Towards 3D Spectra of Galaxies

Ching-Wa Yip
Johns Hopkins University

With Thanks To

- Alex Szalay
- Rosemary Wyse
- László Dobos
- Tamás Budavári
- Istvan Csabai
- Michael Mahoney (Math @ Stanford)
- Emmanuel Candes (Math @ Stanford)
- Vivian Wild
- Samuel Carliles (CS @ JHU)
- Brice Ménard
- Guangtun Zhu
- Tim Heckman
- Robert Kennicutt
- Daniela Calzetti
- Andrew Connolly

Institute for Data Intensive Engineering and Science

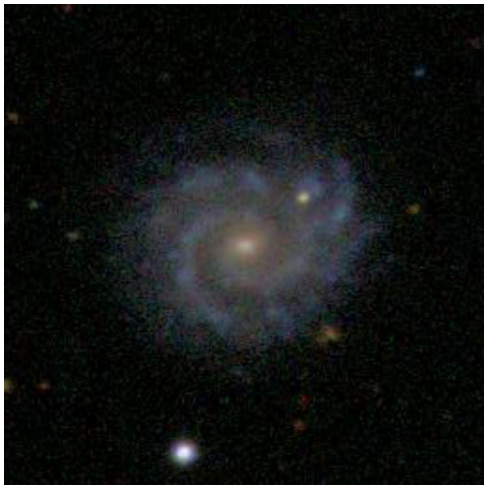
SDSS

NATIONAL VIRTUAL OBSERVATORY

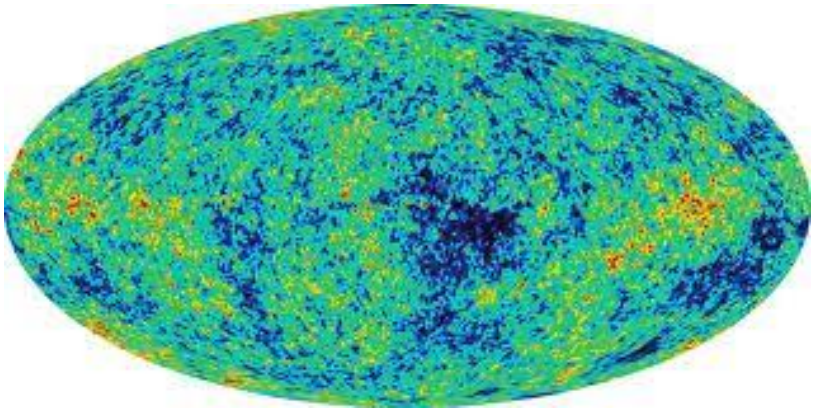
Critical Questions

- How much normal matter are present in galaxies?
- What do they tell us about galaxy formation?

Dark Energy 72%
Dark Matter 23%
Normal Matter 5%

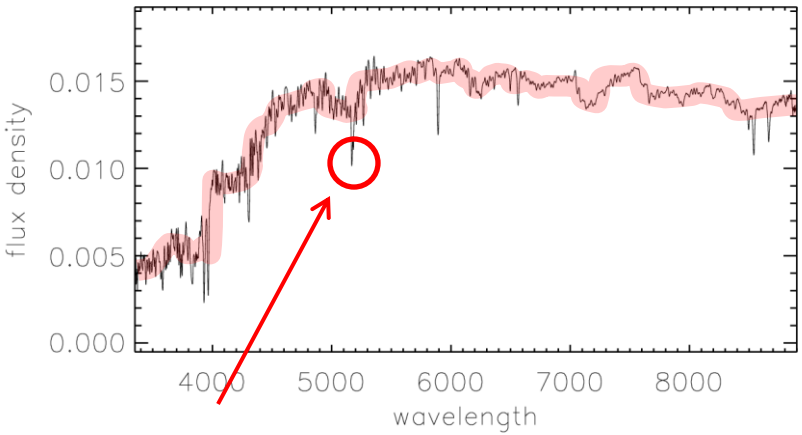


Today



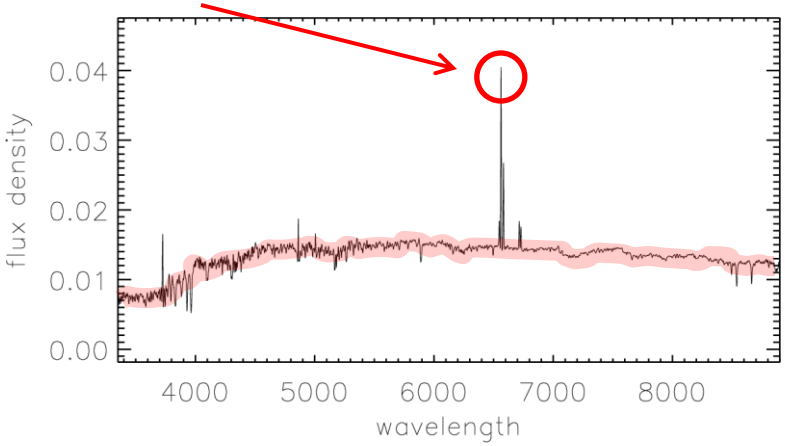
Early Universe

Spectra are the keys to Stars, Gas and Dust in Galaxies



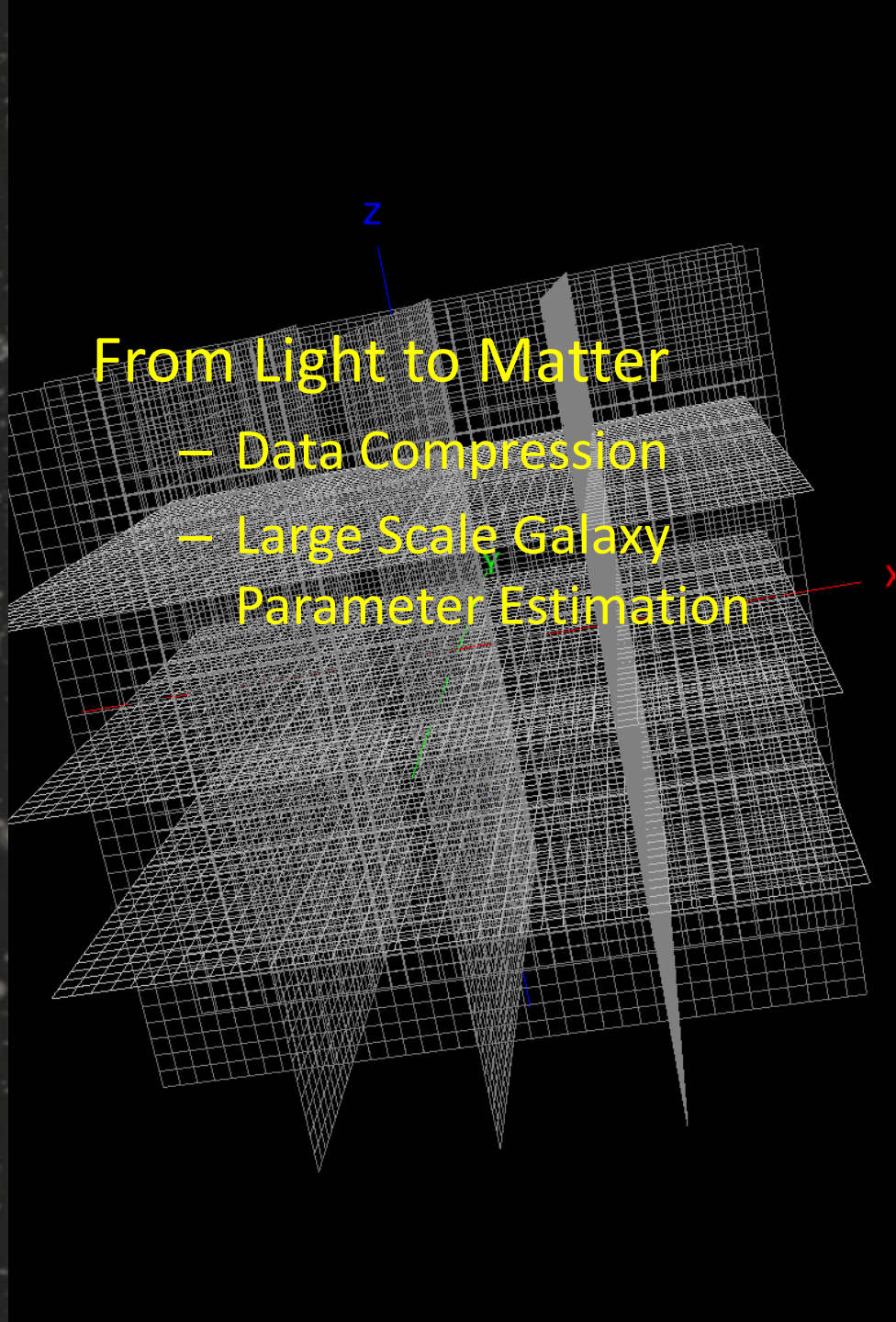
Absorption and Emission Lines

Continuum Emissions



Measure Dust & Light in Galaxies

- (1D) Inclination-Dependent Composites
- 2D Composites
- 3D Composites
  - 3D and High Definition



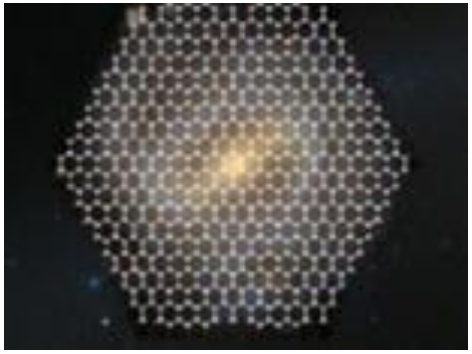
From Light to Matter

- Data Compression
- Large Scale Galaxy Parameter Estimation

SDSS III

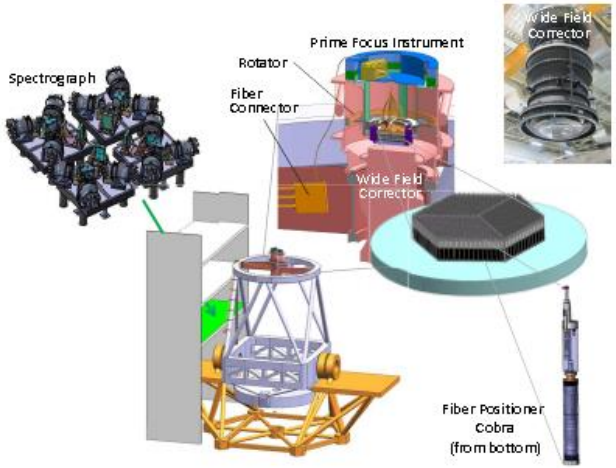
Current & Future Spectroscopic Surveys

- Multiplexing: many objects at a time (SDSS, LAMOST, MaNGA, PFS)
- 2D In Situ: as a function of projected 2D position on galaxies (Integral Field Units, CALIFA, SAMI, MaNGA)



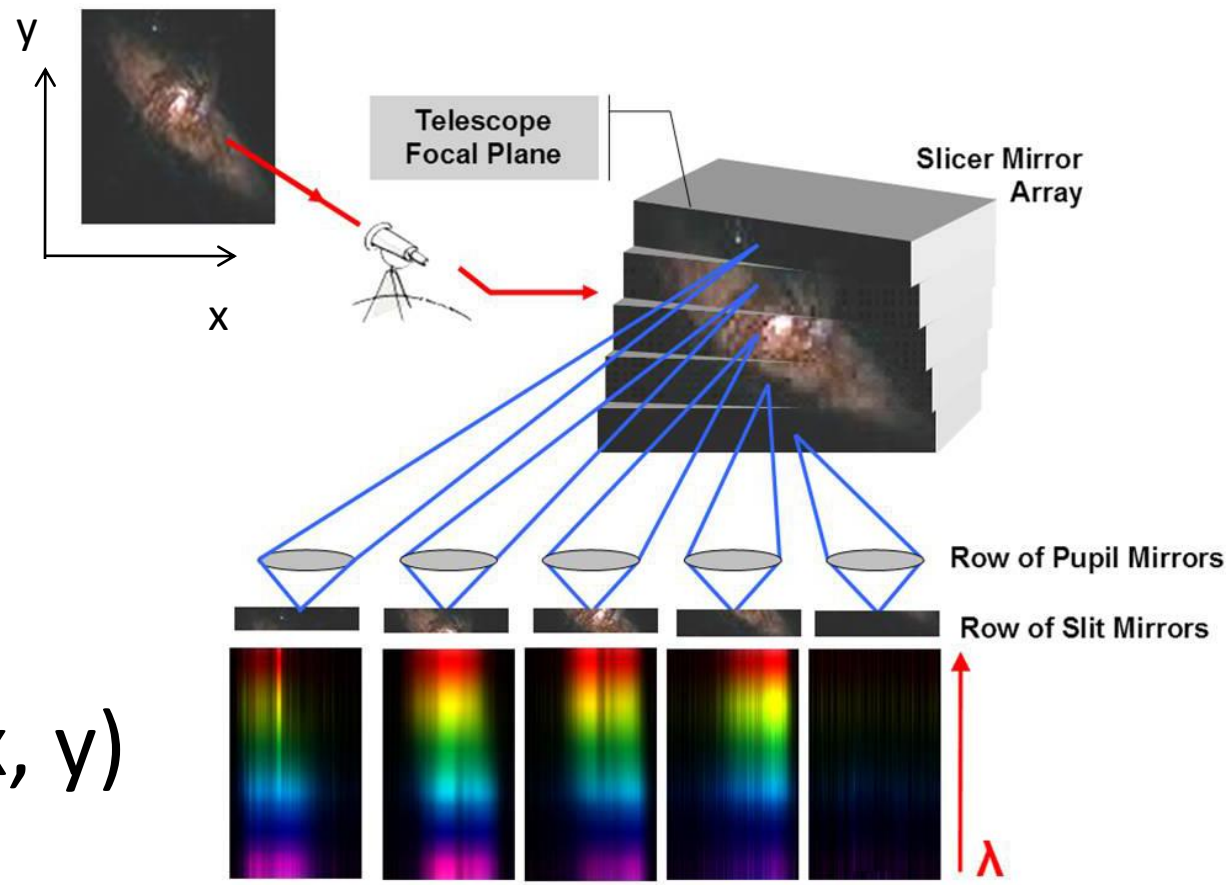
CALIFA

LAMOST



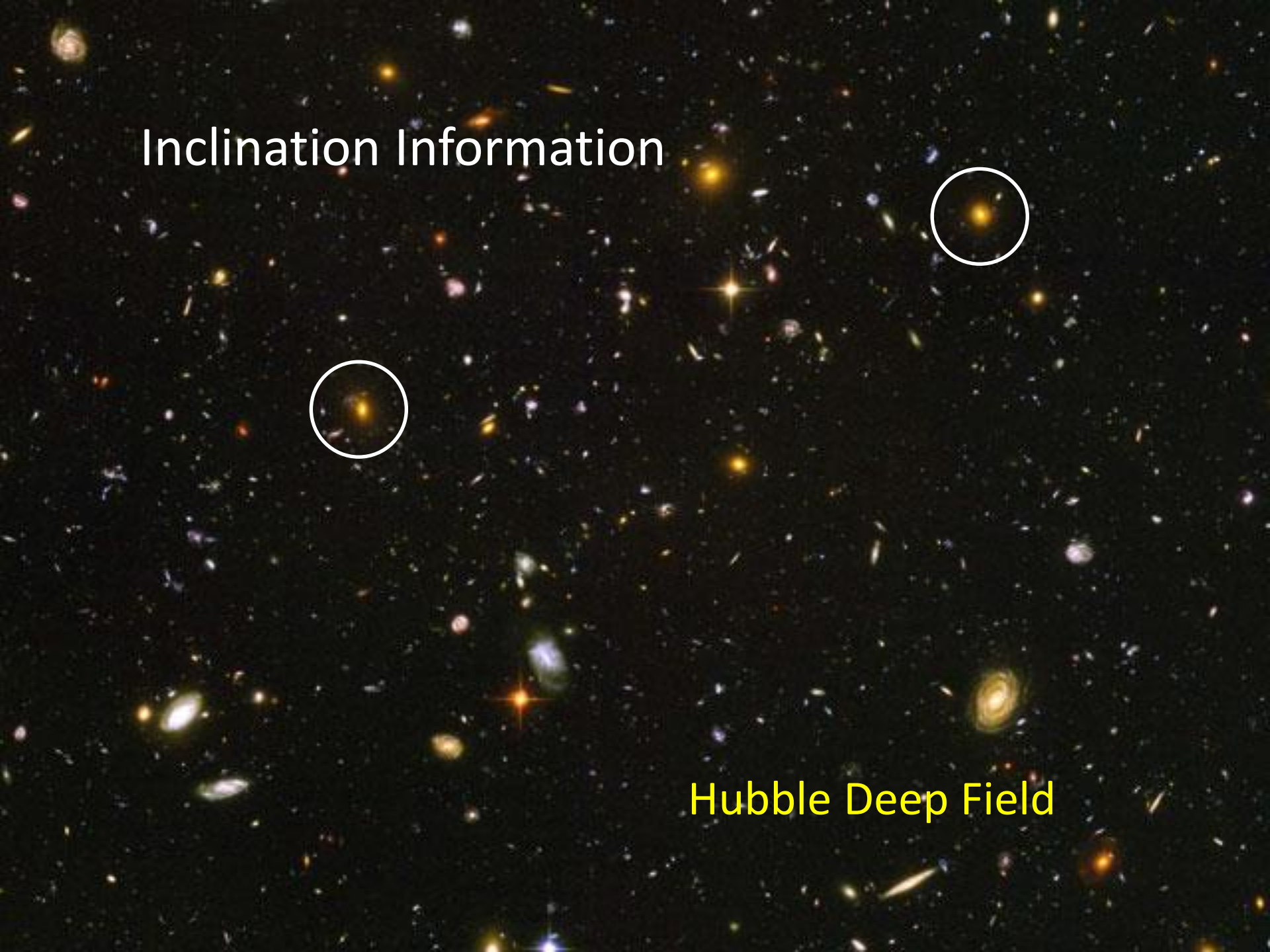
Subaru PFS

Integral Field Spectroscopy (or “3D” Spectroscopy)



$f_\lambda(x, y)$

Ealet et al. 2008

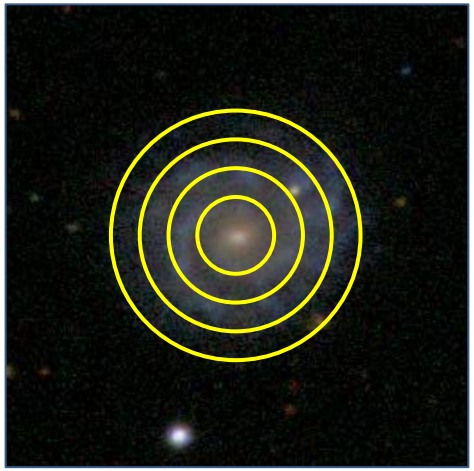

Inclination Information
Hubble Deep Field

Galaxies are located at various inclinations & distances

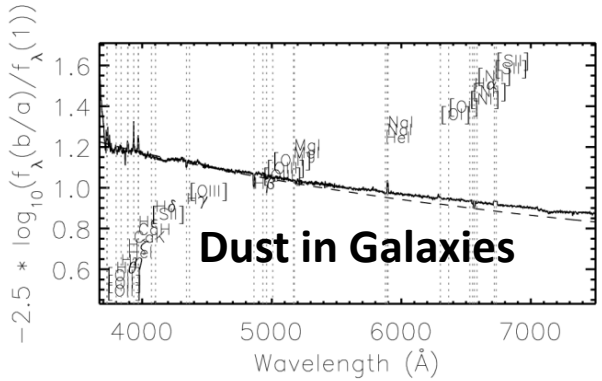
From SDSS JPEG Image Cutout tool.

