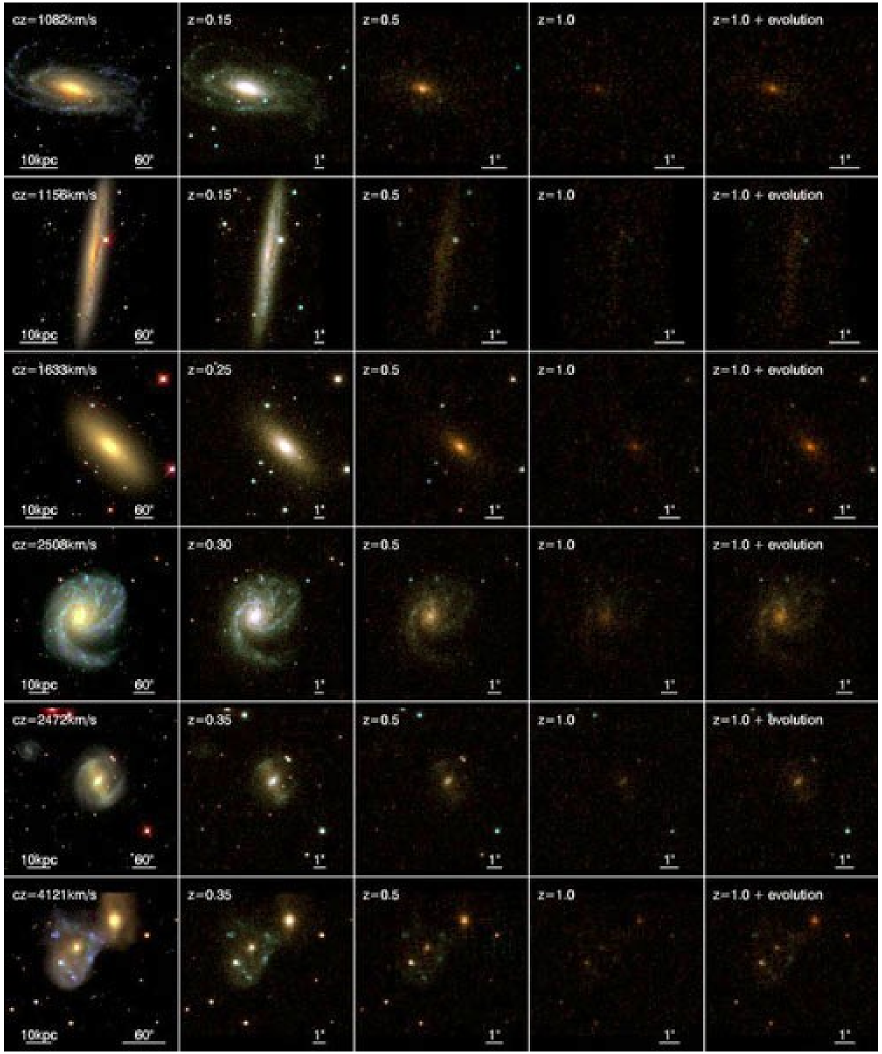
2022 International Workshop in Applied Statistics and Data Science (Cartagena – Colombia) June 29 – July 1, 2022

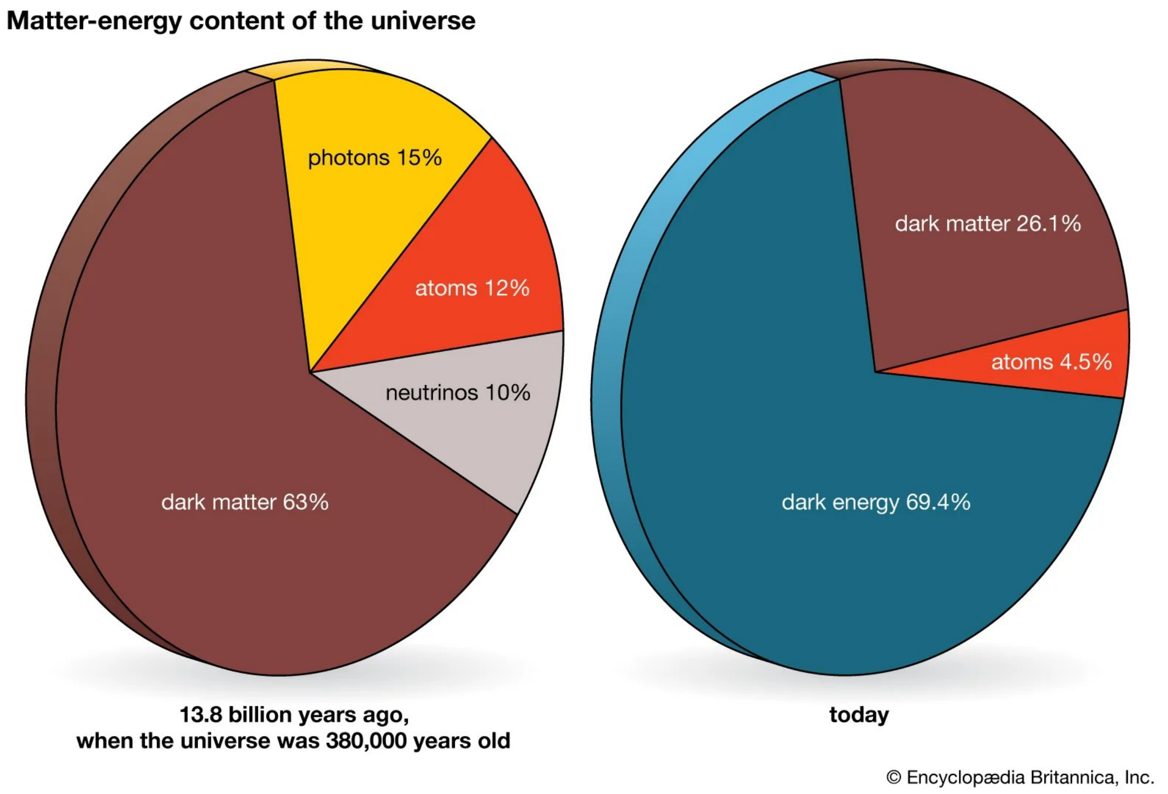
Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS



