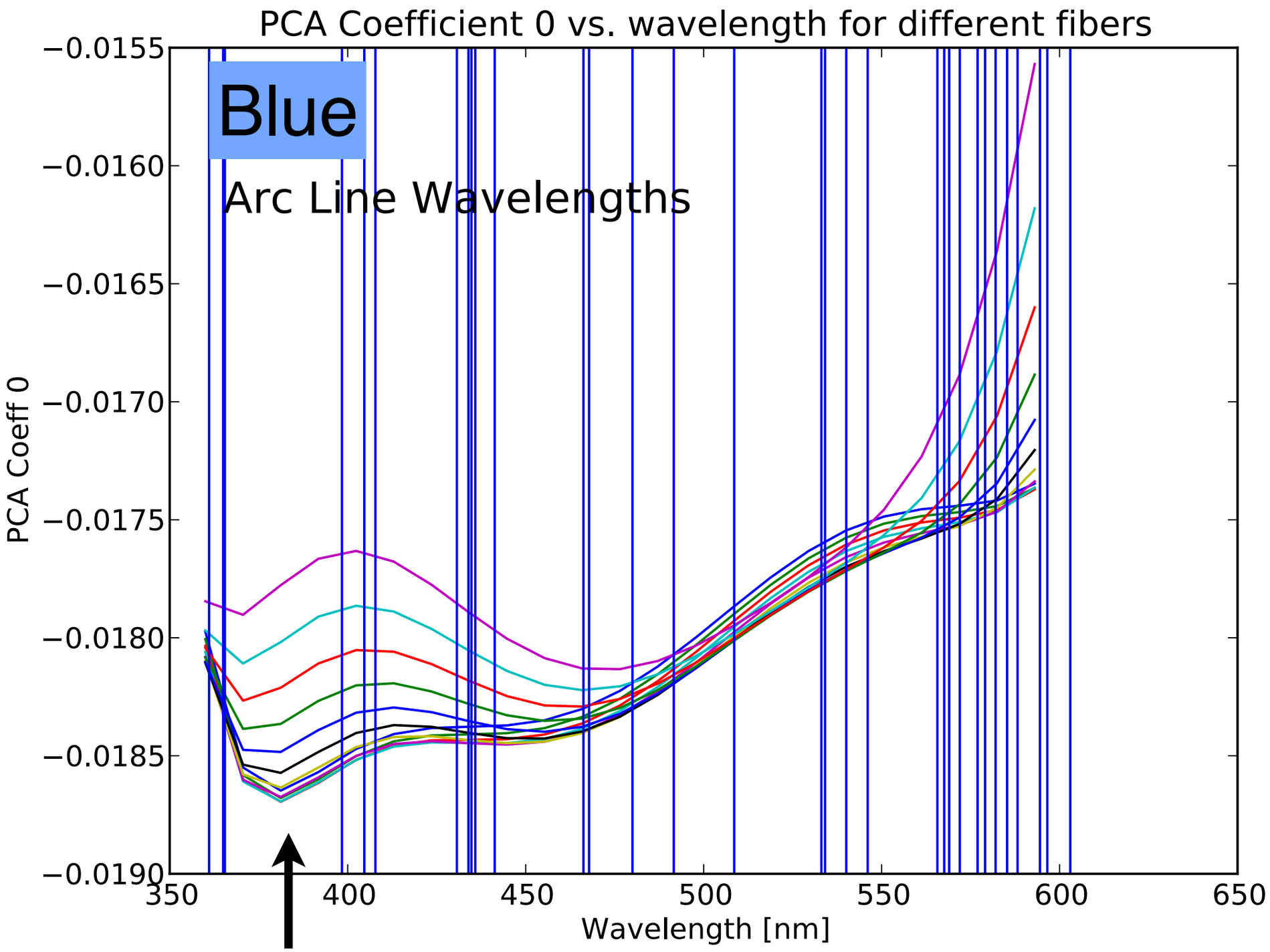
Design study:
Can we predict 2D PSF at *all* λ given arc lamp lines at a few λ?