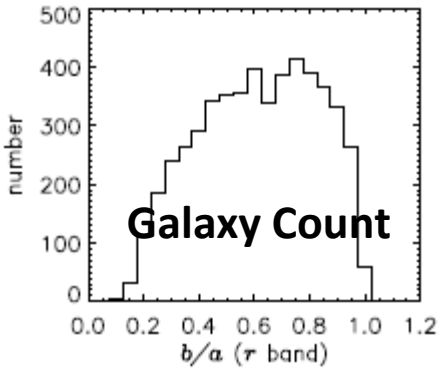
Why 3D Spectrum

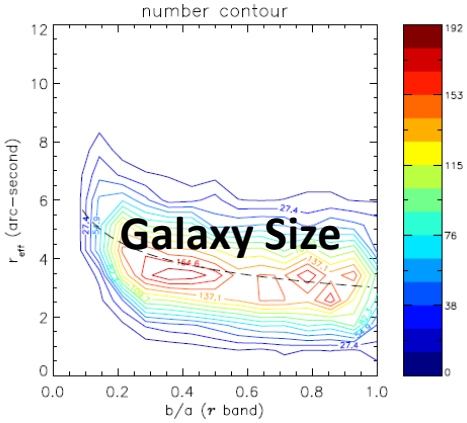
- Galaxies are 3D.
- Impact across many topics:
  - Radial Variance: Galaxy Evolution
  - Inclination Variance: Dust , Galaxy Properties, Cosmology

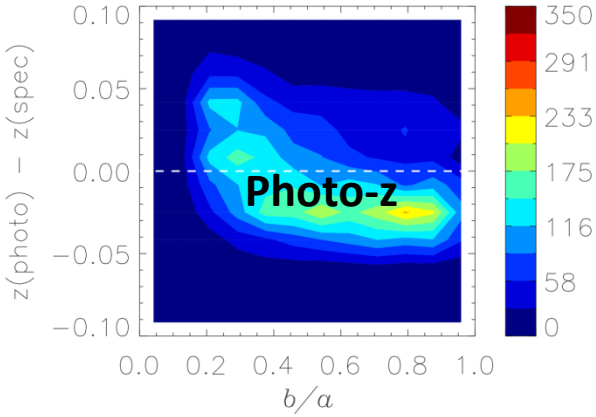


Galaxy Evolution at Sub-galactic Scales







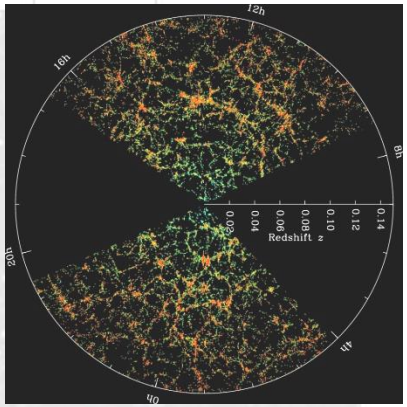


Reconstruction of Galaxy Spectral Templates

- 1D emissivity: f(galaxy center; b/a), inclination variation at galaxy center (Yip et al. 2010)
- 2D emissivity: f(x, y; b/a) by Drizzle2D algorithm (Yip et al. 2014 in prep)
- 3D emissivity: f(x, y, z) by Drizzle3D algorithm (future)

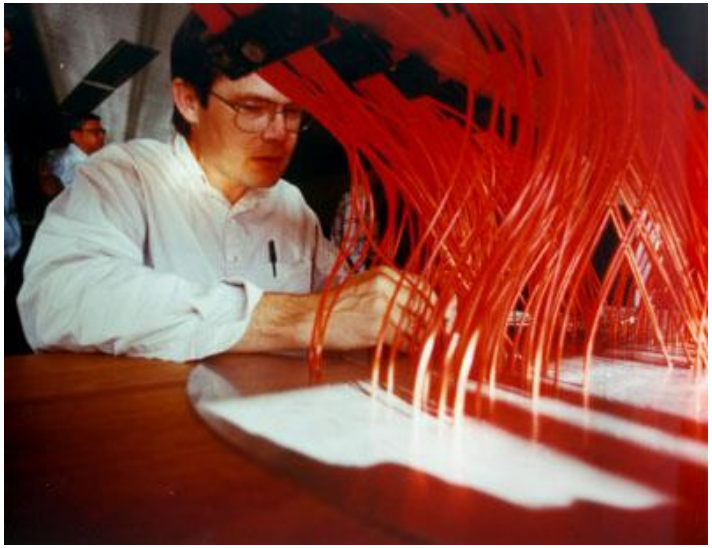
Sloan Digital Sky Survey

- Photometric + Spectroscopic Surveys
  - 11,000 square degree footprint (DR7)
  - 5.9×10^8 u, g, r, i, z photometry
  - 1.6×10^6 fiber spectra
- Phases
  - SDSS I (2000-05)
  - SDSS II (2005-08)
  - SDSS III (2008-14)
  - SDSS 4 (2014-20)
- Data are public
- Web interfaces for data download & exploration
  - SkyServer, DAS, etc.

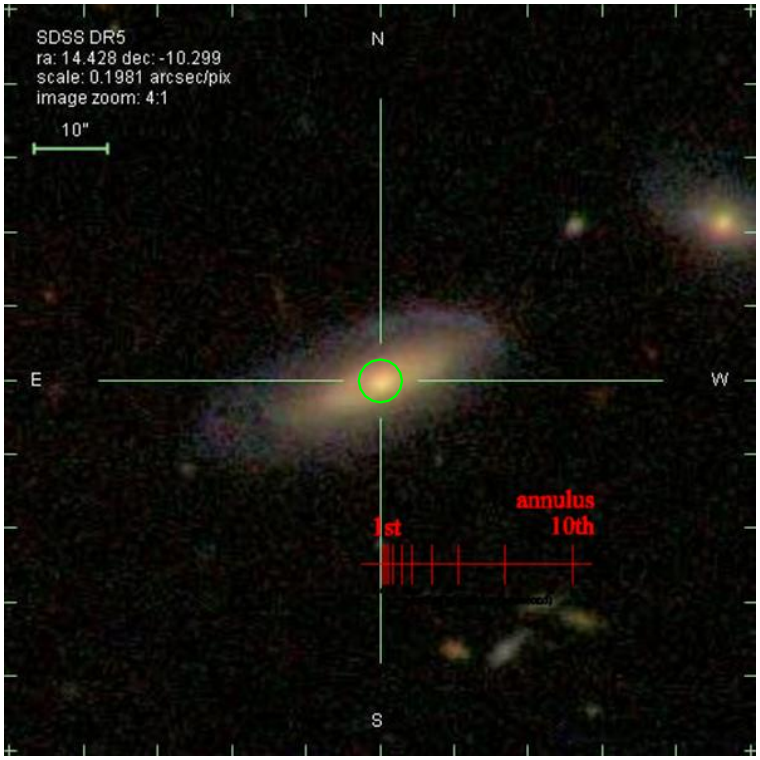


Galaxy light sampled by the SDSS spectroscopy

- SDSS offers fiber spectra at 3800-9200 Å, 69 km/s resolution.
- A fiber projects 3 arcsec-diameter area on the sky.



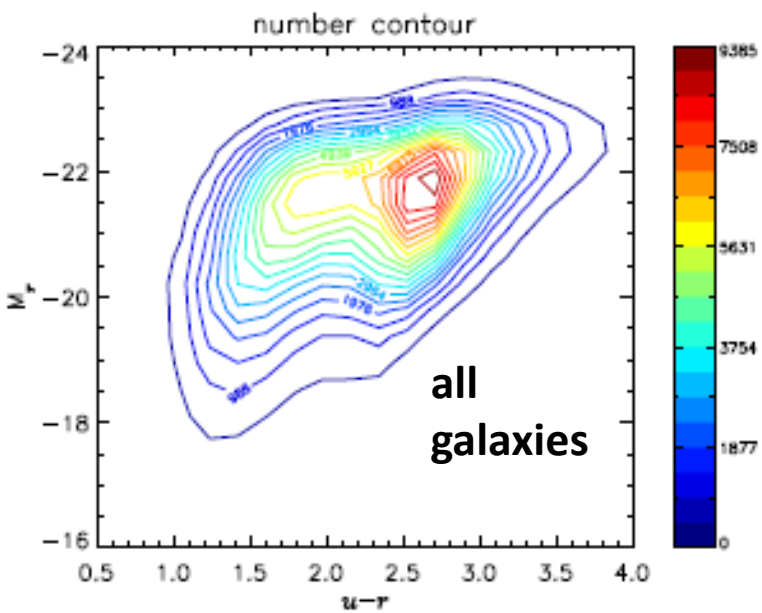
SDSS fiber plug plate.

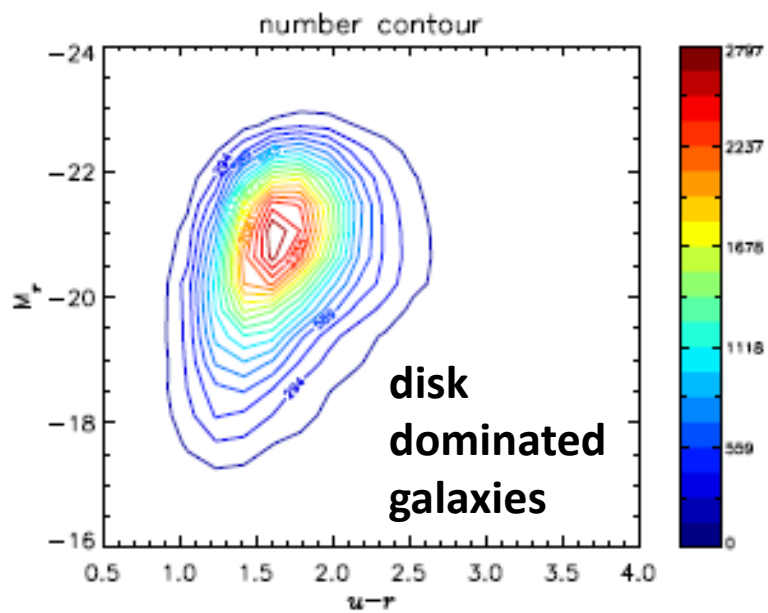


The central 3 arcsec of a galaxy.

Sample: Disk Galaxies in SDSS DR6

- Disk dominated (bulge fraction < 0.1, u – r < 2.4)
- Star forming
- Flux Limited r < 17.7
- Redshift = 0 – 0.2



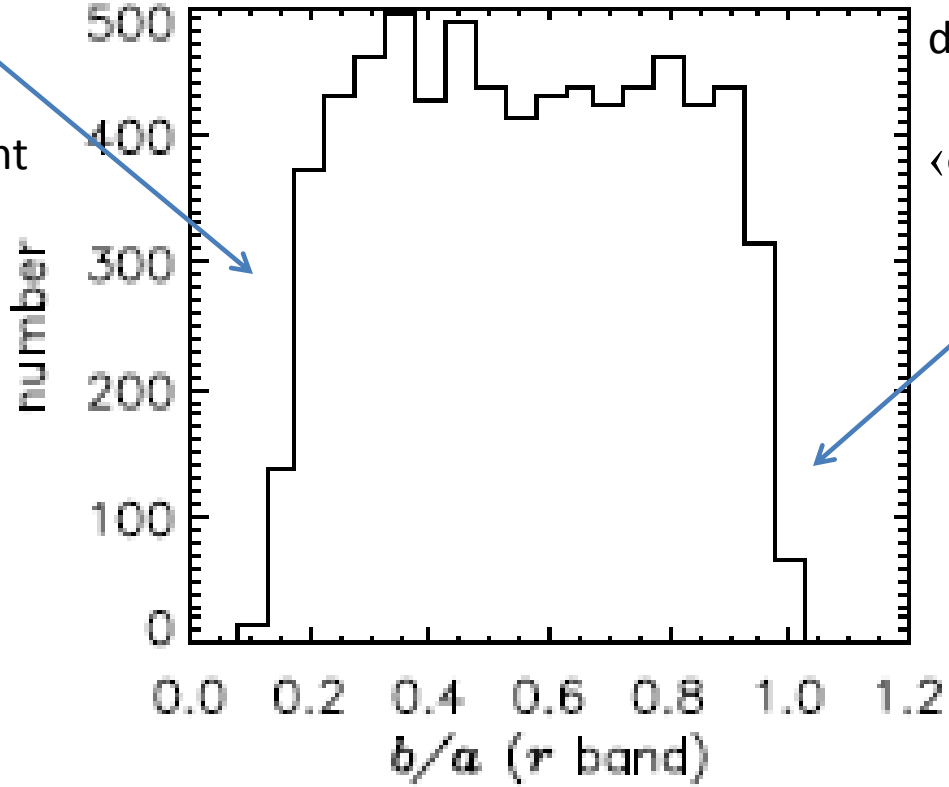


Dynamical Range of Inclination

- Why are we missing galaxies at extreme inclination angles? (see also Shao et al. 07; Unterborn & Ryden 08)

Due to non-zero disk thickness.

Milky Way scale height ~ 90 pc (Mihilas & Binney 81)

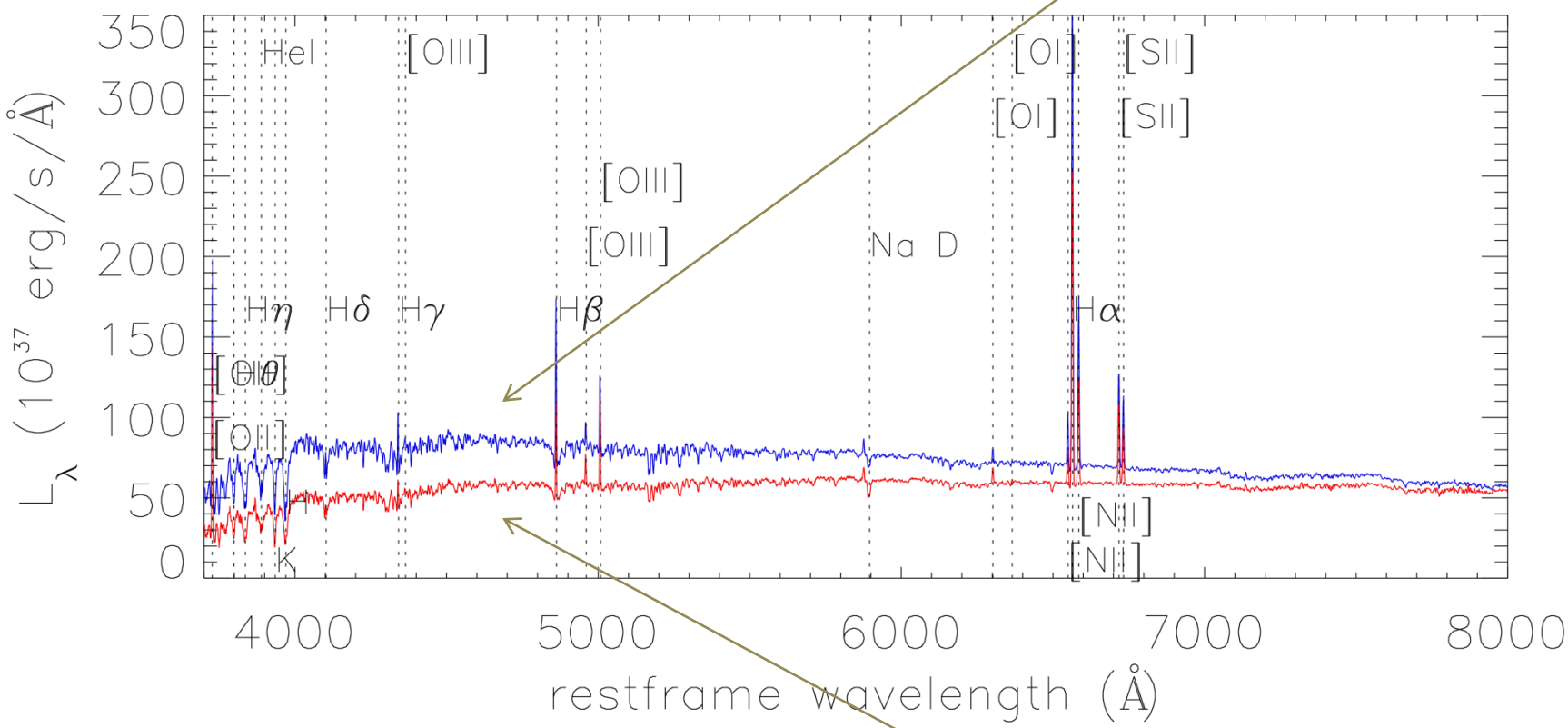


Due to non-zero disk ellipticity.

$\langle \epsilon \rangle = 0.16$ (Ryden 04)

Inclination Dependency of Disk Galaxy Composites

Face-on
<b/a> = 0.94



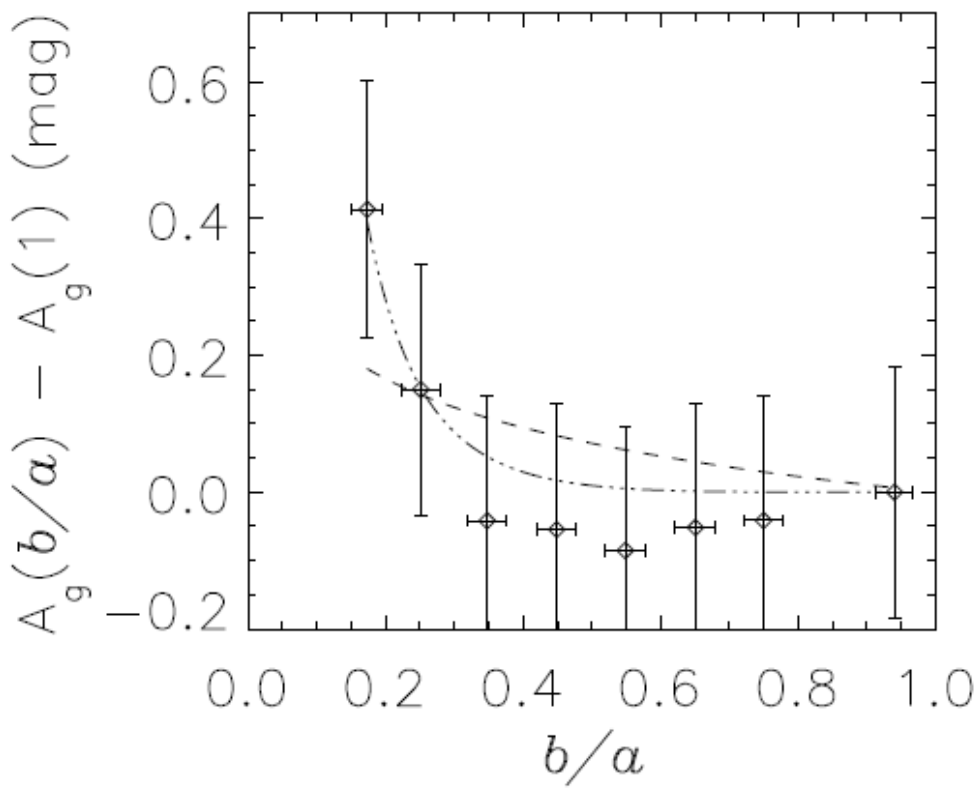
Edge-on
<b/a> = 0.17

Inclination Dependency of Extinction in Stellar Continuum

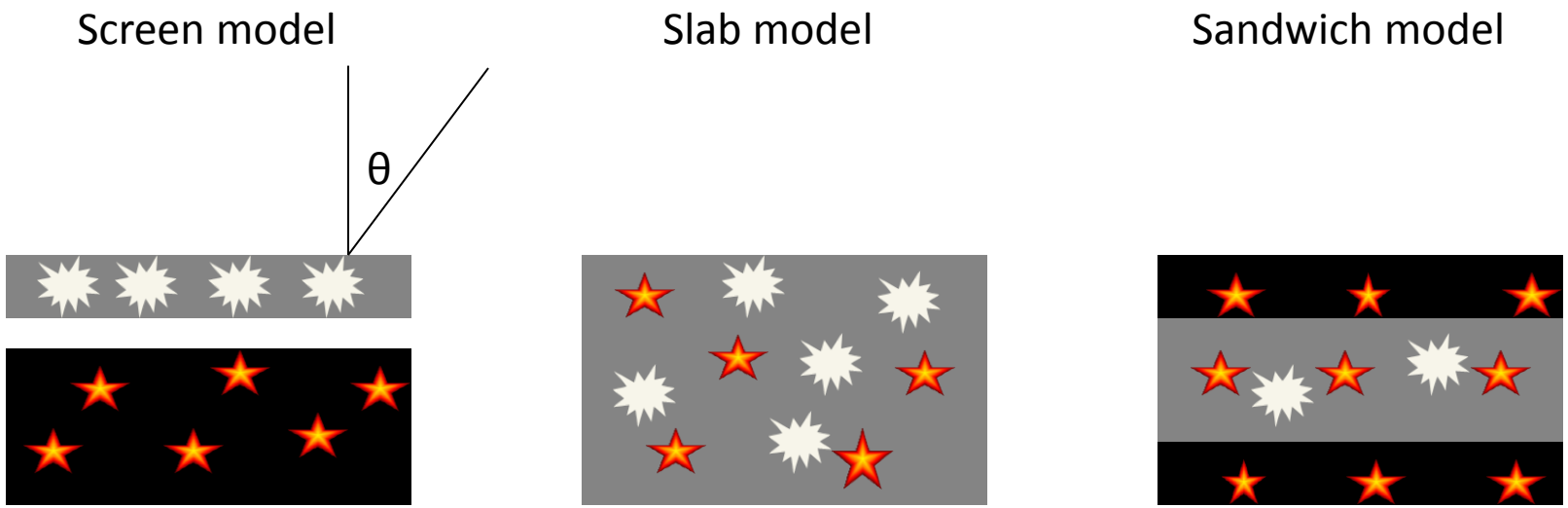
Edge-on galaxies show higher extinction than face-on galaxies.