Machine Learning for Astronomy and Astrophysics

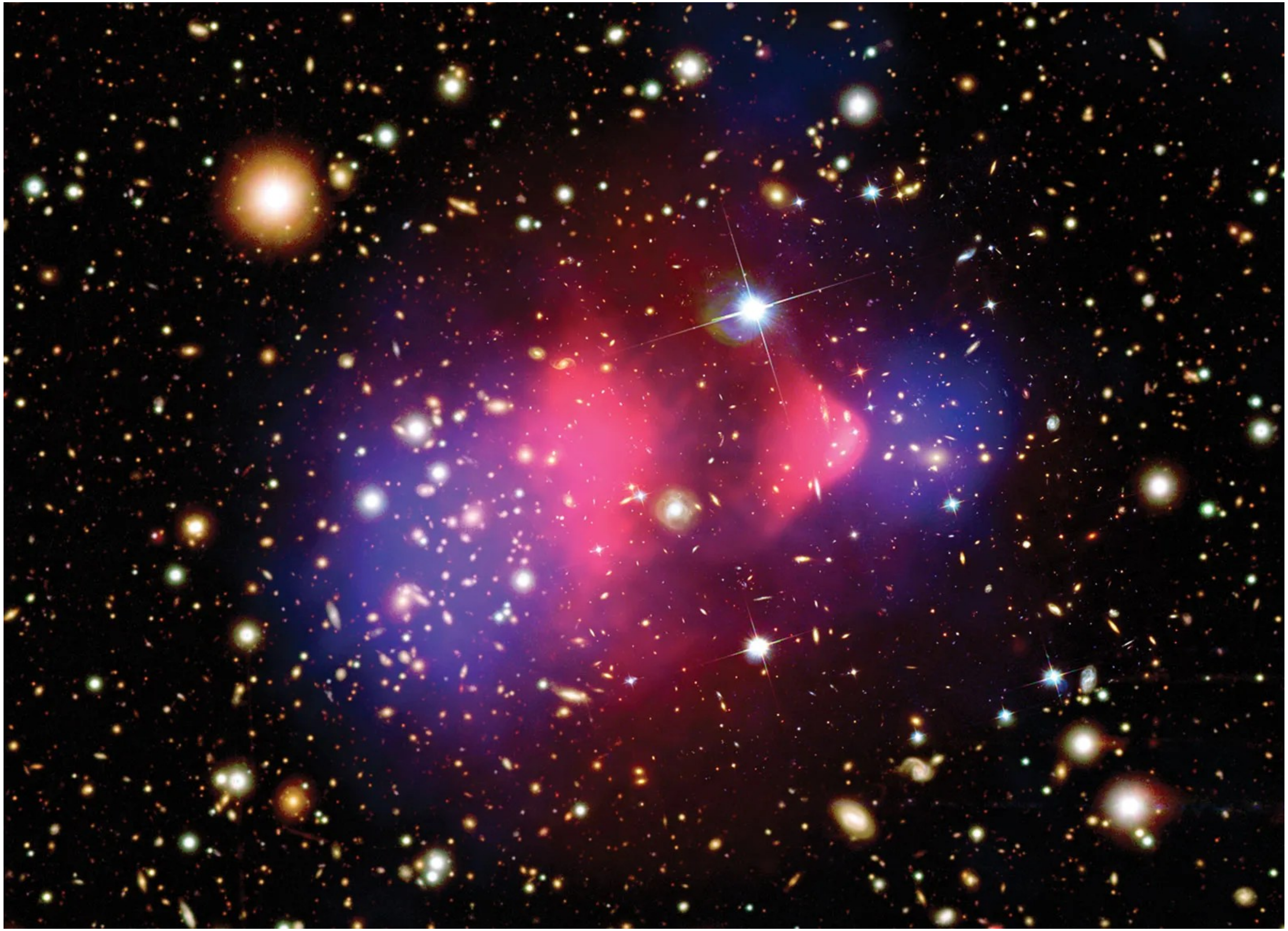
D. Sierra Porta
Faculty of Basic Sciences
Universidad Tecnológica de Bolivar
dporta@utb.edu.co

Dark matter, a component of the universe whose presence is discerned from its gravitational attraction rather than its luminosity. Dark matter makes up 30.1 percent of the matter-energy composition of the universe; the rest is **dark energy** (69.4 percent) and “ordinary” **visible matter** (0.5 percent).



***matter-energy** content of the universeMatter-energy content of the universe.*

Encyclopædia Britannica, Inc.



galaxy cluster 1E0657-56

Composite image showing the galaxy cluster 1E0657-56, the Bullet cluster.

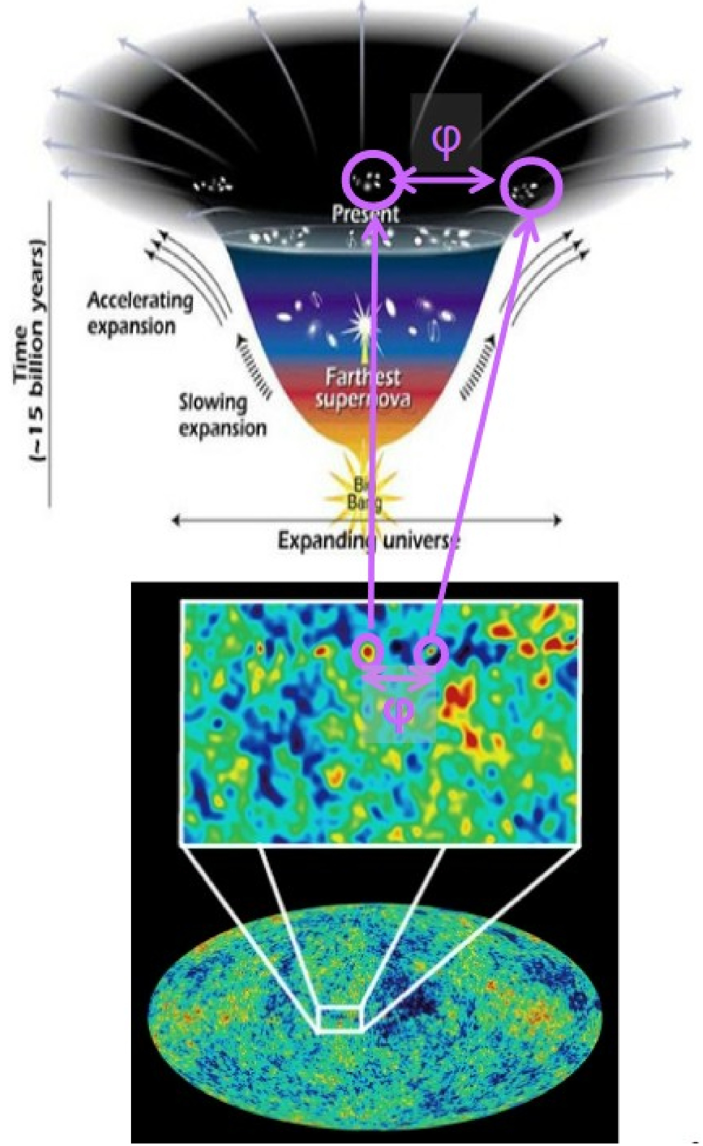
X-ray: NASA/CXC/CfA/M.Markevitch Optical: NASA/STScI; Magellan/U.Arizona/D.Clowe Lensing Map: NASA/STScI; ESO WFI; Magellan/U.Arizona/D.Clowe

Expanding Universe

- Dark energy is even more mysterious, and its discovery in the 1990s was a complete shock to scientists. Previously, physicists had assumed that the attractive force of gravity would slow down the expansion of the universe over time. But when two independent teams tried to measure the rate of deceleration, they found that the expansion was actually speeding up. One scientist likened the finding to throwing a set of keys up in the air expecting them to fall back down-only to see them fly straight up toward the ceiling.
- Scientists now think that the accelerated expansion of the universe is driven by a kind of repulsive force generated by quantum fluctuations in otherwise "empty" space. What's more, the force seems to be growing stronger as the universe expands. For lack of a better name, scientists call this mysterious force dark energy.