Specter: PSF Predictability



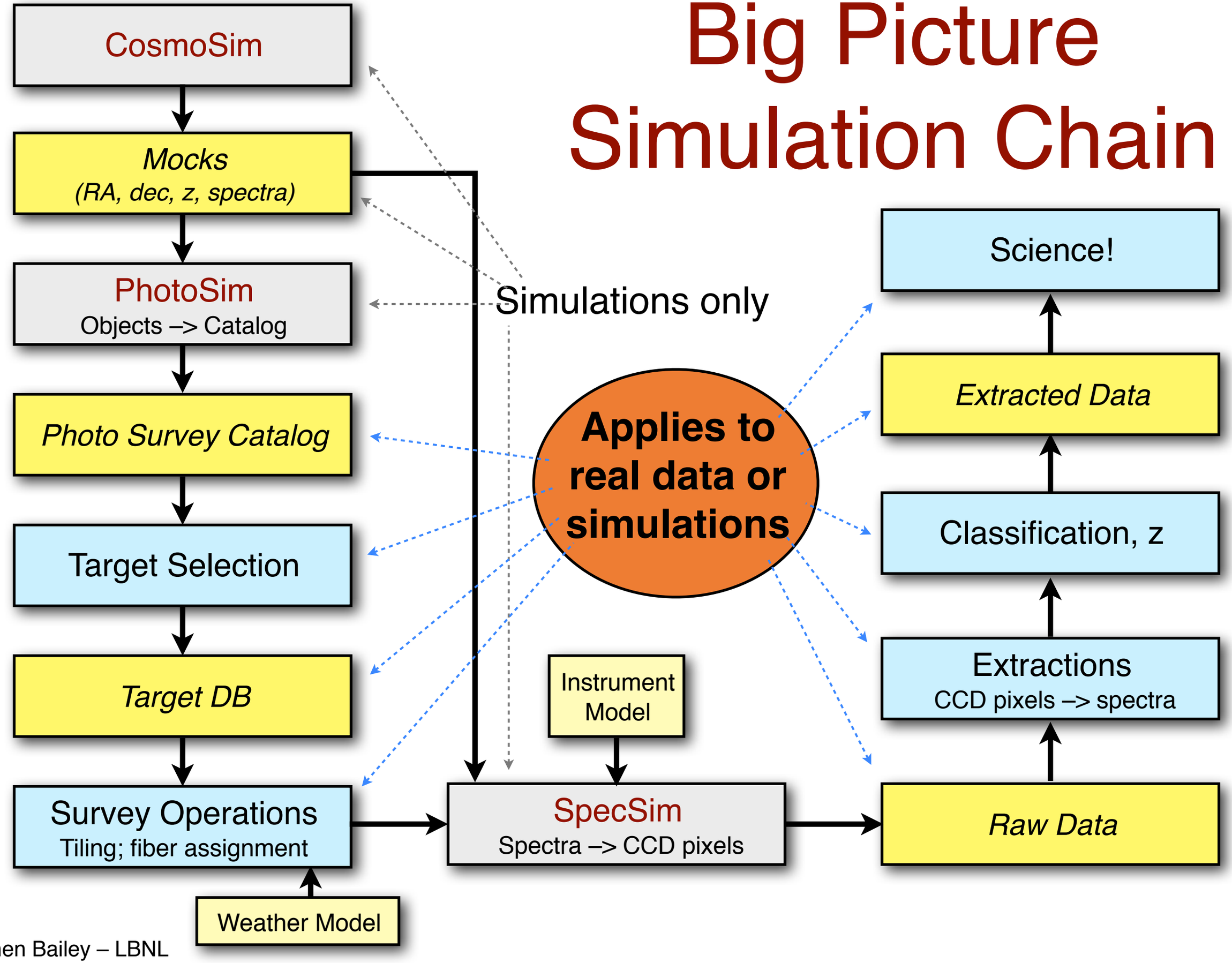


Specter: PSF Predictability



Need extra calibration measurements in the blue

Big Picture Simulation Chain



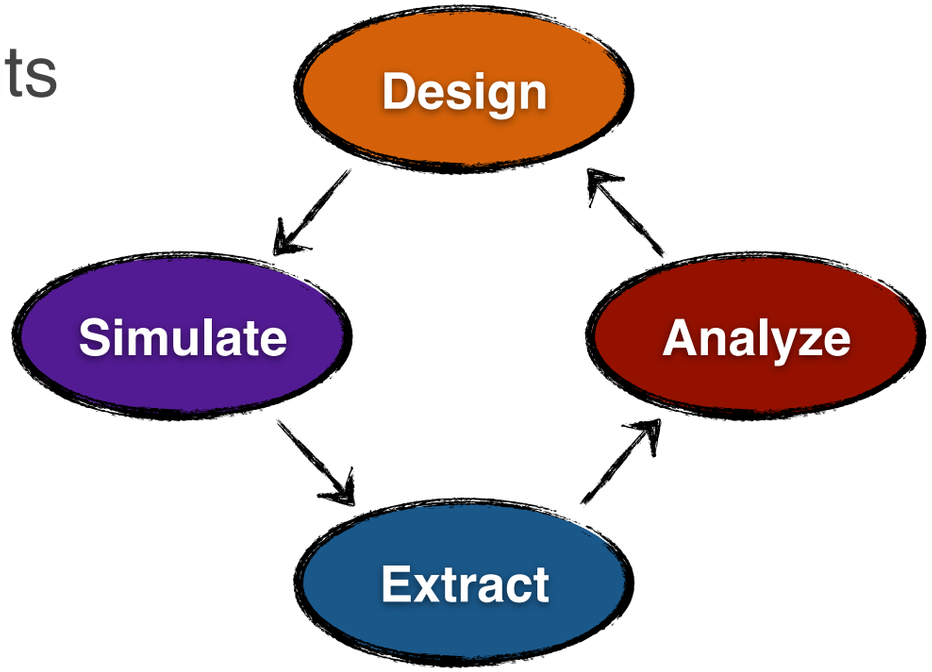
Summary

Spectral Extractions

- Optimizing algorithms to get the most out of the data
  - e.g. full error propagation with diagonalized covariance
- Testing on BOSS data keeps us honest

Spectral Simulations

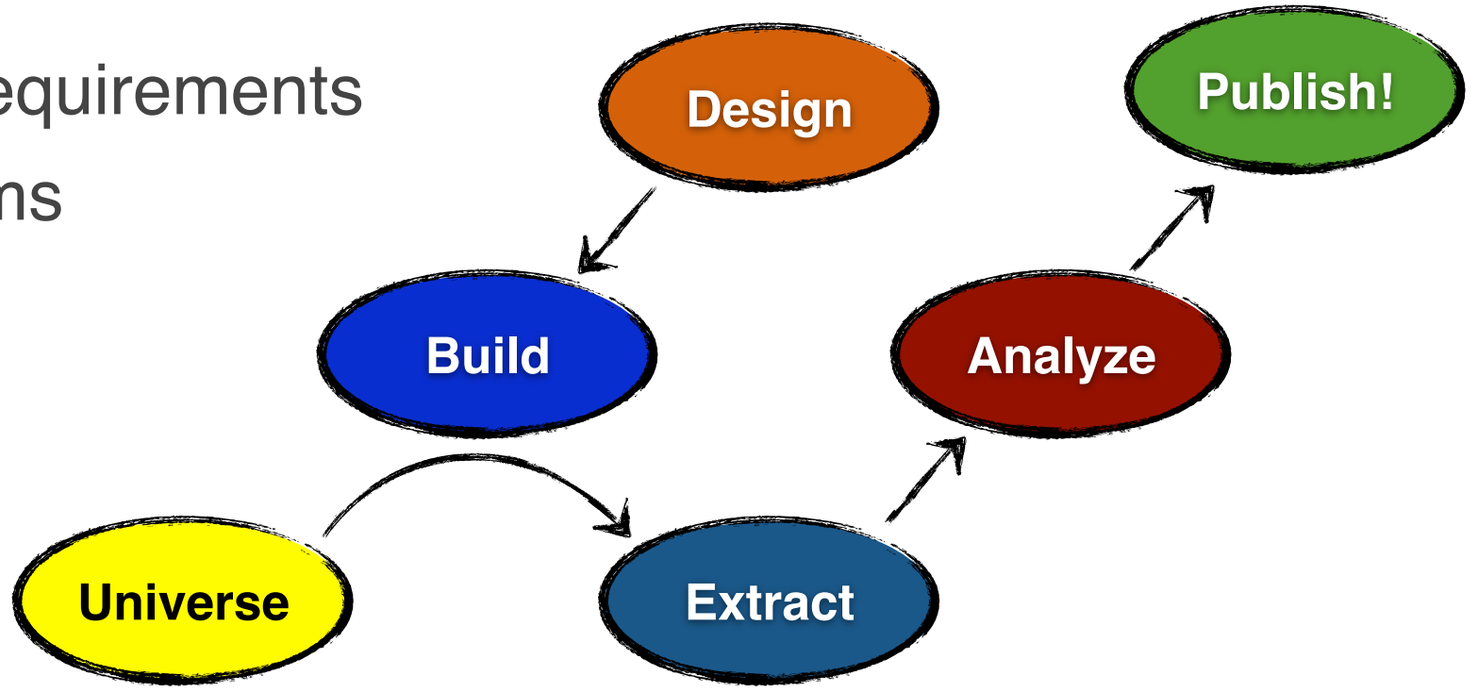
- CCD pixel-level simulations
- Validate designs, develop requirements
- Develop extraction algorithms
- Prepare for real data



Summary

Spectral Extractions

- Optimizing algorithms to get the most out of the data
  - e.g. full error propagation with diagonalized covariance
- Testing on BOSS data keeps us honest

Spectral Simulations

- CCD pixel-level simulations
- Validate designs, develop requirements
- Develop extraction algorithms
- Prepare for real data