The best-fit empirical extinction law is:

$$A_x(b/a) - A_x(1) = \eta_x \log_{10}^{4}(b/a)$$



Optical Thickness of Stellar Continuum of Galaxies



- Best-fit theoretical model is the slab model.
- Best-fit face-on extinction is 0.2 mag (SDSS g band).

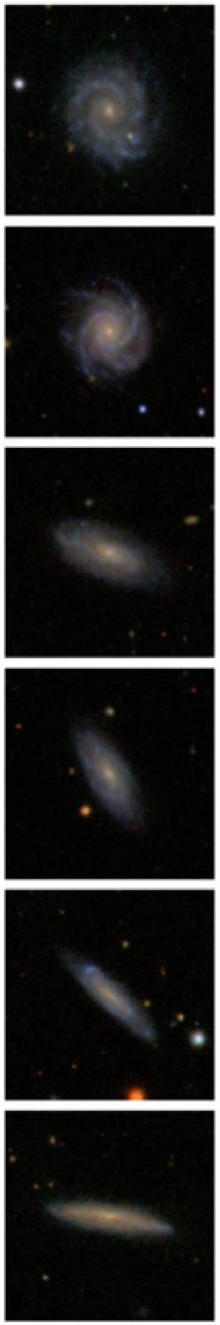
Tie Extinction to Inclination

$$\tau_\lambda (H_\alpha / H_\beta) \propto \tau_\lambda (b/a)$$

u(0.1) - u(1) 1.1 mag
g(0.1) - g(1) 0.8 mag
r(0.1) - r(1) 0.4 mag
i(0.1) - i(1) 0.2 mag
z(0.1) - z(1) 0.0 mag

Yip, Szalay, Carliles, et al. 2011

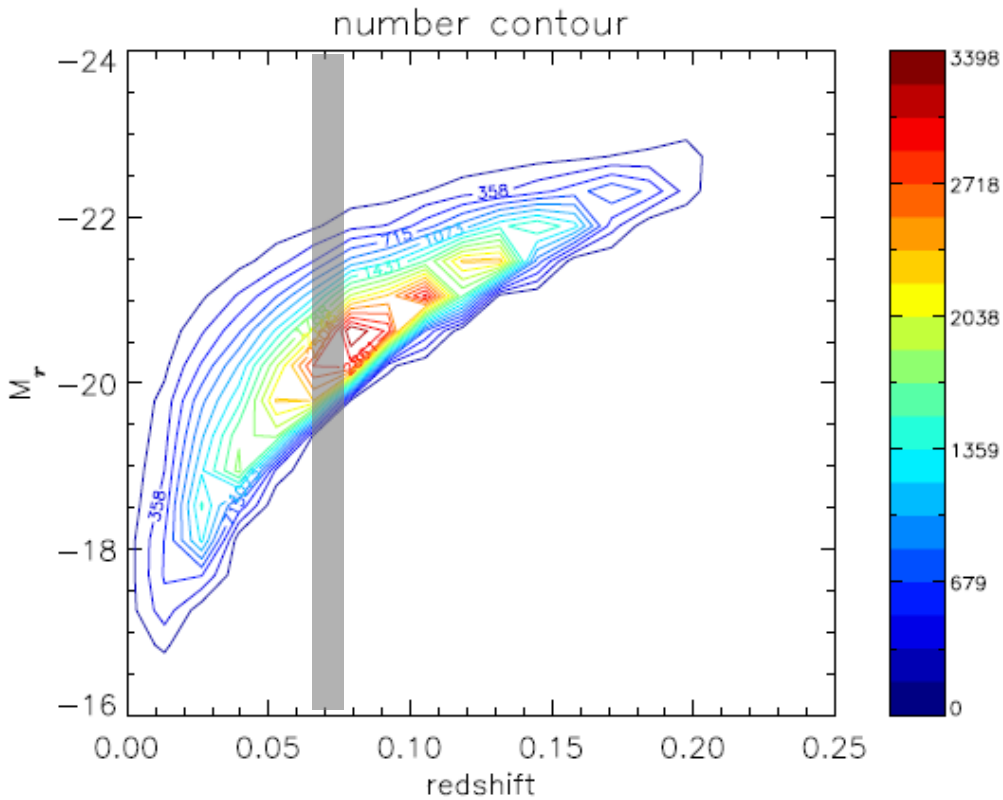
Broader Impact of Inclination Effect

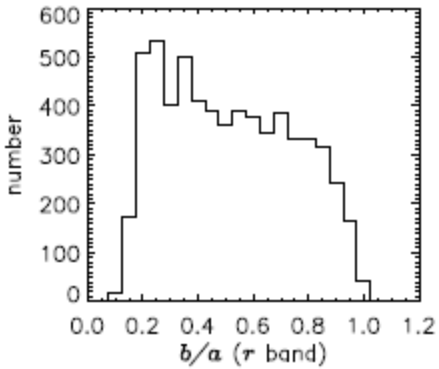


- Inclination changes the observed color and luminosity of galaxies.
- Affected galaxy properties:
  - Galaxy Count
  - Galaxy Size
  - Galaxy Luminosity Function
  - Photo-z
  - Derived parameters

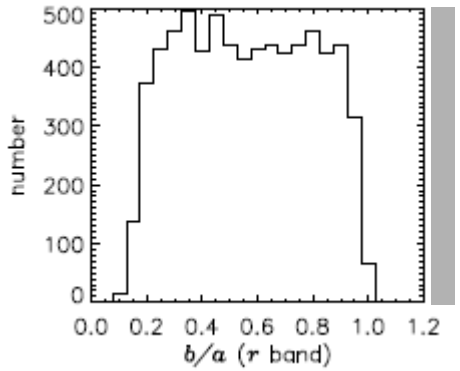
Galaxy Count

Missing edge-on galaxies at higher redshift.

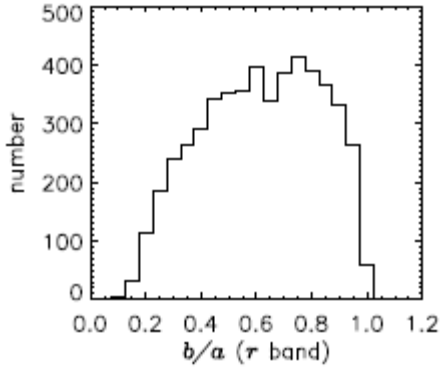




Lower Redshift 0.03–0.04



Middle Redshift 0.065–0.075



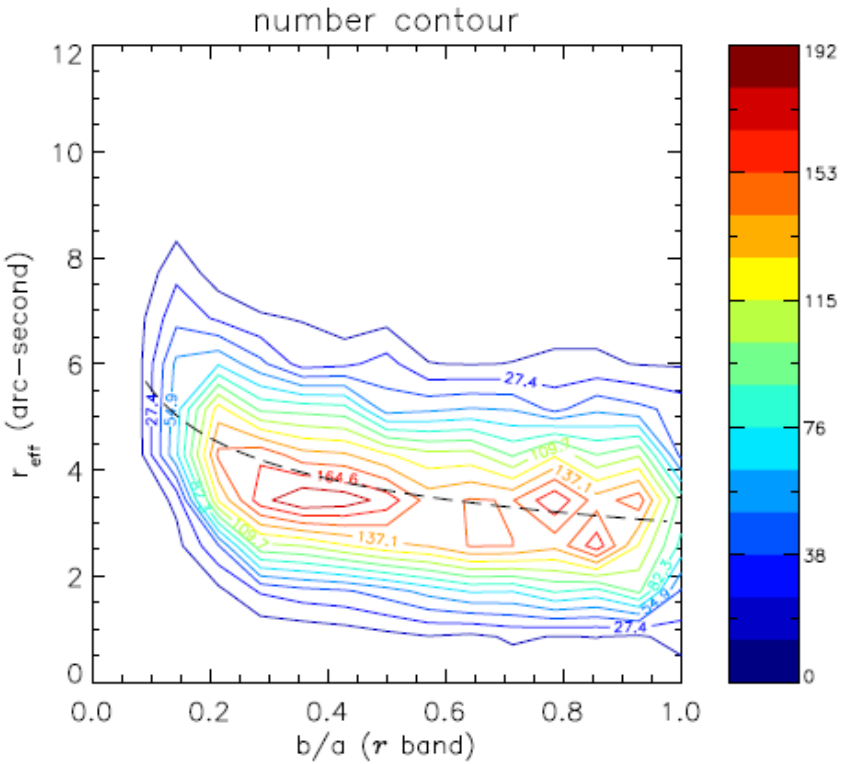
Higher Redshift 0.1–0.11

Galaxy Size

- Disk galaxies appear larger when inclined:

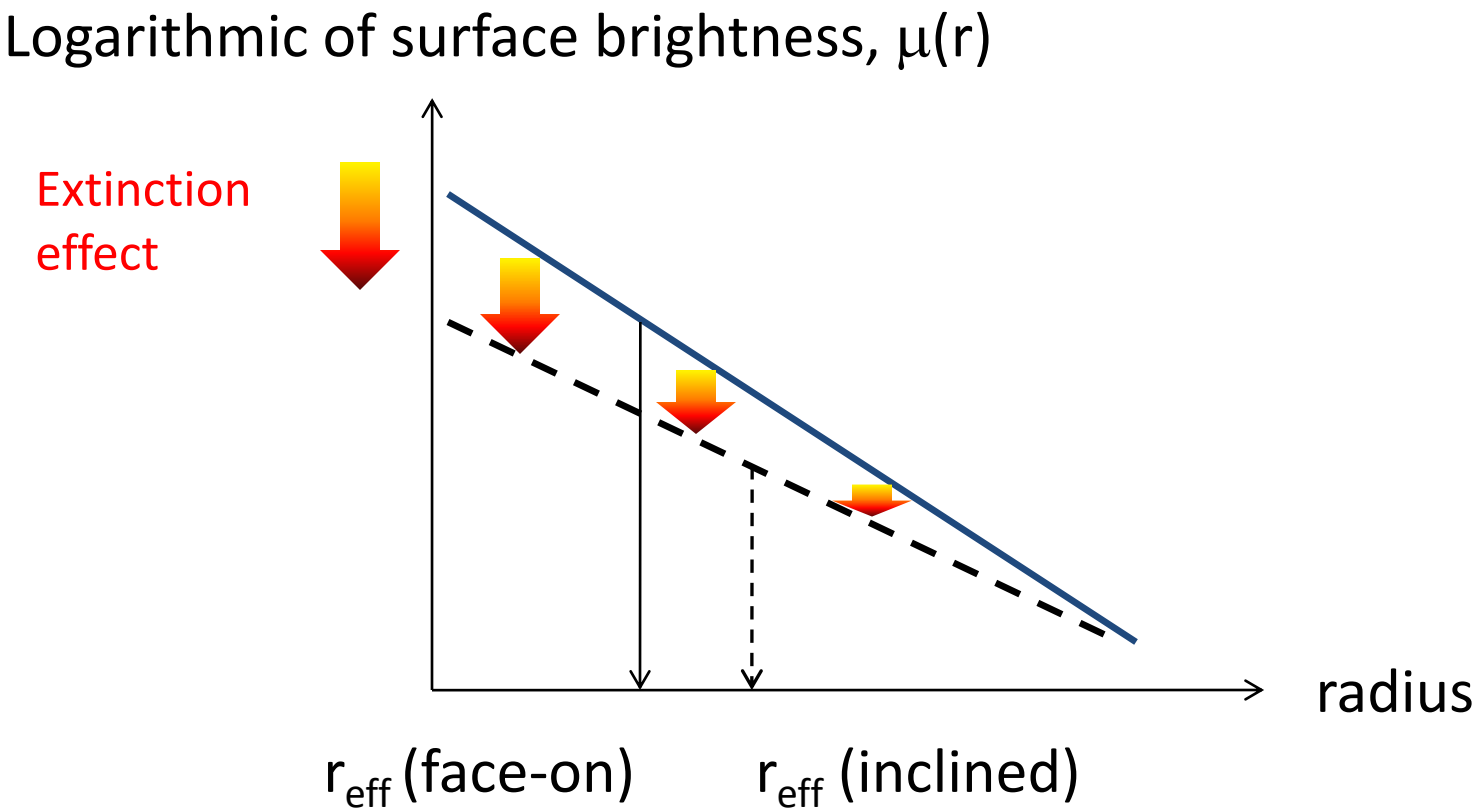
$$\log_{10}(r_{\mathrm{eff}}^{b/a}) = \log_{10}(r_{\mathrm{eff}}^{1}) - \beta_r \log_{10}(b/a)$$

- Found also in previous studies (e.g., Huizinga & van Albada 92, Mollenhoff et al. 2006, Maller et al. 09)

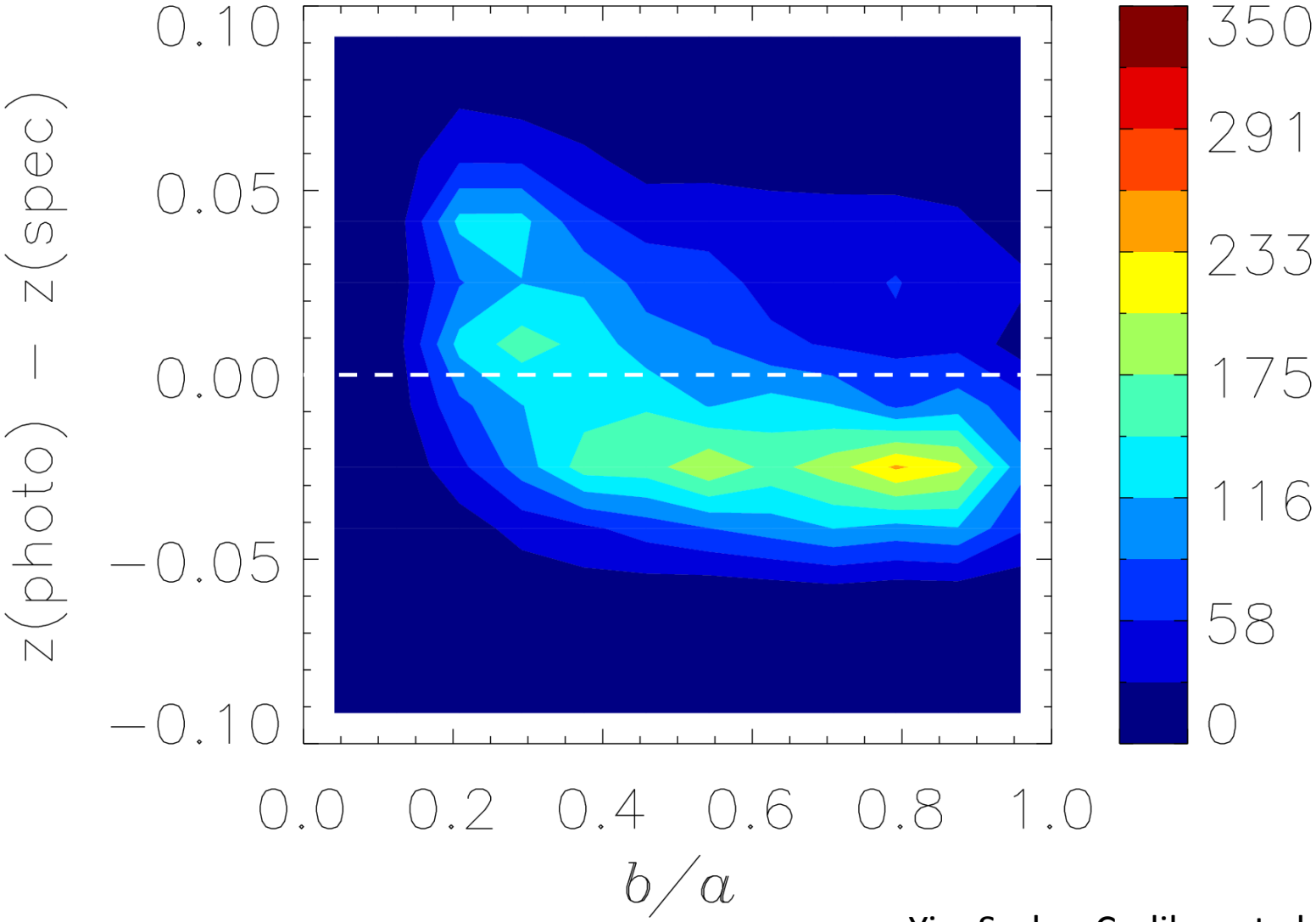


Why disk galaxies appear larger when inclined

- Because of the presence of extinction radial gradient in the galaxies.