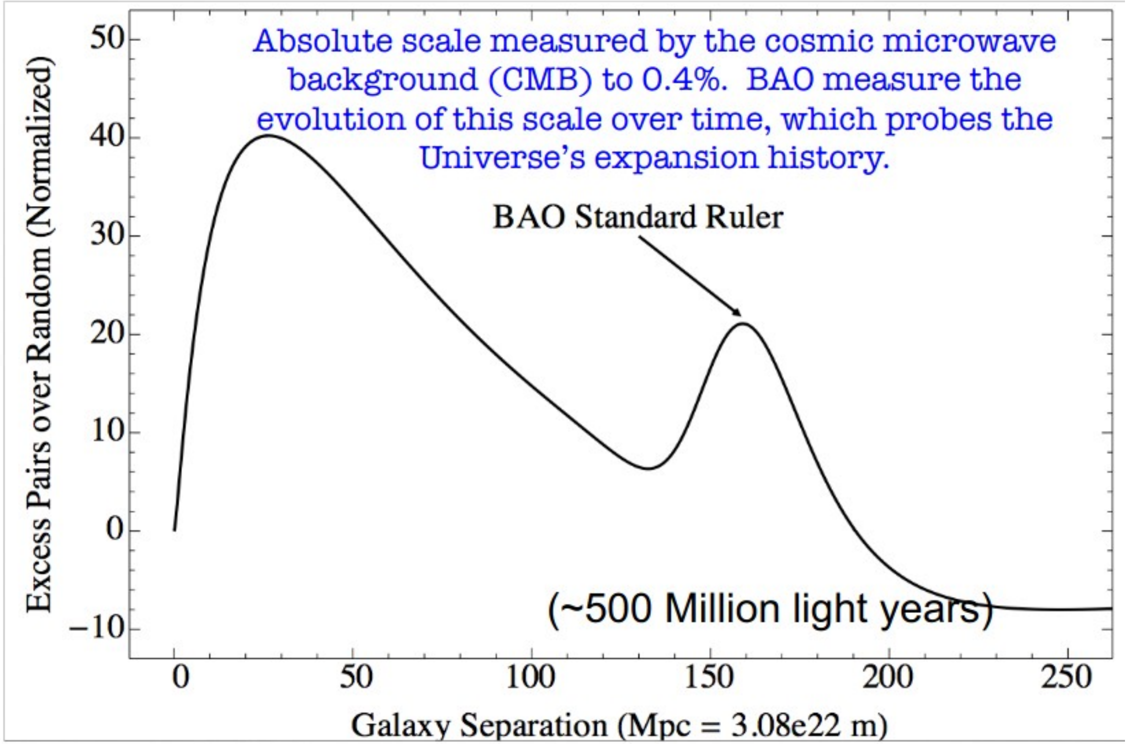
Mystery of Dark Energy

- What is the source of the accelerated expansion of the universe
  - Cosmological constant (w= -1)?
  - New long-range repulsive force?
  - Modification of gravity?
  - Other ideas?
- Experimentally we study two major aspects of our universe
  - The expansion history of the universe
  - The growth of structures (such as galaxies and galaxy clusters)
- DESI focuses primarily on measuring the expansion history using a standard ruler:
  - Baryon Acoustic Oscillations (BAO)



Baryon Acoustic Oscillations

Sound waves in the early Universe (~ 400,000 yrs after the big bang) produce a peak in the clustering of matter that shows up in the distribution of galaxies we see today ~ 13 billion years later. Standard Ruler!



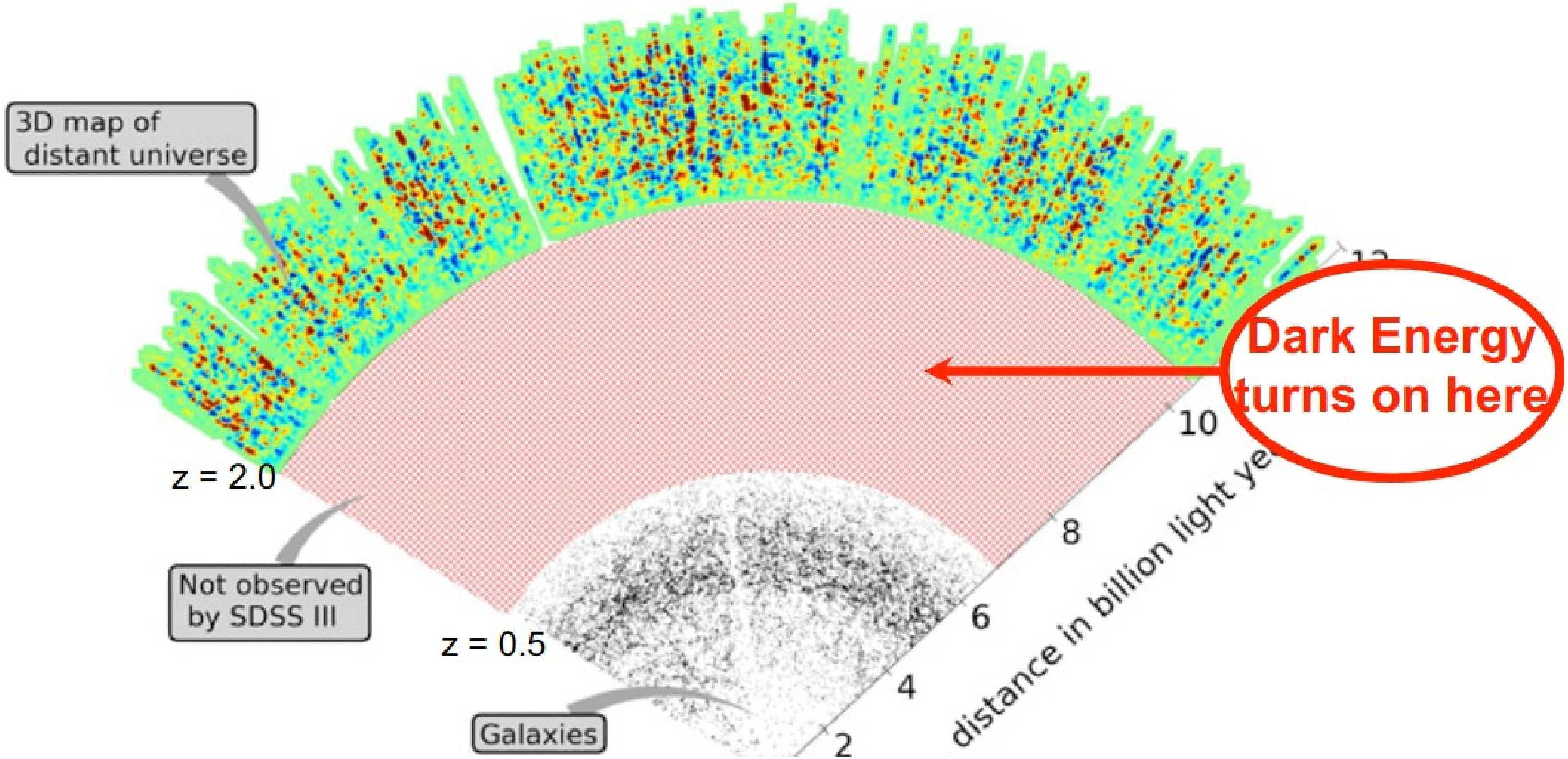
BAO is fundamentally a measurement of the separations between galaxies, not the properties of the galaxies themselves:

- low systematic
- uncertainties

Clustering of galaxies tells us about gravity and the neutrino mass

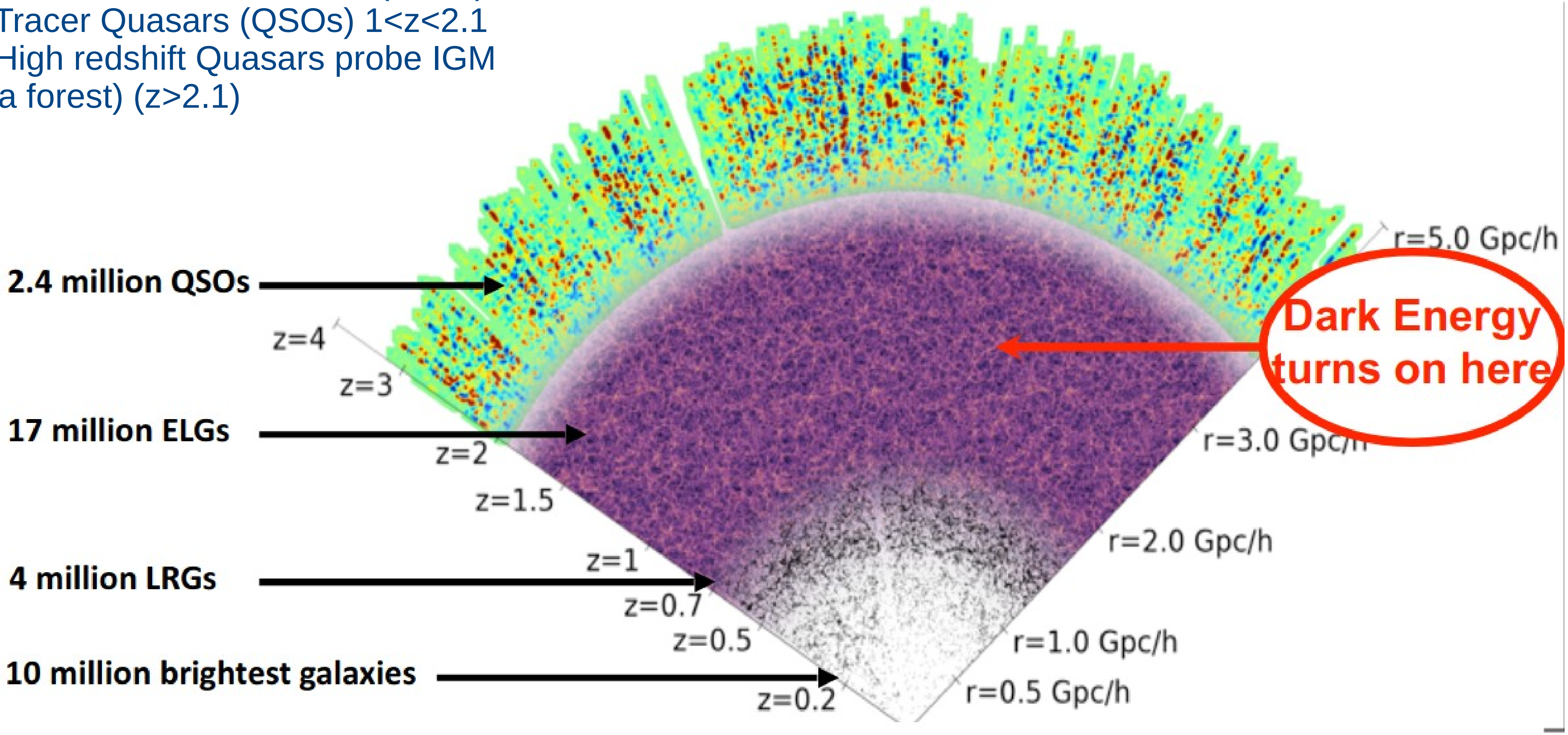
SDSS, SDSS II and SDSS III (BOSS) have shown the power of spectroscopic surveys but had limited redshift range (<0.7 , >2)