Inclination Effect on Photo-z



Yip, Szalay, Carliles, et al. 2011

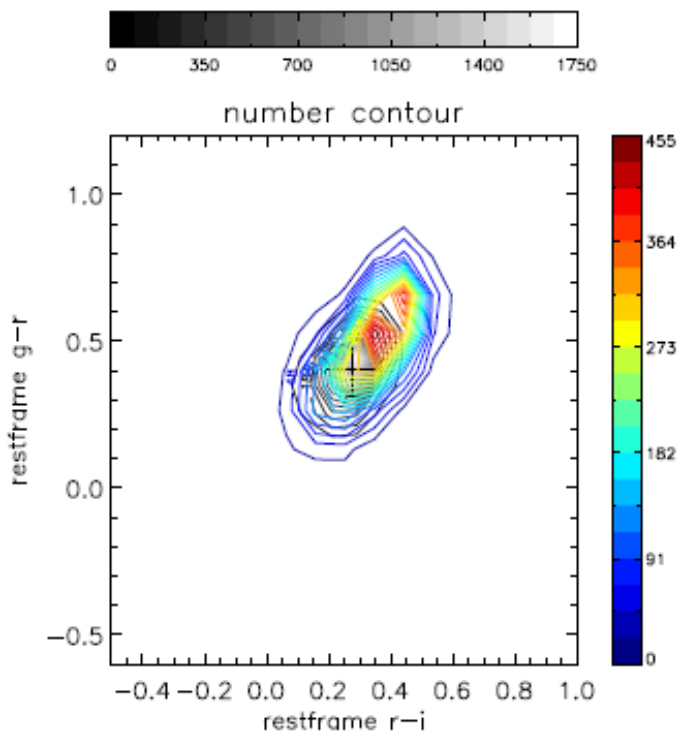
Inclination Effect on Observed Colors

- Color-color diagram of galaxies before and after inclination correction (g-r vs r-i):

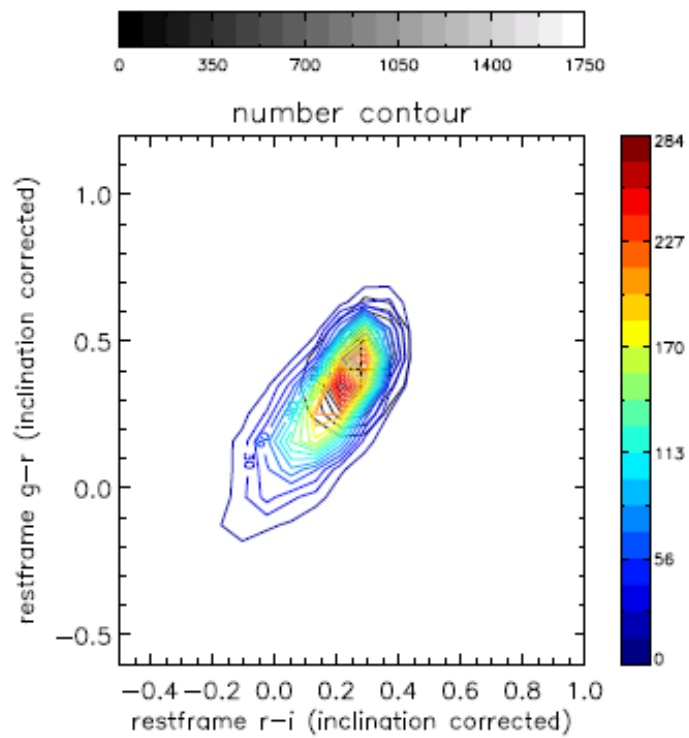
Gray: face-on galaxies

Color: edge-on galaxies

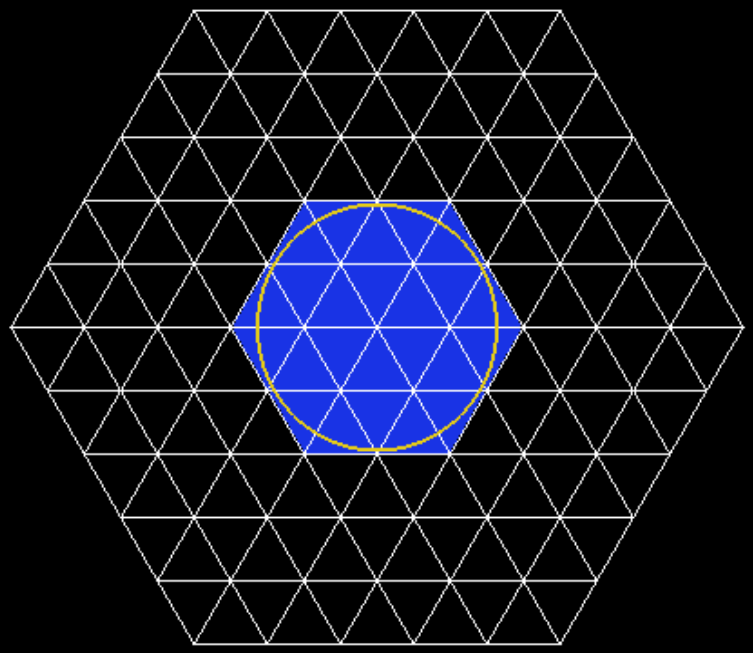
Before



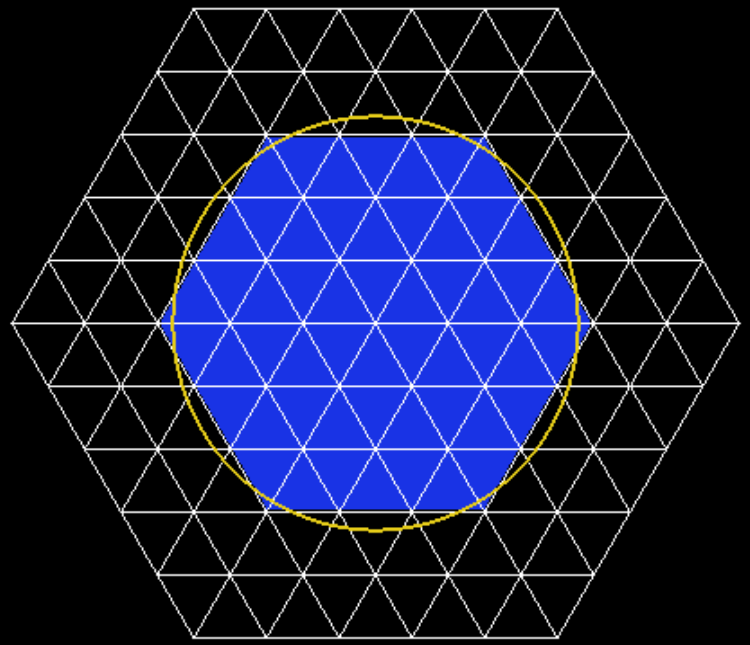
After



Drizzle2D Algorithm:
Radial Reconstruction per b/a bin using fiber spectra

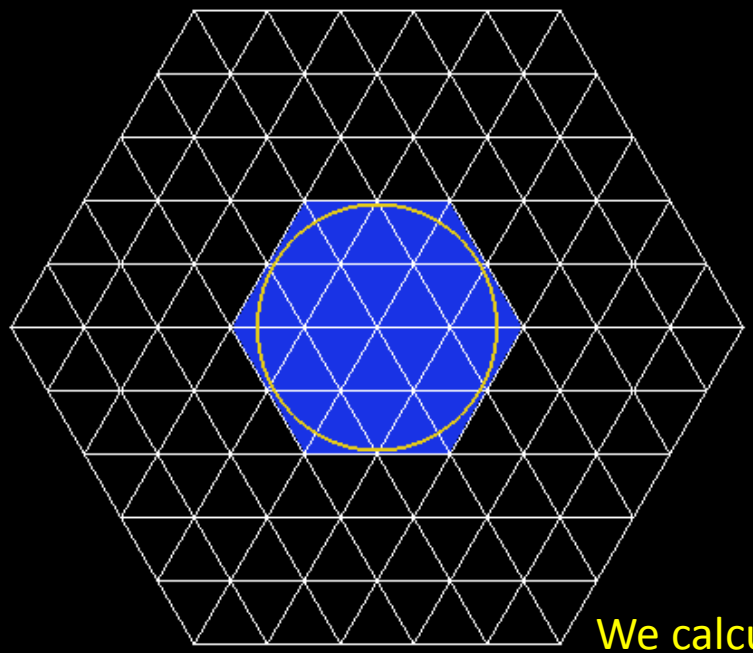


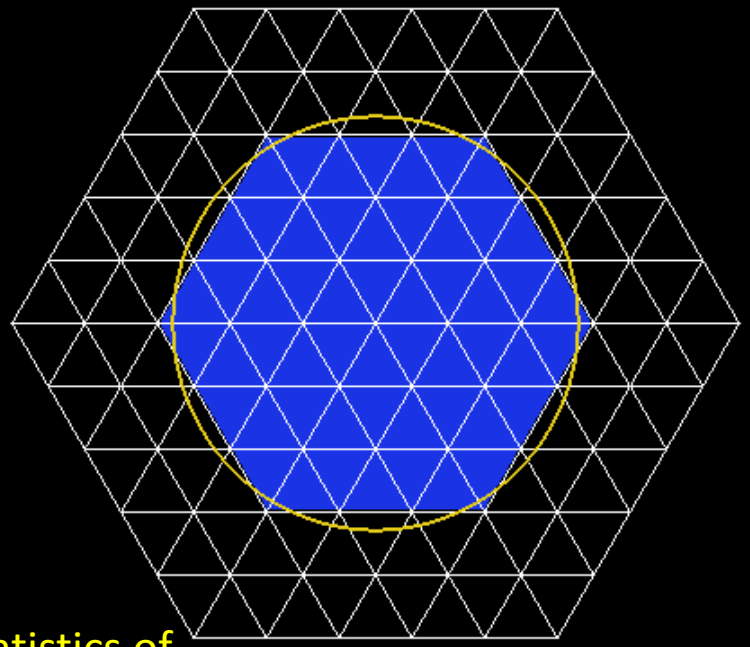
$r(\text{fiber})/r_{\text{eff}} = 0.6$



$r(\text{fiber})/r_{\text{eff}} = 1.1$

Drizzle2D Algorithm:
Radial Reconstruction per b/a bin using fiber spectra



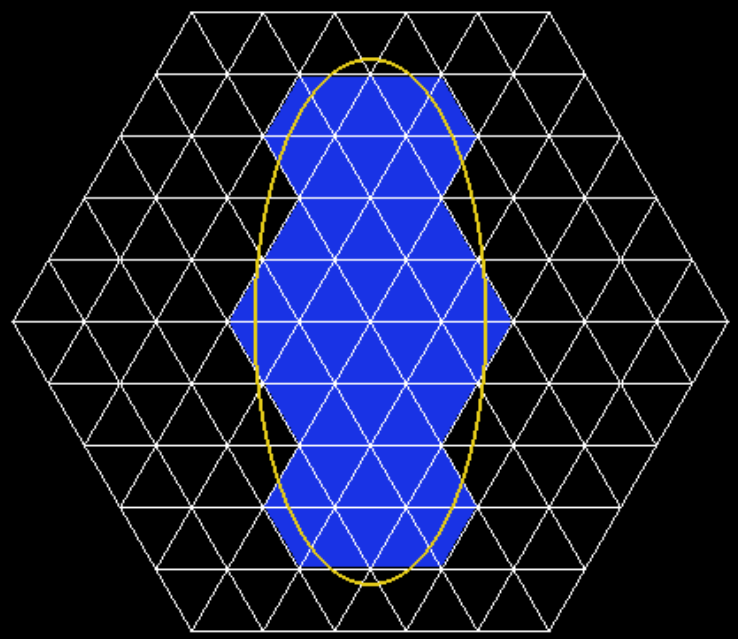


We calculate statistics of a list of spectra associated with a pixel.

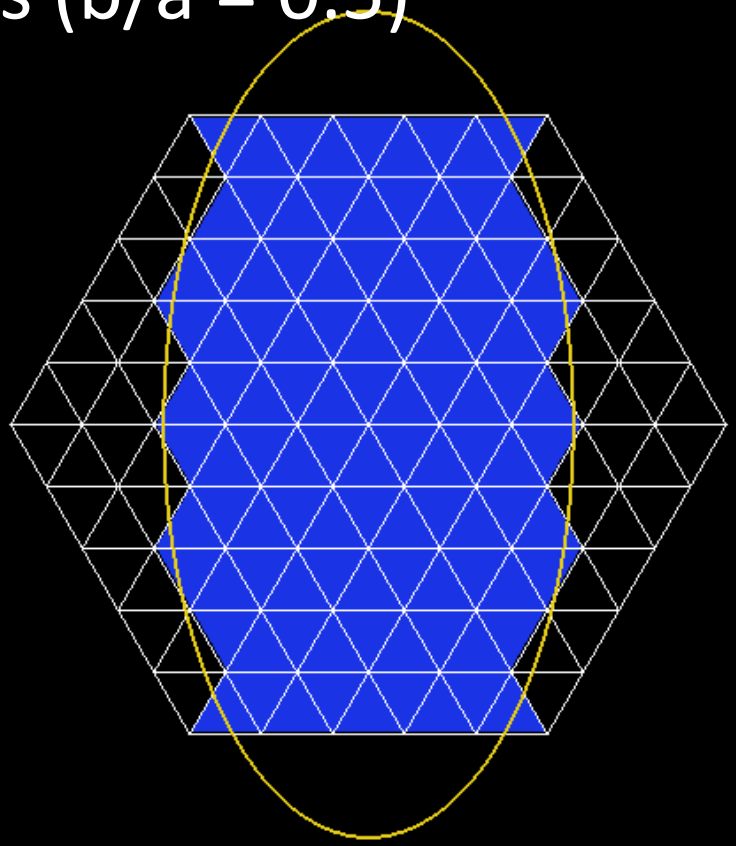
$r(\text{fiber})/r_{\text{eff}} = 0.6$

$r(\text{fiber})/r_{\text{eff}} = 1.1$

Inclined Galaxies (b/a = 0.5)



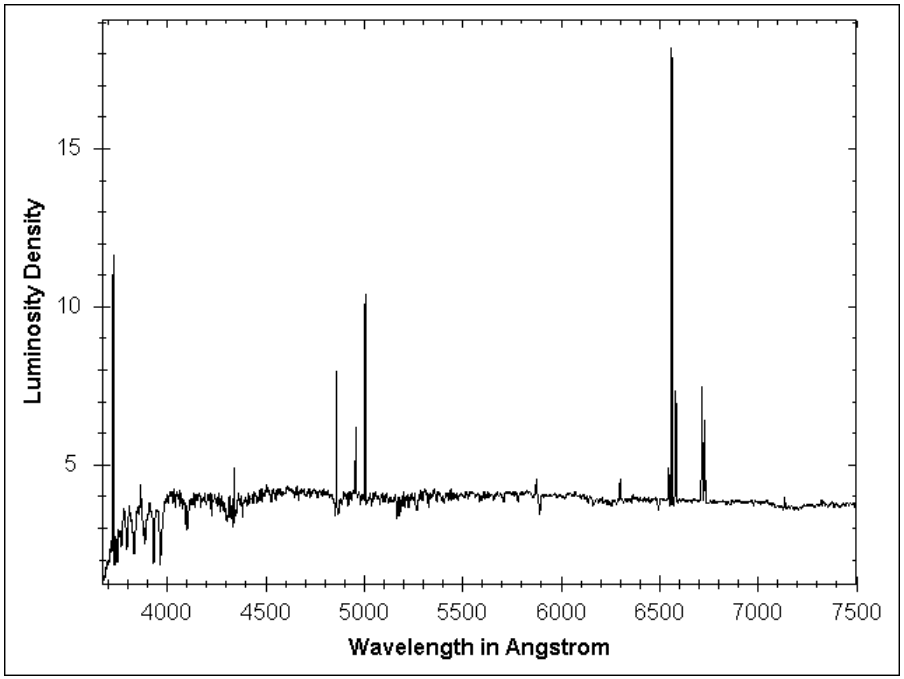
r(fiber)/r_{eff} = 0.6



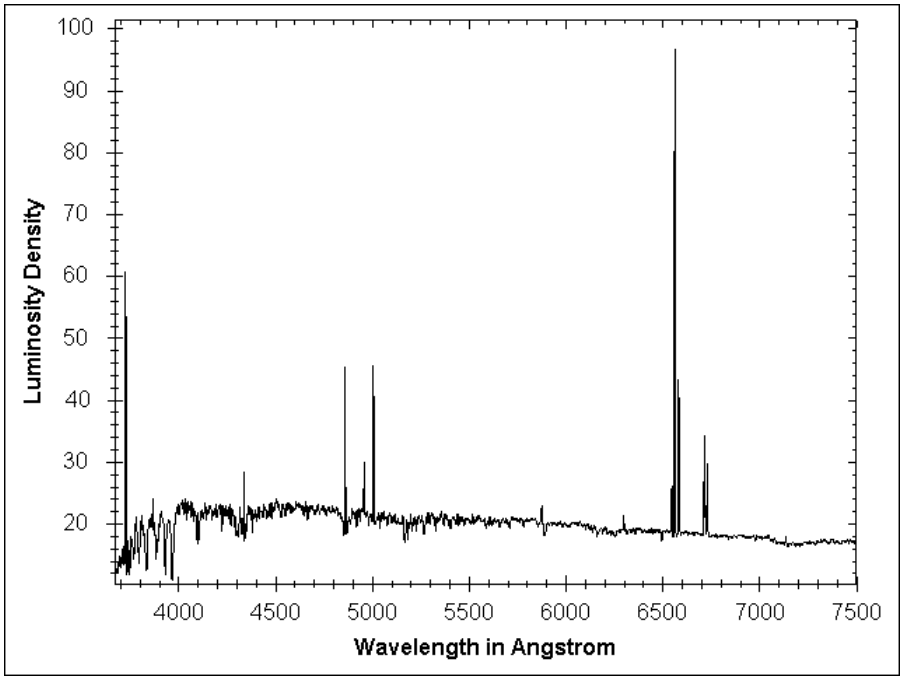
r(fiber)/r_{eff} = 1.1

Extinction at Galaxy Center (r/r_{eff} = 0.23): 1.9 mag

Edge-on
b/a = 0.1 – 0.2



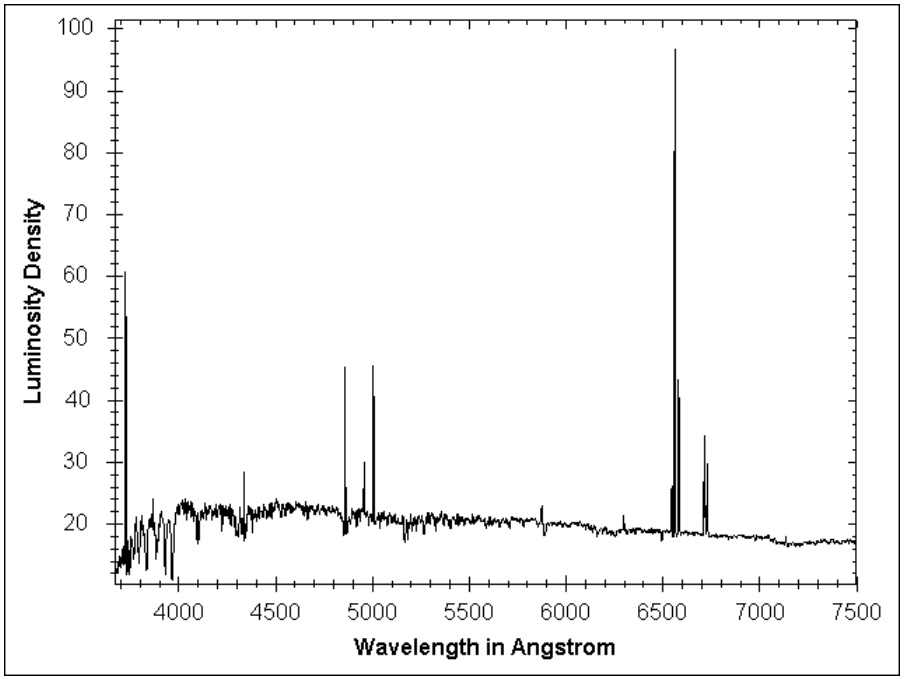
Face-on
b/a = 0.9 – 1.0



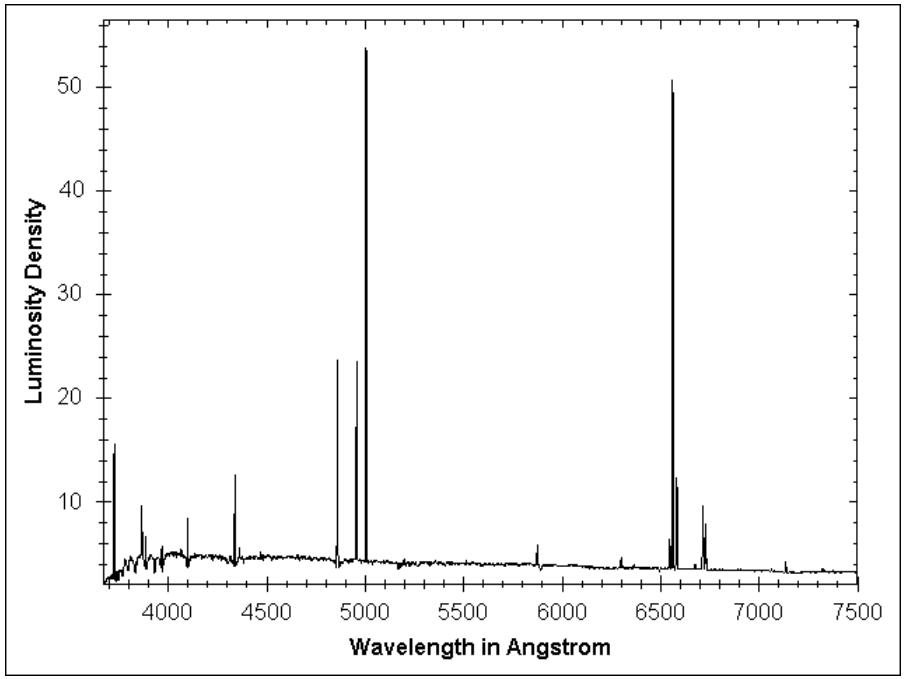
Yip et al. 2014 in prep.