SDSS ~2h^{-3}Gpc^3 ⇒ BOSS ~6h^{-3}Gpc^3 ⇒ DESI ~ 50 h^{-3}Gpc^3



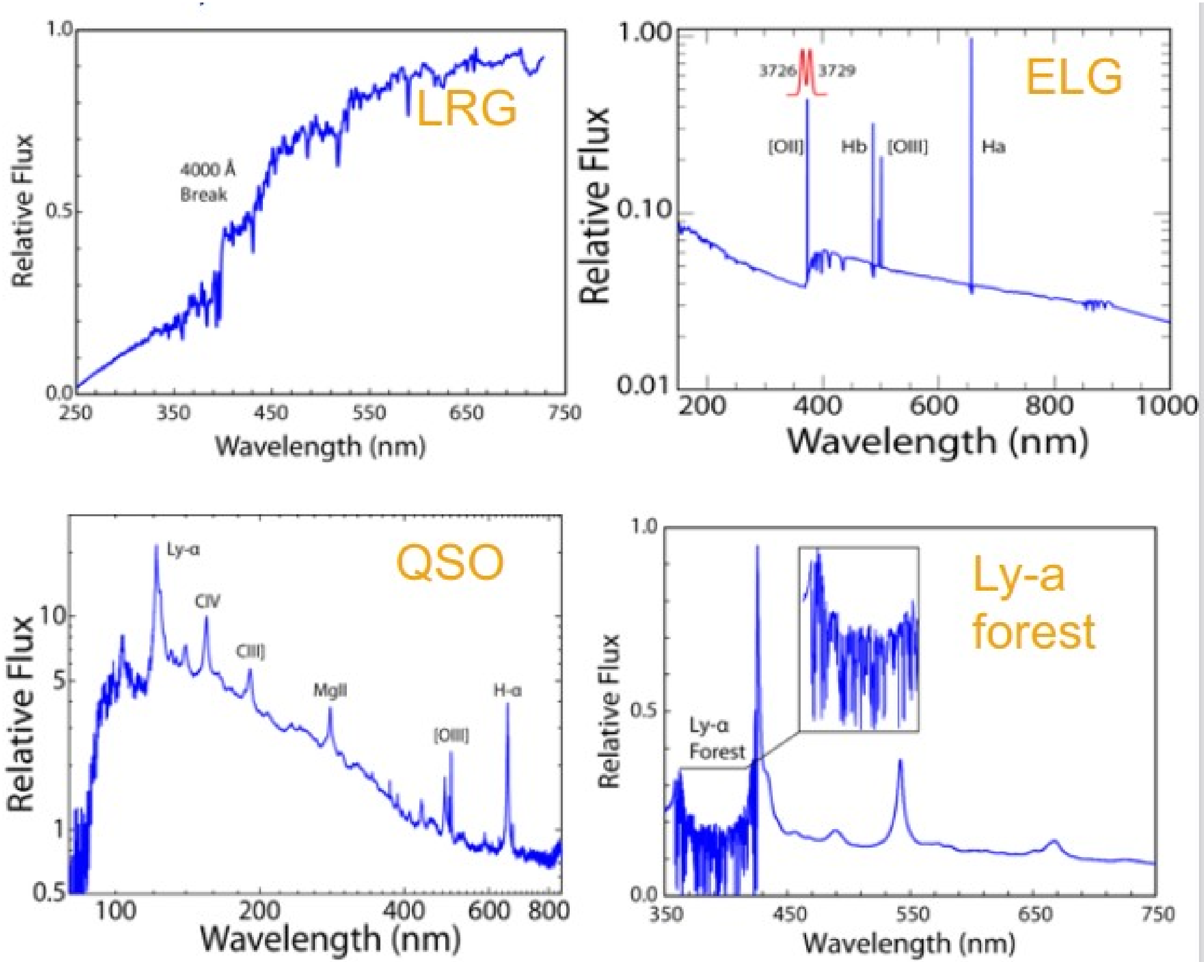
DESI Survey: ~ 34M Galaxies, 14K deg2

- 10 million Bright Galaxies 0.0<z<0.4
- 4 million Luminous Red Galaxies (LRGs) 0.4<z<1
- 17.1 million Emission Line Galaxies (ELGs) 0.6<z<1.6
- 1.7 million Tracer Quasars (QSOs) 1<z<2.1
- 0.7 million High redshift Quasars probe IGM (Lyman-alpha forest) (z>2.1)



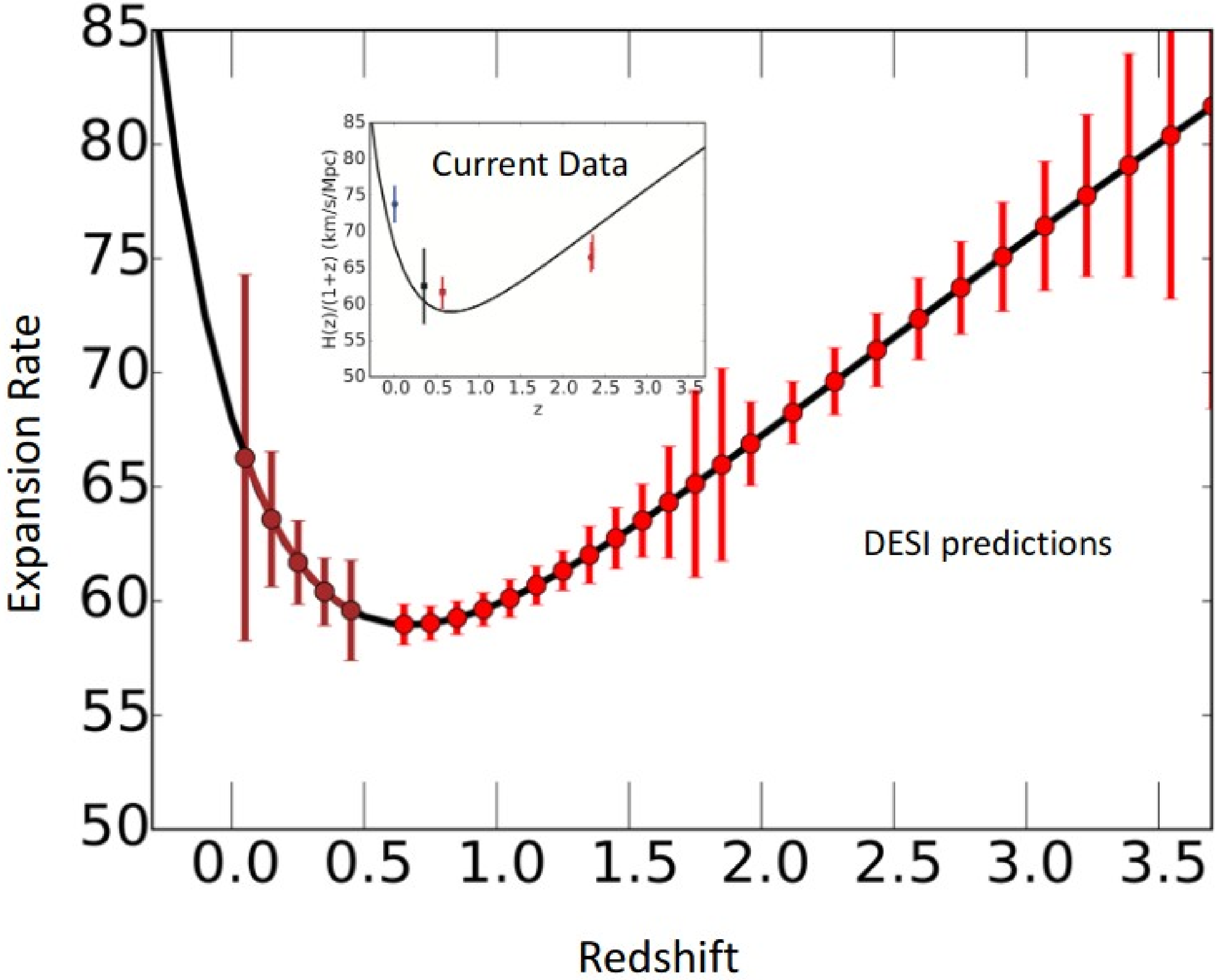
DESI Science Requirements

- Identify spectral features for each type of target
  - Bandpass from 360 – 980 nm, Δz/(1+z) ~0.0005, λ/Δλ resolution ~ 2000-4000
  - 3 arm spectrographs to cover full range
- 5 year survey (~ 100% of the time)
- 14000 deg2 survey size
- 5000 spectra per exposure
- 34 Million targets
- ~20 min exposure times.
- ~10,000 total number of exposures
- Goal: <0.3% uncertainty in the distance scale at z ~ 1 and <1% for z < ~3



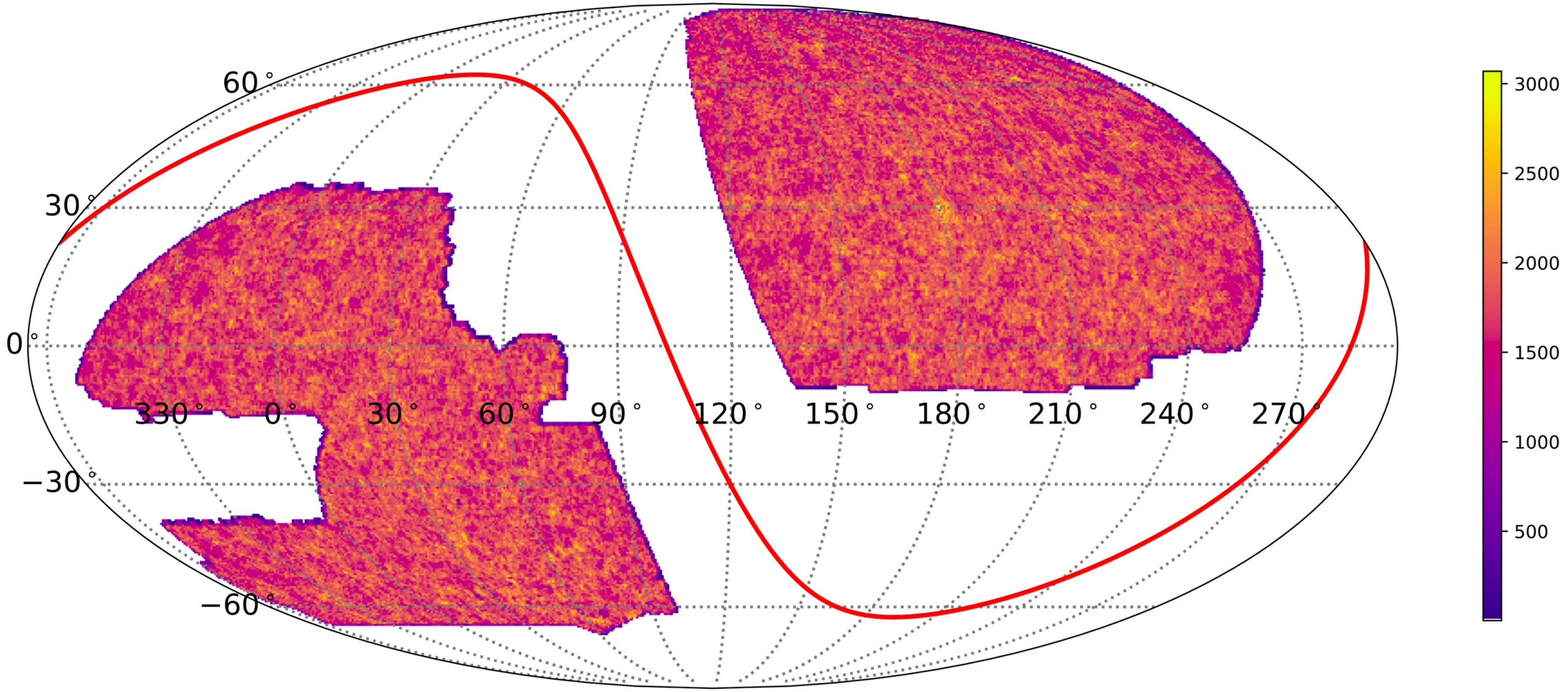
DESI Hubble Diagram:

- Estimated Errors after 5 year survey

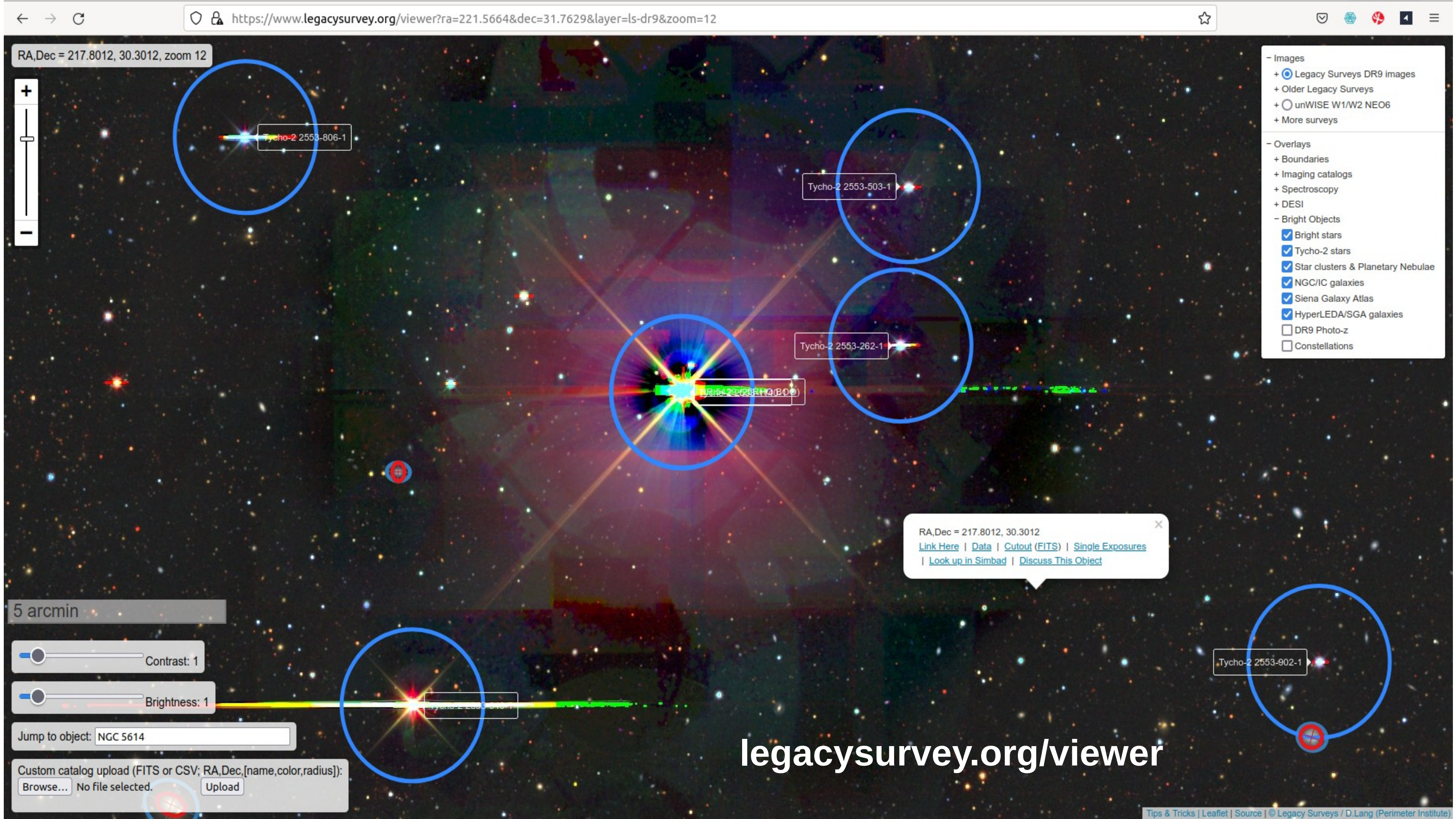


DESI Survey: Targets

• DESI collaboration Imaging surveys (BASS, MzLS and DECALs) will provide targets for DESI.



DECALS uses DECam to survey the southern 9000 deg2 of the DESI area Northern area is being covered by a Mayall z and BASS Regular public data releases



legacysurvey.org/viewer

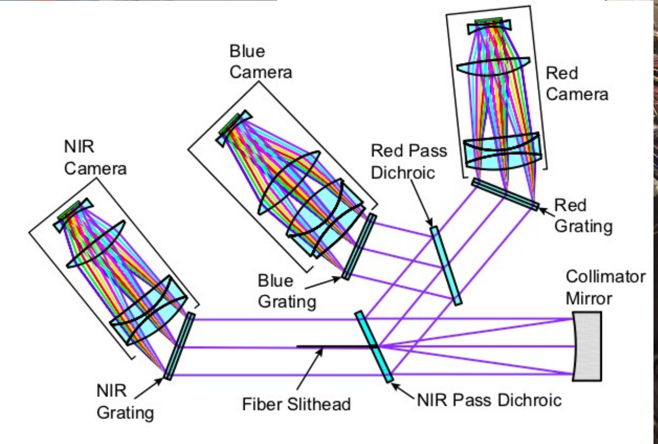
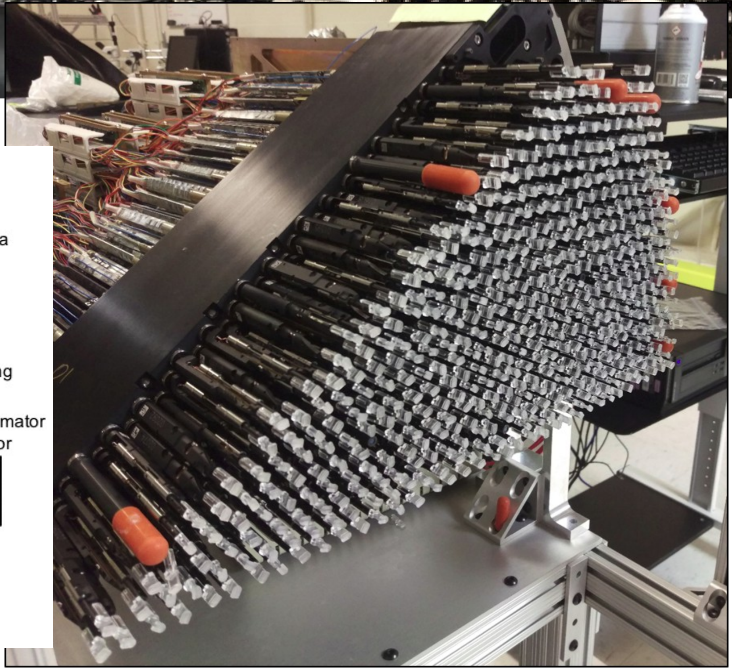
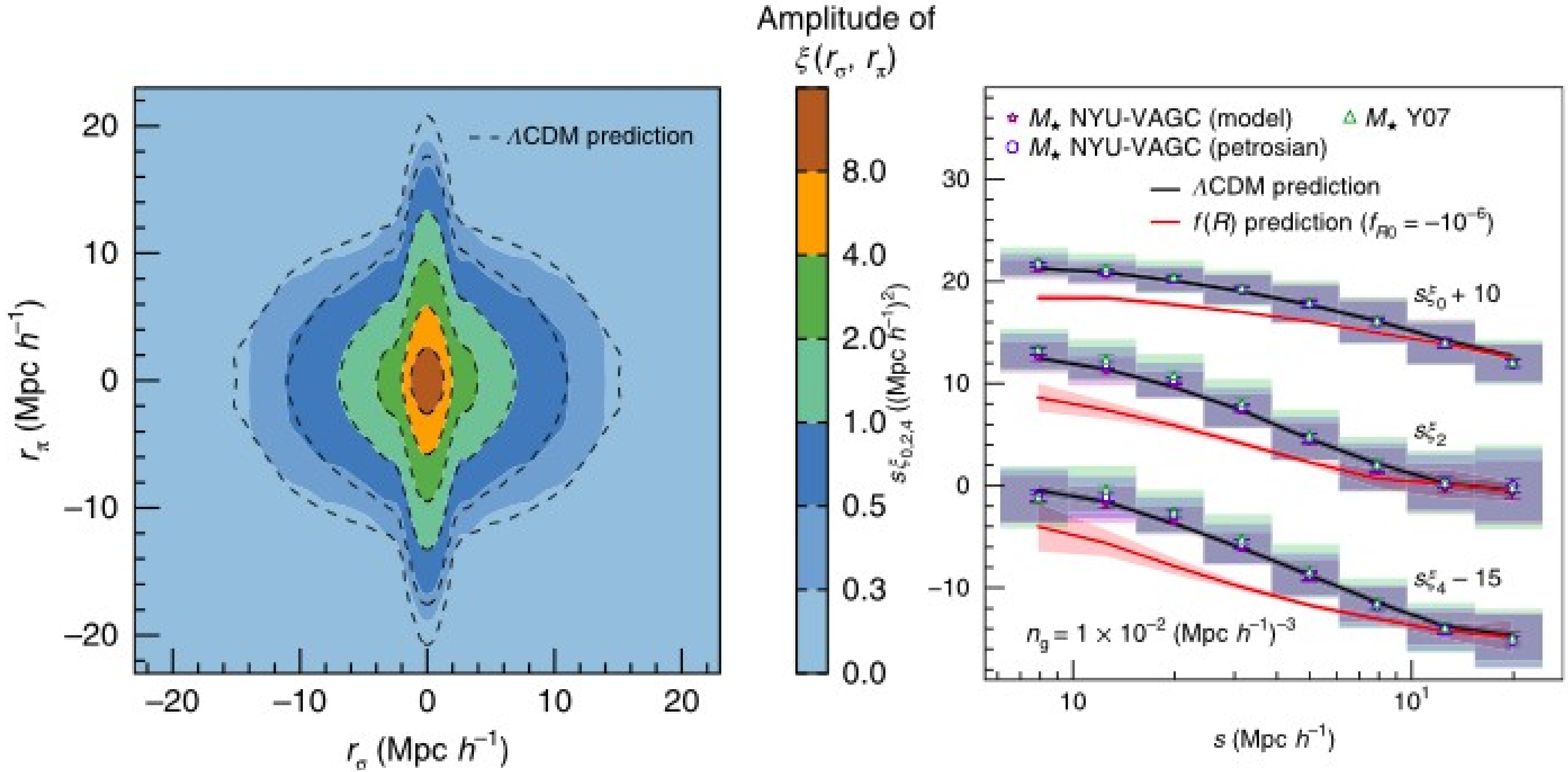