Radial Profile of Galaxy (b/a = 0.9 – 1.0)

Galaxy Center
$r/r_{eff} = 0.23$

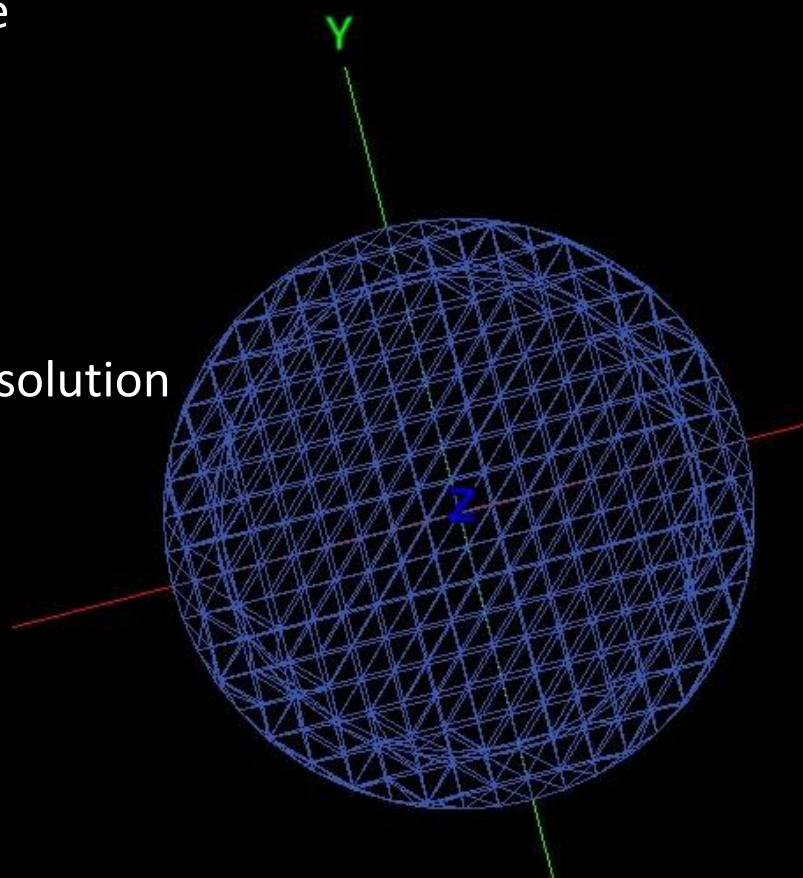


Galaxy Outer Region
$r/r_{eff} = 1.15$



Future: Drizzle3D Algorithm

- Deproject 2D composites f(x, y; b/a) to get 3D f(x, y, z)
- Much like 3D human body reconstruction using X-ray CAT scans
- Key components:
  - 3D Galaxy Mesh & Tessellation Scheme
  - Galaxy opacity profile
  - Parameterization of voxel spectra
    - Principal Component Analysis
  - Flux convergence in individual voxels
  - Interpolation for higher-than-native resolution
    - With Emmanuel Candes @ Stanford
  - 3D Visualization

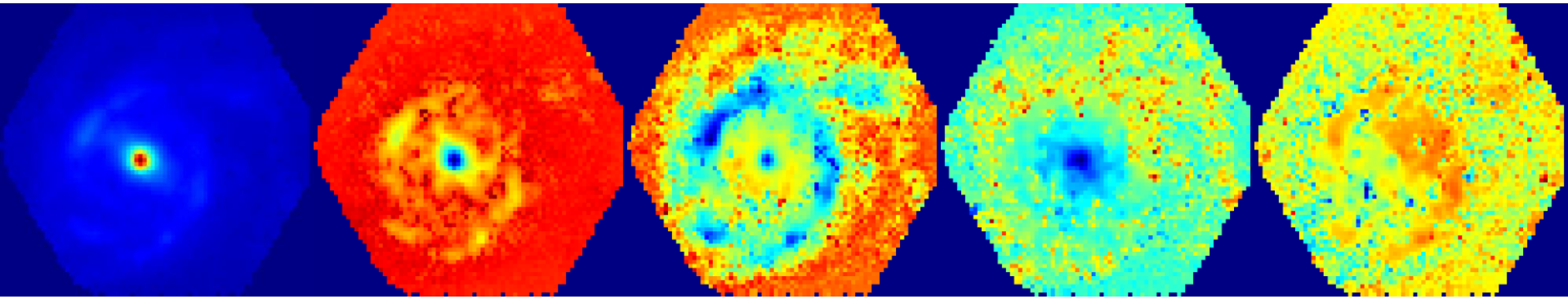


Parameterization of Spectra using Principal Component Analysis

- Compress datacube from 78x73x1700 to 78x73x5.

NGC 5409 SDSS image.

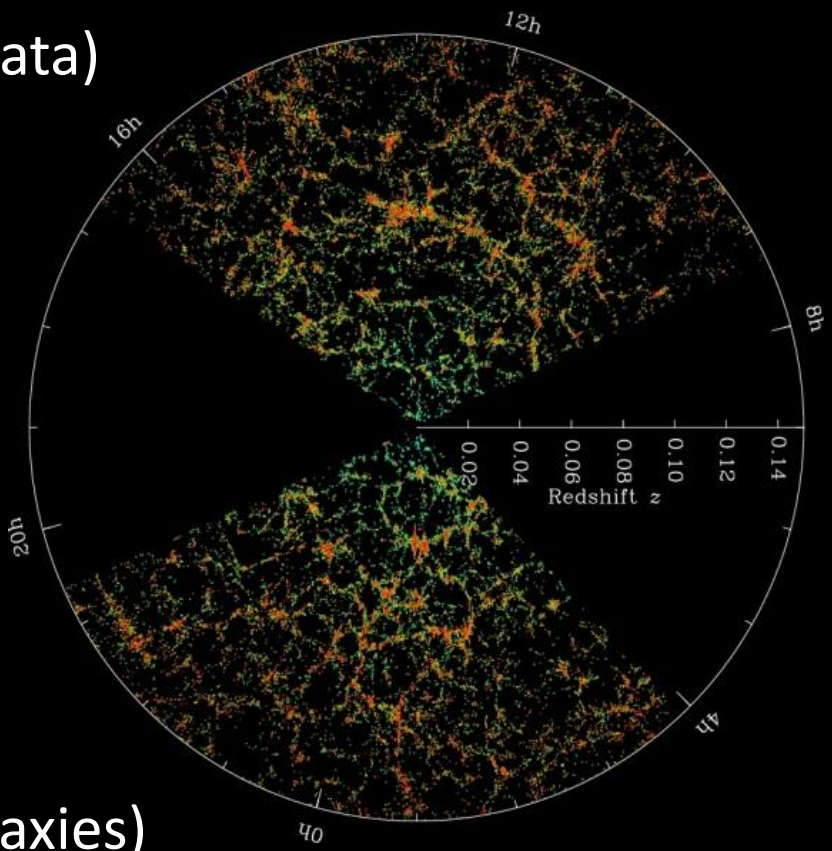
Top 5 PCA Eigencoefficient map.



IFU data from CALIFA DR1: 3650 – 4840 Å at 0.7 Å / pixel.
Husemann et al. 2012, Sanchez et al. 2012, Walcher et al. in prep.

Challenges in Probing Galaxies' Properties

- Many Objects/Spectra (Big Data)
- Many Parameters
- Noisy Data



SDSS & BOSS (z = 0 – 0.7, 2.5M galaxies)
LAMOST (z = 0 – 0.2, 10M galaxies)
Prime Focus Spectrograph (z = 1 – 2, 200K galaxies)

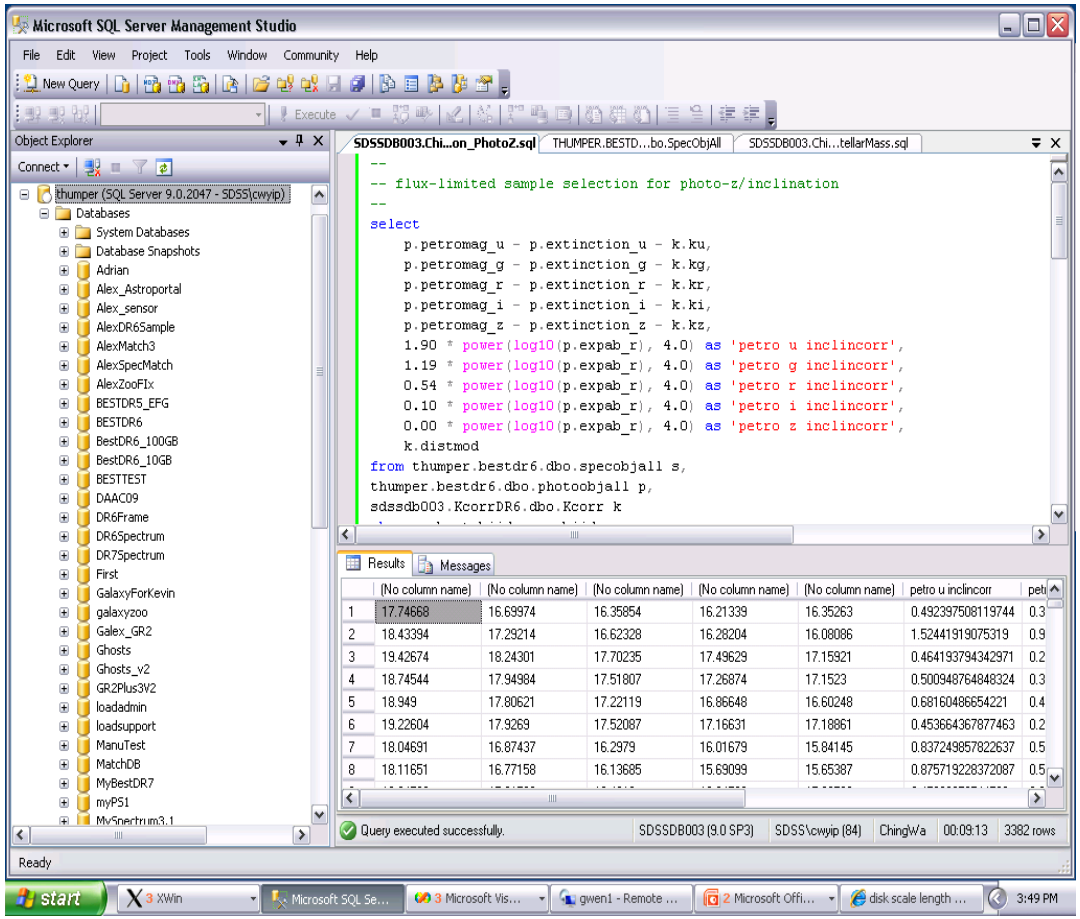
Data Analysis using Database

- Automated data analysis:

Select data from DB using C# routines with SQL scripts embedded

Perform computations

Output results to DB, if necessary

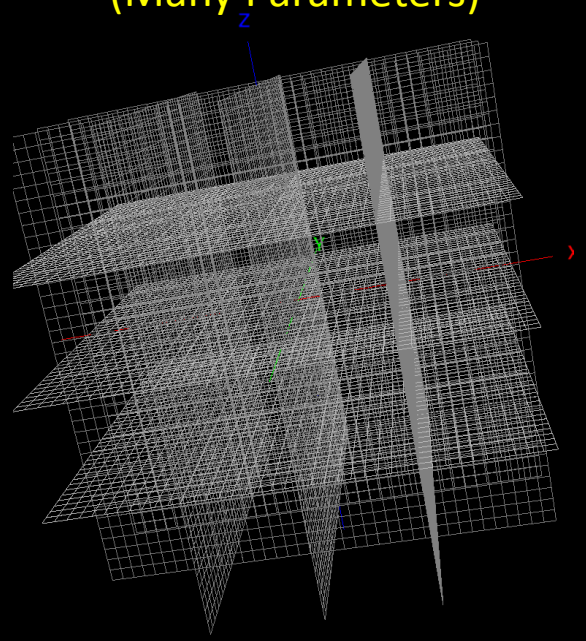


MS SQL Server. Source: Alex Szalay

Probing Galaxies' Properties

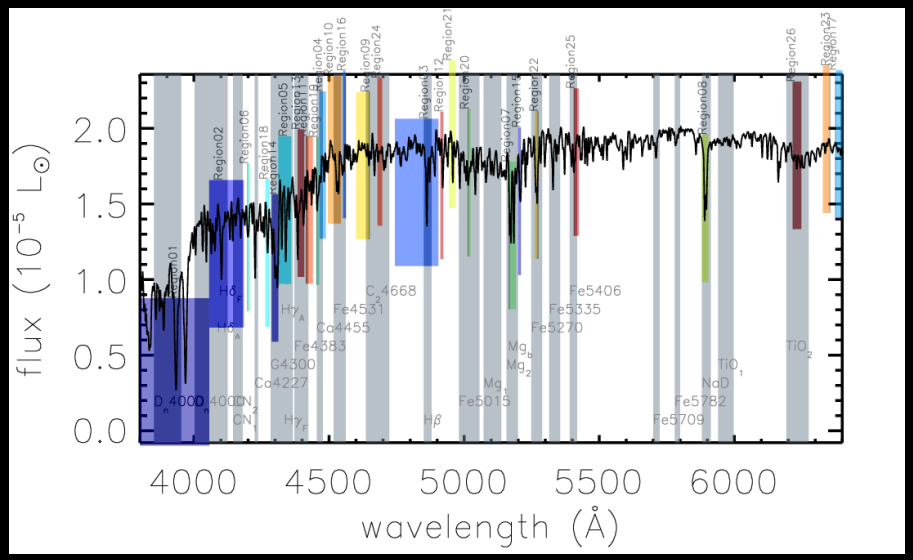
- Model Fitting of Spectra (SDSS/JHU Catalog: Kauffmann, Heckman, Tremonti, Chen, et al.)
- Empirical Relations

(Many Parameters)



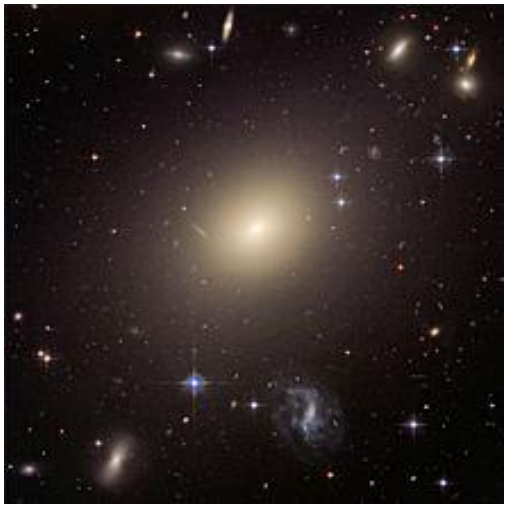
N-Dimensional Model Grid
(Yip 2010)

(Many Objects)



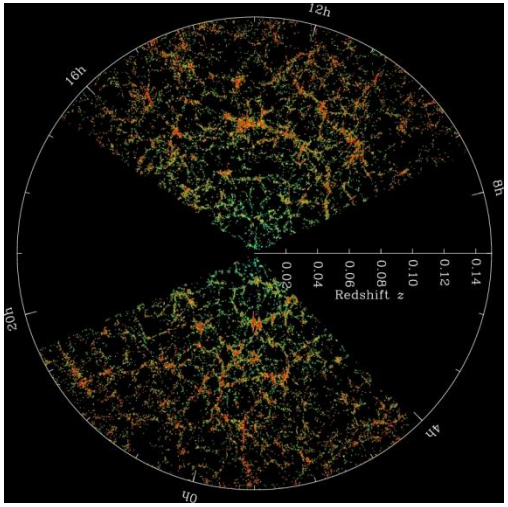
Informative Wavelength Regions using CUR
(Yip, Mahoney, Szalay, et al. 2014 AJ in press)

Probing Galaxies' Properties (Many Objects)



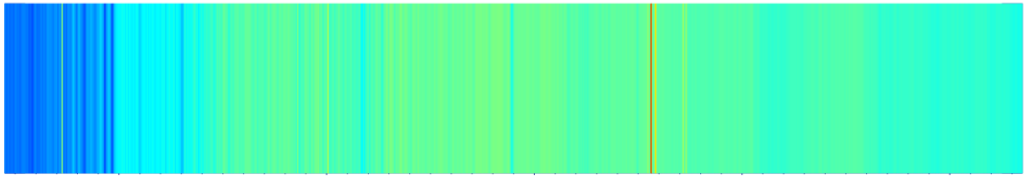
Galaxy Parameter Estimation:

- Stellar Age
- Stellar Metallicity
- Star Formation Rate
- Dust Extinction
- etc.



Brute force Bayesian approach:

N = 1 Million	Galaxy Spectra
M = 4,000	Pixels
P = 4	Parameters
T = 20 years	CPU time



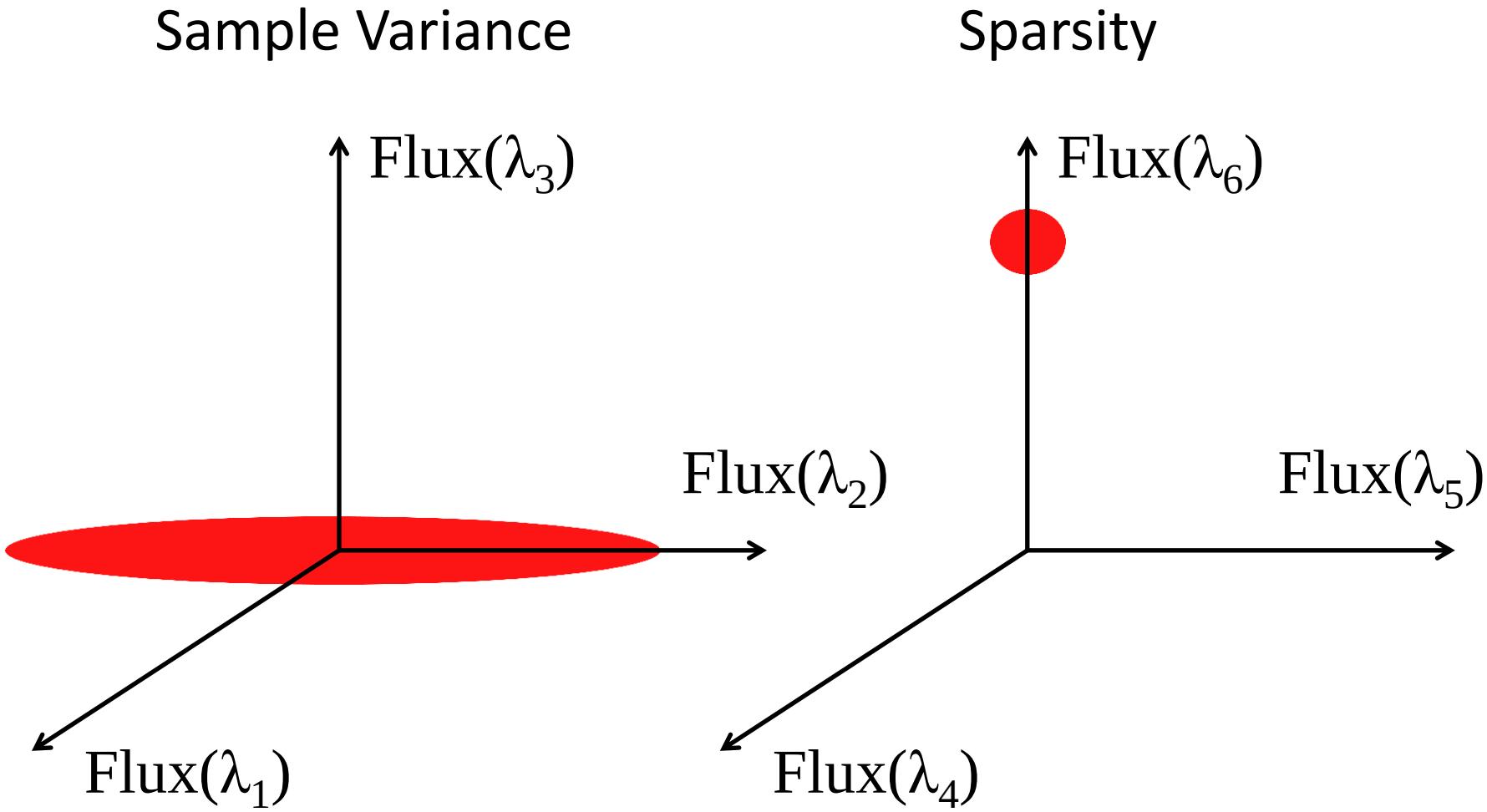
Lick Indices

Worthey et al. 94;
Trager et al. 98

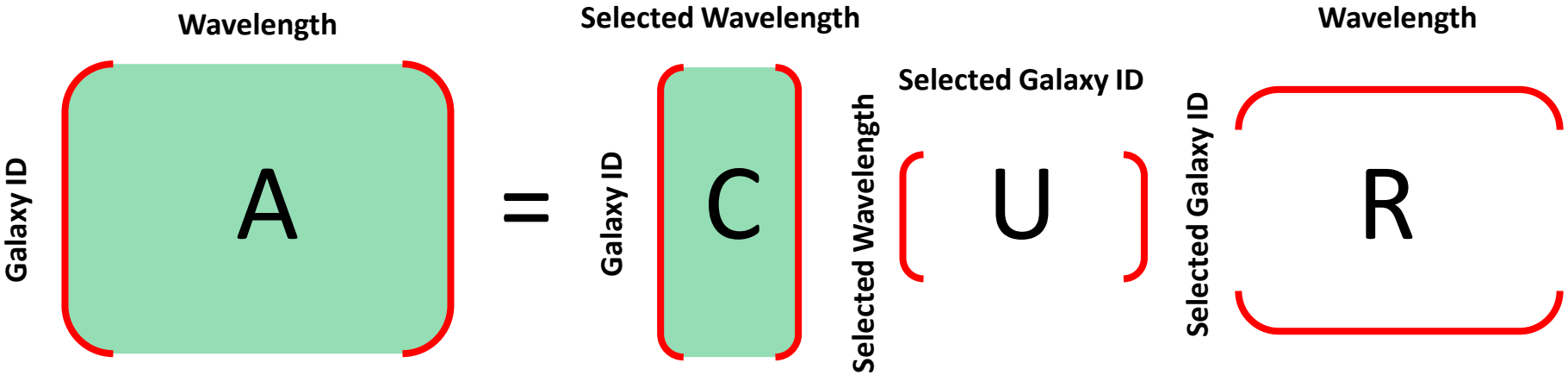
INDEX DEFINITIONS

	Name	Index Bandpass	Pseudocontinua	Units	Measures	Error[1]	Notes
01	CN_1	4143.375-4178.375	4081.375-4118.875 4245.375-4285.375	mag	CN, Fe I	0.021	
02	CN_2	4143.375-4178.375	4085.125-4097.625 4245.375-4285.375	mag	CN, Fe I	0.023	2
03	Ca4227	4223.500-4236.000	4212.250-4221.000 4242.250-4252.250	Å	Ca I, Fe I, Fe II	0.27	2
04	G4300	4282.625-4317.625	4267.625-4283.875 4320.125-4336.375	Å	CH, Fe I	0.39	
05	Fe4383	4370.375-4421.625	4360.375-4371.625 4444.125-4456.625	Å	Fe I, Ti II	0.53	2
06	Ca4455	4453.375-4475.875	4447.125-4455.875 4478.375-4493.375	Å	Ca I, Fe I, Ni I, Ti II, Mn I, V I	0.25	2
07	Fe4531	4515.500-4560.500	4505.500-4515.500 4561.750-4580.500	Å	Fe I, Ti I, Fe II, Ti II	0.42	2
08	Fe4668	4635.250-4721.500	4612.750-4631.500 4744.000-4757.750	Å	Fe I, Ti I, Cr I, Mg I, Ni I, C_2	0.64	2
09	Hβ	4847.875-4876.625	4827.875-4847.875 4876.625-4891.625	Å	Hβ, Fe I	0.22	3
10	Fe5015	4977.750-5054.000	4946.500-4977.750 5054.000-5065.250	Å	Fe I, Ni I, Ti I	0.46	2,3
11	Mg_1	5069.125-5134.125	4895.125-4957.625 5301.125-5366.125	mag	MgH, Fe I, Ni I	0.007	3
12	Mg_2	5154.125-5196.625	4895.125-4957.625 5301.125-5366.125	mag	MgH, Mg *b*, Fe I	0.008	3
13	Mg *b*	5160.125-5192.625	5142.625-5161.375 5191.375-5206.375	Å	Mg *b*	0.23	3
14	Fe5270	5245.650-5285.650	5233.150-5248.150 5285.650-5318.150	Å	Fe I, Ca I	0.28	3
15	Fe5335	5312.125-5352.125	5304.625-5315.875 5353.375-5363.375	Å	Fe I	0.26	3
16	Fe5406	5387.500-5415.000	5376.250-5387.500 5415.000-5425.000	Å	Fe I, Cr I	0.20	2,3
17	Fe5709	5698.375-5722.125	5674.625-5698.375 5724.625-5738.375	Å	Fe I, Ni I, Mg I Cr I, V I	0.18	2
18	Fe5782	5778.375-5798.375	5767.125-5777.125 5799.625-5813.375	Å	Fe I, Cr I Cu I, Mg I	0.20	2
19	Na D	5878.625-5911.125	5862.375-5877.375 5923.875-5949.875	Å	Na I	0.24	
20	TiO_1	5938.375-5995.875	5818.375-5850.875 6040.375-6105.375	mag	TiO	0.007	
21	TiO_2	6191.375-6273.875	6068.375-6143.375 6374.375-6416.875	mag	TiO	0.006	

Some Wavelengths are More Informative: Leverage Score



CUR Matrix Decomposition (Mahoney & Drineas 2009): Best Rank-k Approximation of Data Matrix



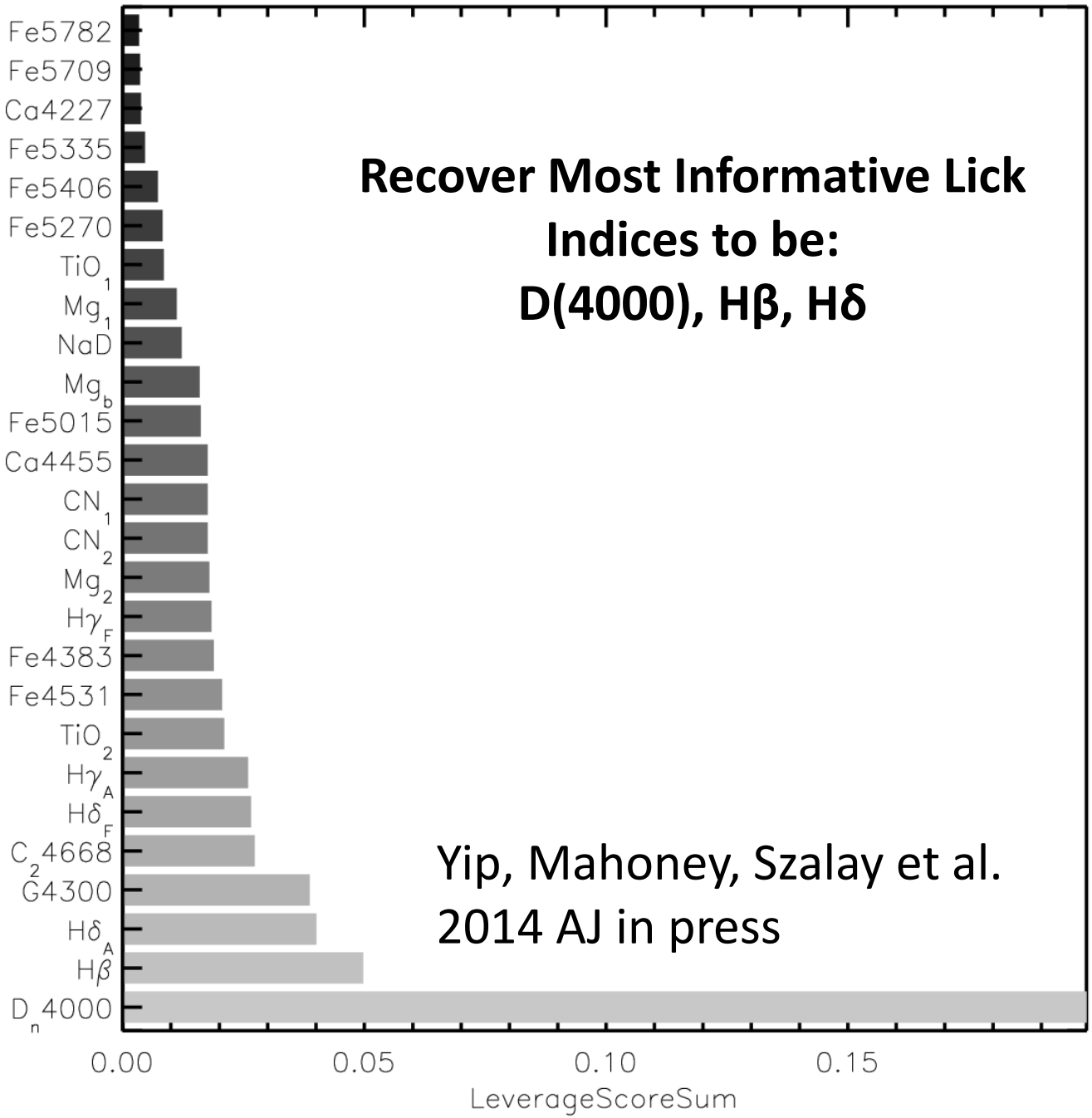
CUR approximates data matrix A as such:

$$min\|A - CUR\|_F$$

Identifying Objective Wavelength Regions using CUR Matrix Decomposition

- Simple Stellar Populations (SSP) defined by stellar age (0 – 20 Gyr) and stellar metallicity (Z = 0.0001, 0.0004, 0.004, 0.008, 0.02, 0.05).
  - Bruzual and Charlot 2003
  - Absorption Lines Only, No Emission Lines

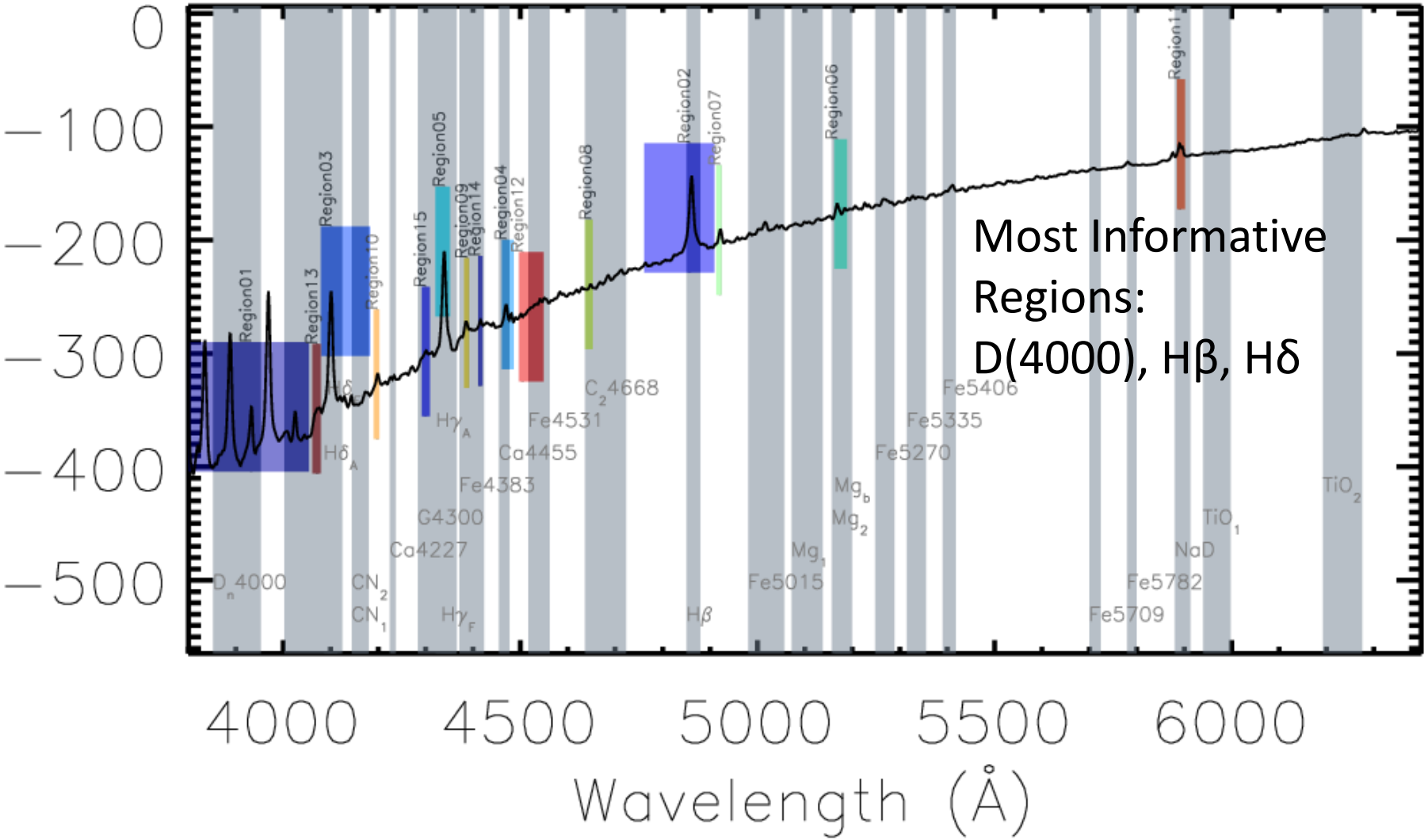
- 2 configurations:
  - Lick: 3800-6400 Å with 9 Å resolution
  - SDSS: 3450-8350 Å with 3 Å resolution



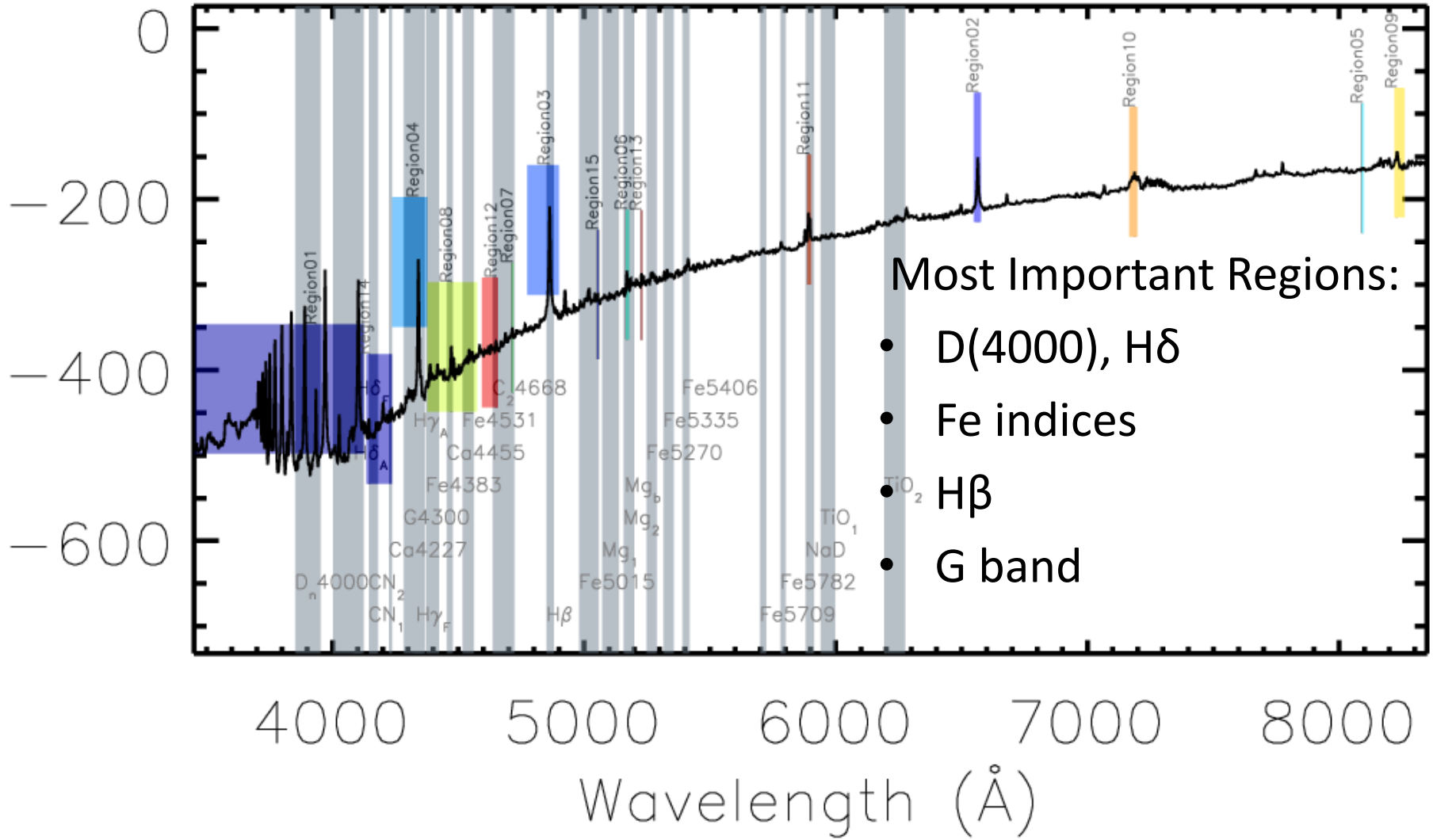
Recover Most Informative Lick Indices to be: D(4000), Hβ, Hδ

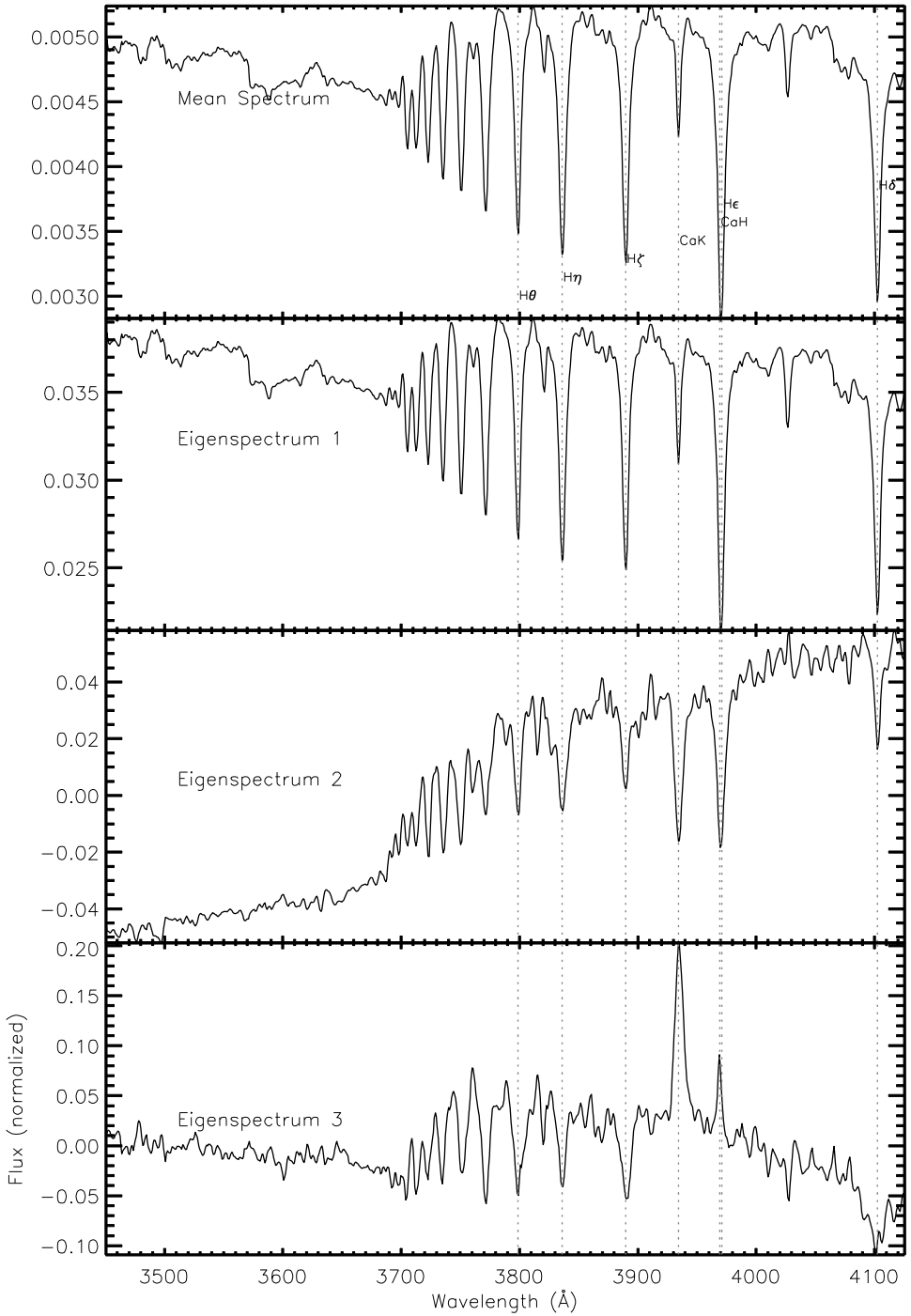
Yip, Mahoney, Szalay et al. 2014 AJ in press

9Å Resolution



3Å Resolution





PCA Eigenspectra of 1st Informative Region

Larger a1 implies:

Younger stellar populations

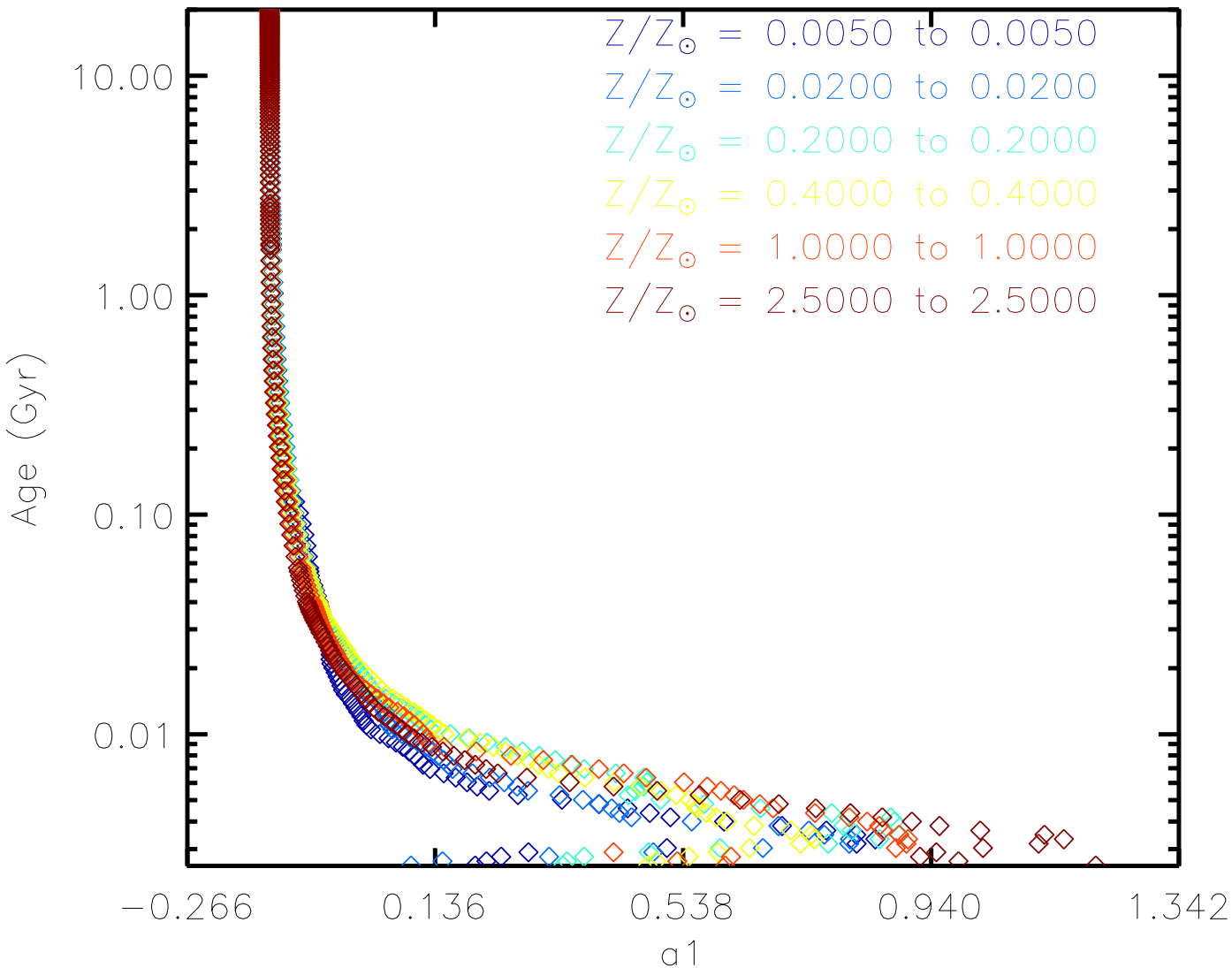
Larger a2 implies:

Younger, More Metal Rich

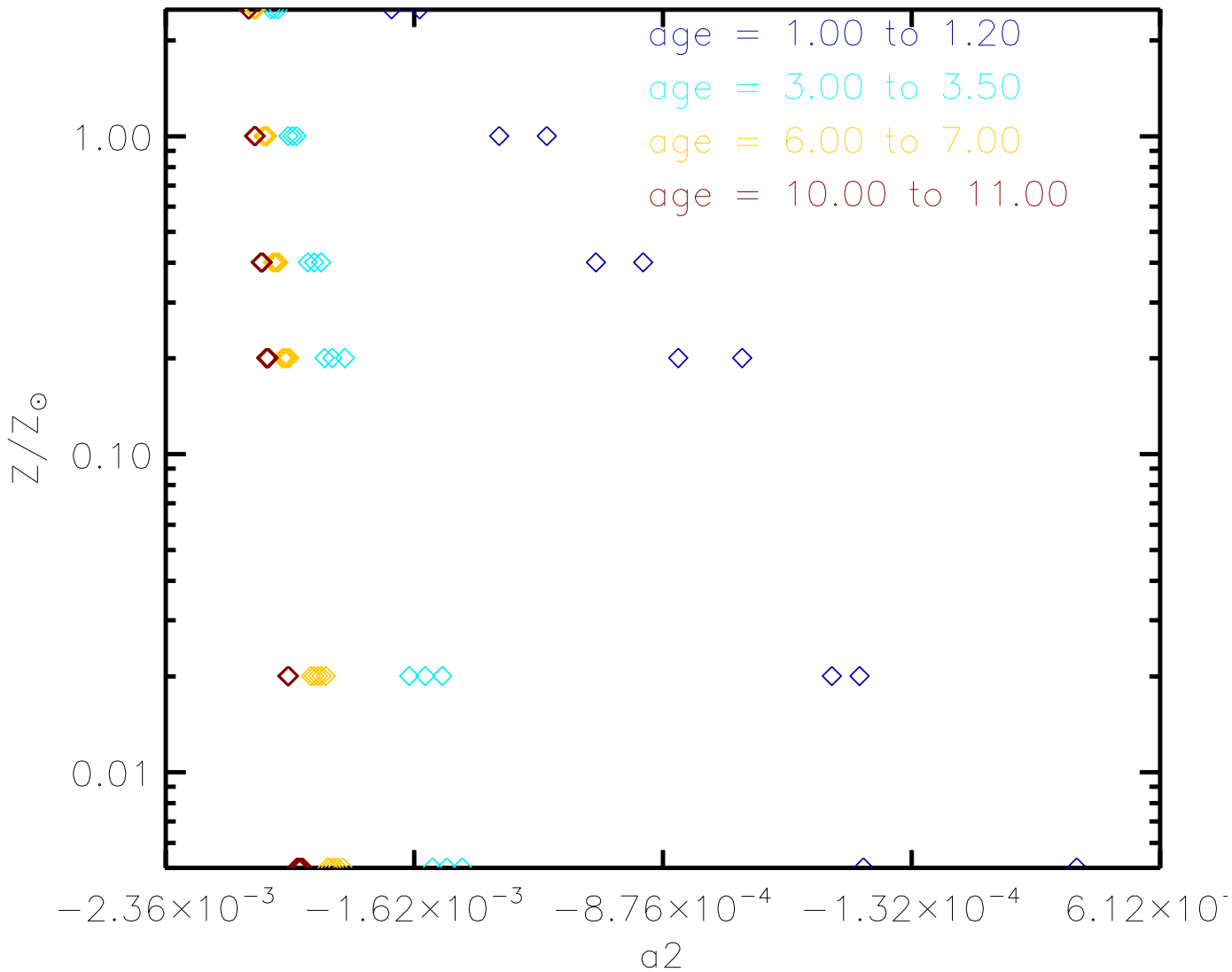
OR

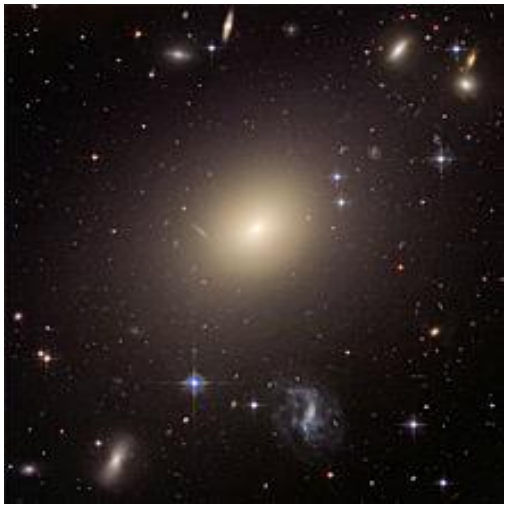
Older, Less Metal Rich

Stellar Age Sensitivity of Wavelength Regions



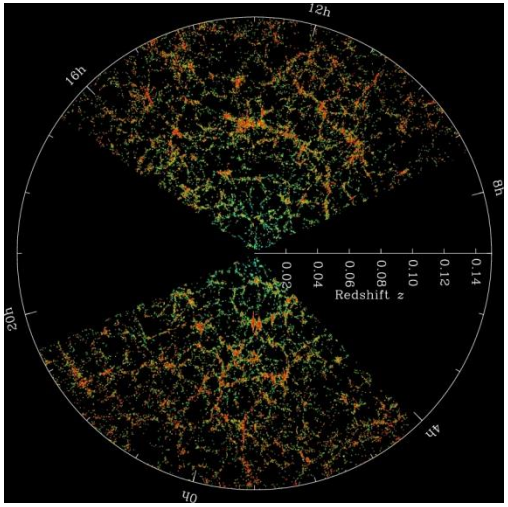
Stellar Metallicity Sensitivity of Wavelength Regions (Seeing Age-Metallicity Degeneracy!)





Galaxy Parameter Estimation:

- Stellar Age
- Stellar Metallicity
- Star Formation Rate
- Dust Extinction
- etc.



Brute force Bayesian approach:

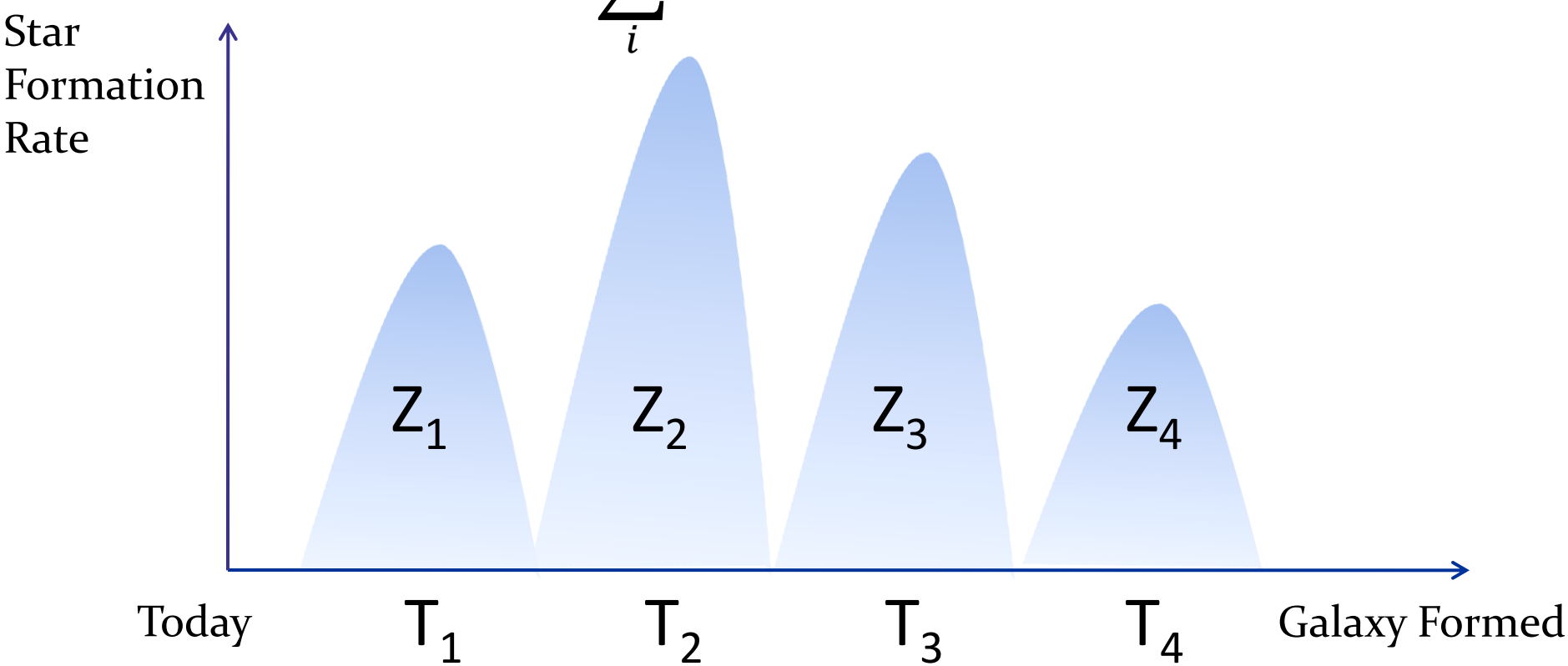
N = 1 Million	Galaxy Spectra
M = 4,000	Pixels
P = 4	Parameters
T = 20 years	CPU time

2 weeks!

Model Stellar Light in Galaxy Spectra

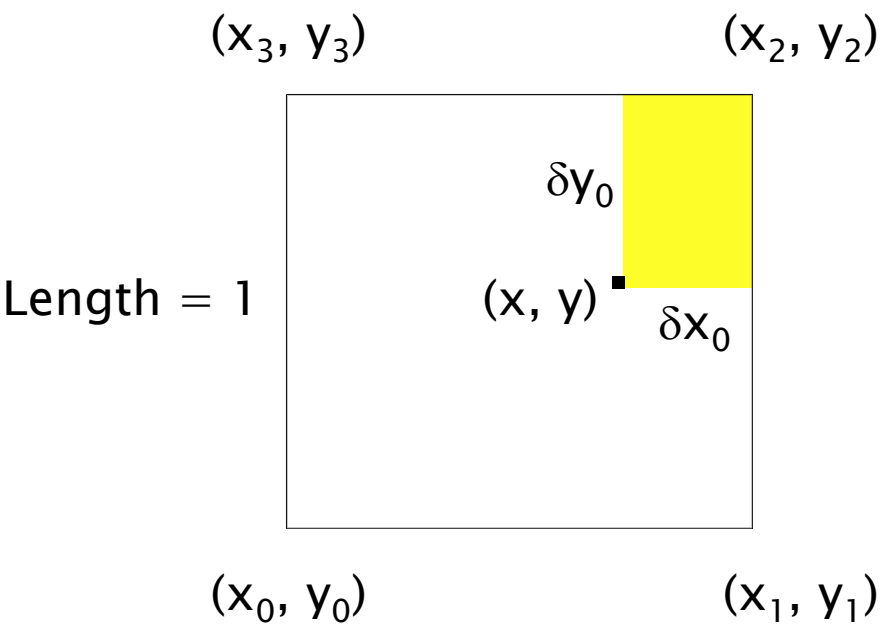
- We model each galaxy spectrum by a linear combination of Single Stellar Populations (Tinsley 70'):

$$f_i^{galaxy} = \sum_i a_i \, f_i^{star}(Age, Metallicity)$$



Galaxy Composition through N-D Parameter Estimation

- N-D hypercube
- Multi-linear interpolation to achieve arbitrary computational resolution in parameters



2-D:

$$
\begin{aligned}
f(x, y) = \; & f(x_0, y_0) * (\delta x_0 * \delta y_0) \\
& + f(x_1, y_1) * (\delta x_1 * \delta y_1) \\
& + f(x_2, y_2) * (\delta x_2 * \delta y_2) \\
& + f(x_3, y_3) * (\delta x_3 * \delta y_3)
\end{aligned}
$$

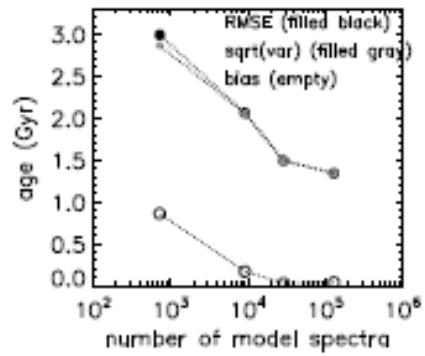
N-D:

$$f(x) = \Sigma\, f(zi)\, \Pi\, (1 - |xj - zij|\,)$$

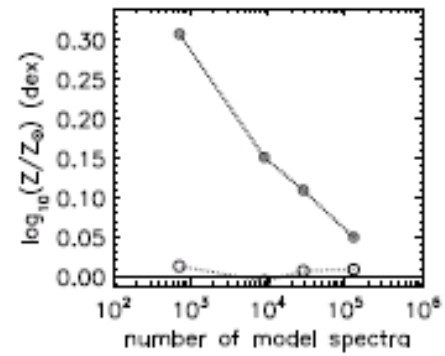
where zi are the neighboring parameter points

Mean Square Error vs. Parameter Grid Resolution

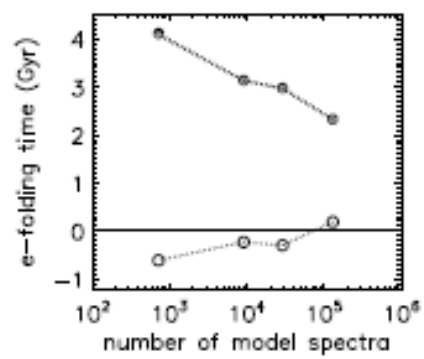
- Mean square error of the parameter estimates decreases with grid resolution.
- Improvement is simultaneous for all parameters.



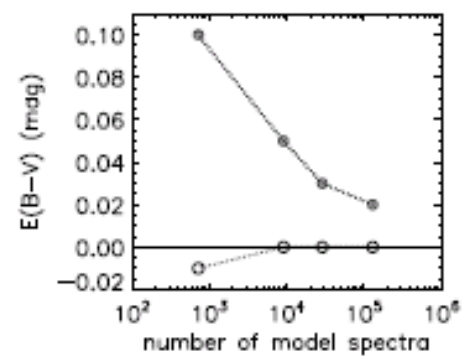
Age of oldest stars



Stellar Metallicity



e-folding time of star formation

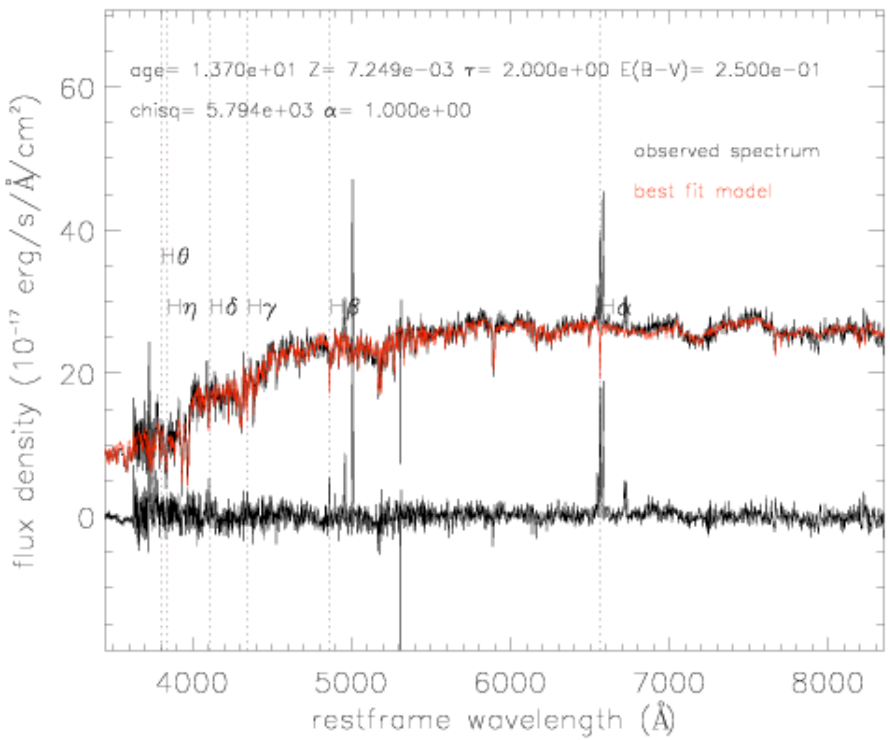


Dust reddening

N-D Parameter Estimation on Spectra

4D stellar population model

- On-the-fly generation of 300,000 model spectra
- Parameter uncertainties are estimated on object-to-object basis
- 10 minutes (SDSS spectral resolution 69km/s)



Age of the oldest stars = 13 Gyr

Metallicity by mass = 0.36 solar

e-folding time of star formation = 2 Gyr

E(B-V) = 0.25 mag

Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)

- 4m segmented telescope, 5° FOV (the Moon spans 0.5 °)
- 4,000 fiber spectra into 16 spectrographs
- 10 million fiber spectra, 10x more than SDSS

Spectral resolutions:

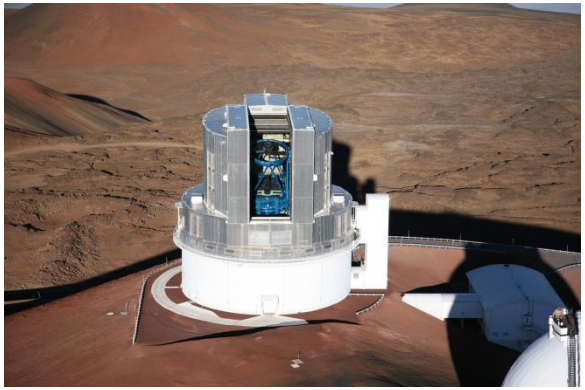
medium-low
R = 1000 - 2000

medium
R = 5000 – 10000

Xinglong Station,
180 km north of Beijing

Subaru Prime Focus Spectrograph (PFS; 2017)

- High redshift version of SDSS
- 2,400 fiber array, 1.3° FOV
- 200,000 galaxy spectra (1.4 < z < 2.2)
- 140,000 Lyα emitters (2 < z < 7)
- 50,000 QSOs (3 < z < 7)

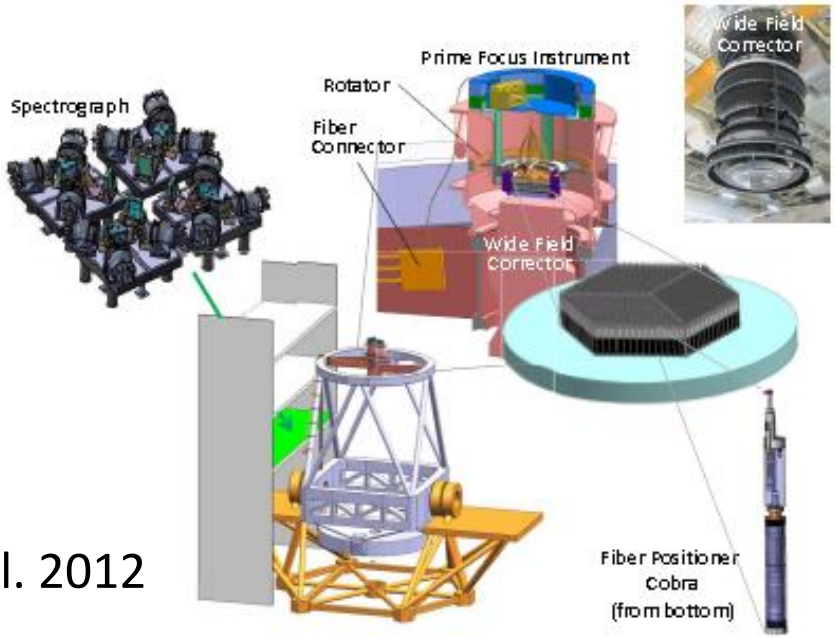


Spectral resolutions:

3800-6700 Å
R = 1900

6500-10000 Å
R = 2400

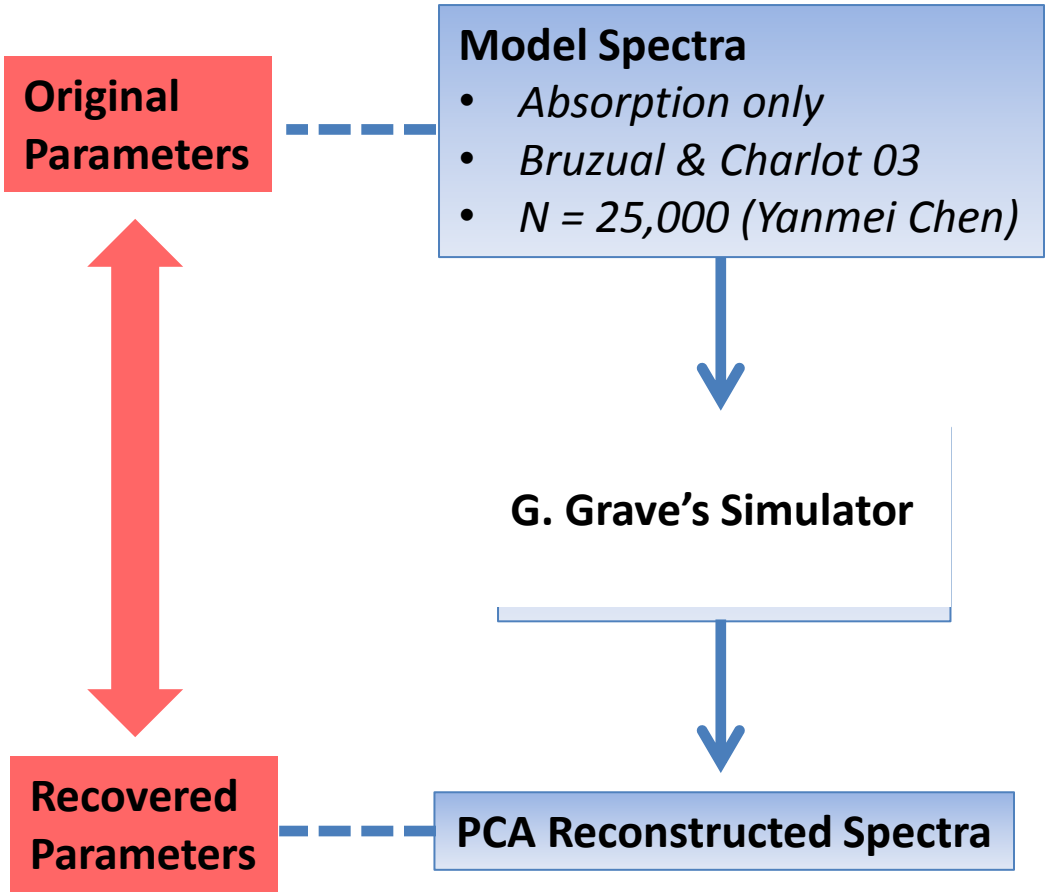
9700-13000 Å
R = 3500



Ellis et al. 2012

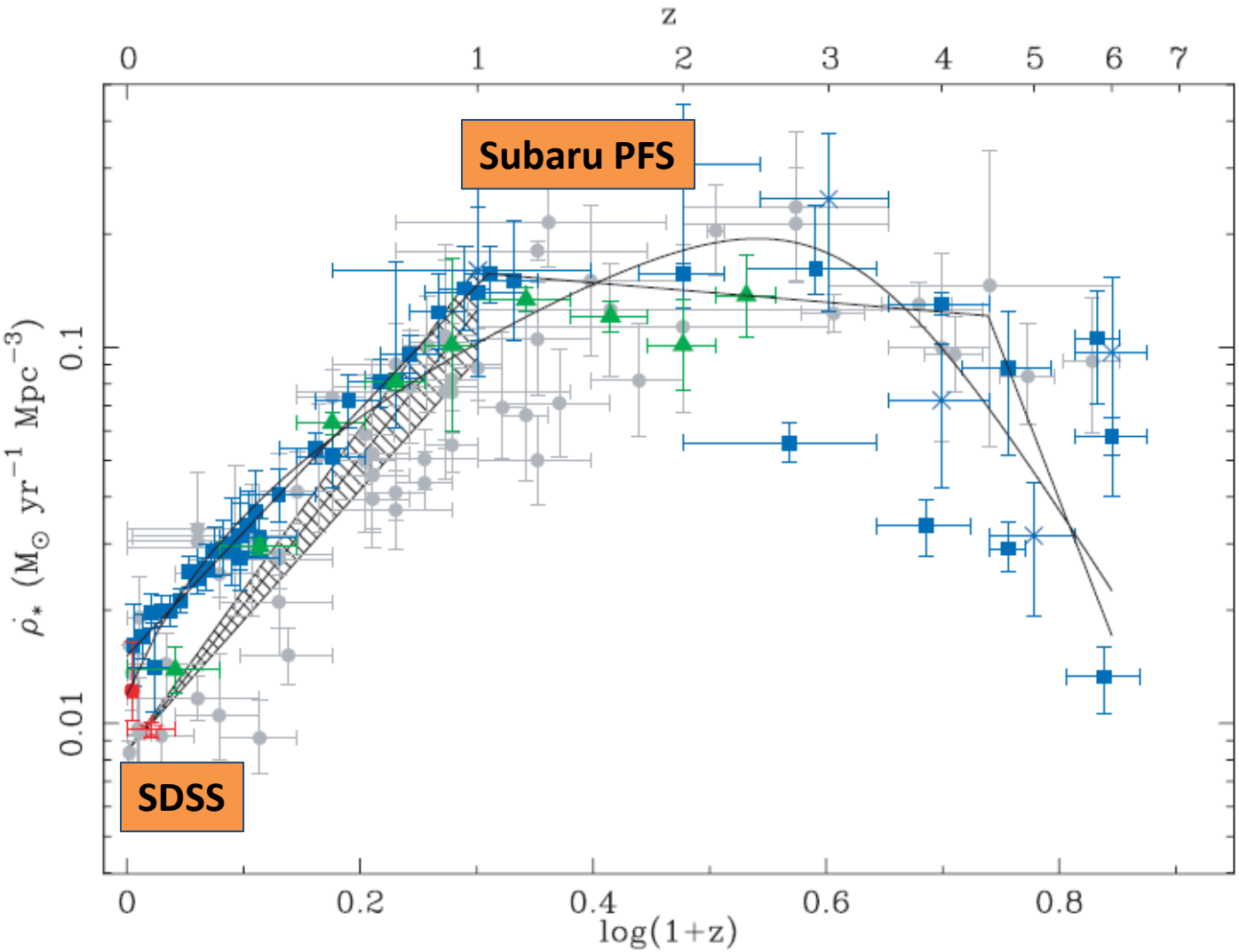
Galaxy Parameters from Prime Focus Spectrograph (PFS) Spectra

- 25,000 model spectra simulation
- Realistic star formation history
- 5 model parameters
- 18 derived parameters
- Focus on $D_n(4000)$



Yip & Heckman, et al.

Cosmic Evolution of Dust and SFR using Inclination Method



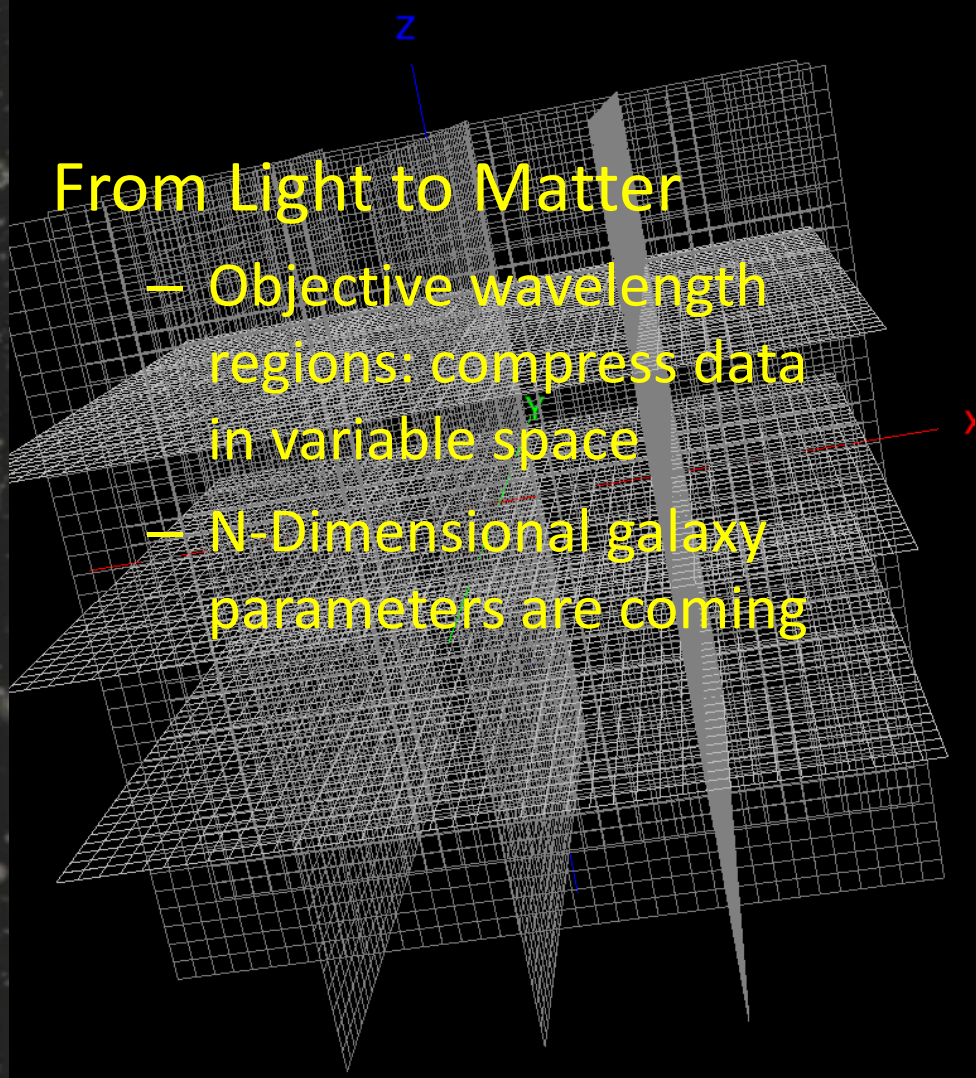
Hopkins 2006

Measure Dust & Light in Galaxies

- Radial Dependency: Galaxy Evolution
- Inclination Dependency: True Color/Luminosity of Galaxies
- 3D Spectra marry both

From Light to Matter

- Objective wavelength regions: compress data in variable space
- N-Dimensional galaxy parameters are coming



SDSS III