Figure 5.1: The spectrograph schematic.





ΛCDM prediction
Amplitude of $\xi(r_\sigma, r_\pi)$
8.0
4.0
2.0
1.0
0.5
0.3
0.0
r_π (Mpc h^{-1})
r_σ (Mpc h^{-1})
$M_\star$ NYU-VAGC (model)
$M_\star$ Y07
$M_\star$ NYU-VAGC (petrosian)
ΛCDM prediction
$f(R)$ prediction ($f_{R0} = -10^{-6}$)
$s\xi_0 + 10$
$s\xi_2$
$s\xi_4 - 15$
$n_g = 1 \times 10^{-2}$ (Mpc h^{-1})$^{-3}$
$s\xi_{0,2,4}$ ((Mpc h^{-1})^2)
s (Mpc h^{-1})

Galaxy Zoo: reproducing galaxy morphologies via machine learning* FREE

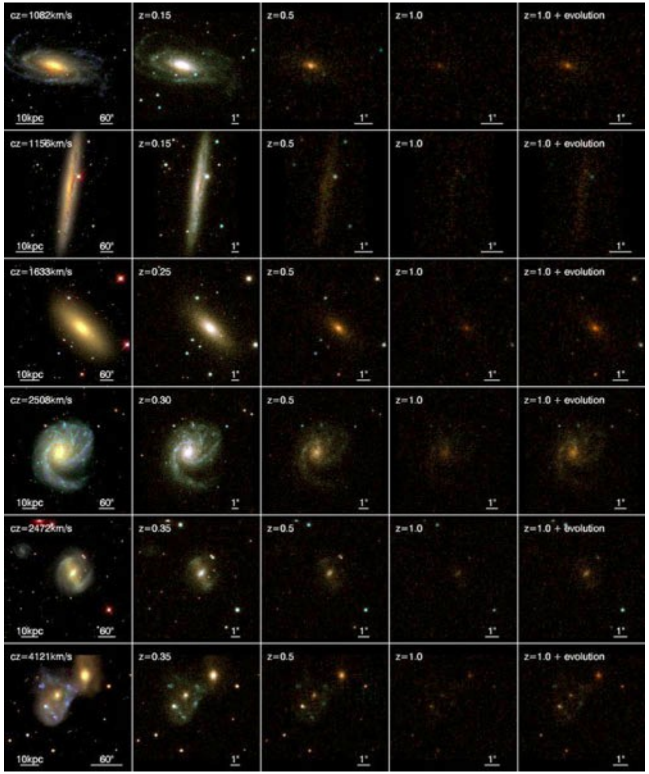
Manda Banerji, Ofer Lahav, Chris J. Lintott, Filipe B. Abdalla, Kevin Schawinski, Steven P. Bamford, Dan Andreescu, Phil Murray, M. Jordan Raddick, Anze Slosar ... Show more

Monthly Notices of the Royal Astronomical Society, Volume 406, Issue 1, July 2010, Pages 342–353, https://doi.org/10.1111/j.1365-2966.2010.16713.x

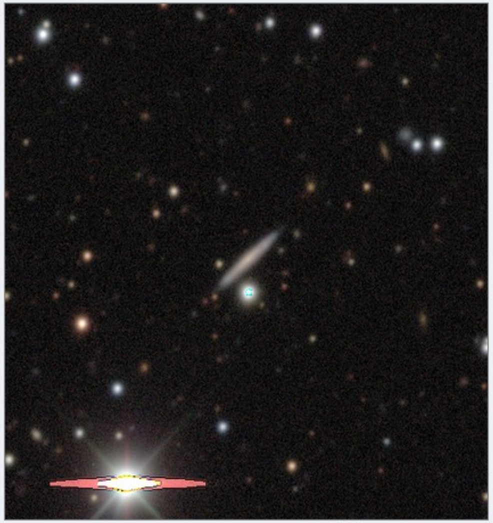
Published: 21 July 2010 **Article history ▾**

We present morphological classifications obtained using machine learning for objects in SDSS DR6 that have been classified by Galaxy Zoo into three classes, namely early types, spirals and point sources/artifacts. An artificial neural network is trained on a subset of objects classified by the human eye and we test whether the machine learning algorithm can reproduce the human classifications for the rest of the sample. We find that the success of the neural network in matching the human classifications depends crucially on the set of input parameters chosen for the machine-learning algorithm. The colours and parameters associated with profile-fitting are reasonable in separating the objects into three classes. However, these results are considerably improved when adding adaptive shape parameters as well as concentration and texture. The adaptive moments, concentration and texture parameters alone cannot distinguish between early type galaxies and the point sources/artifacts. Using a set of twelve parameters, the neural network is able to reproduce the human classifications to better than 90% for all three morphological classes. We find that using a training set that is incomplete in magnitude does not degrade our results given our particular choice of the input parameters to the network. We conclude that it is promising to use machine-learning algorithms to perform morphological classification for the next generation of wide-field imaging surveys and that the Galaxy Zoo catalogue provides an invaluable training set for such purposes.

Key words:
Galaxy - morphologies, Methods - data analysis.



Galaxy Zoo



You should sign in!

TASK | TUTORIAL

Is the galaxy simply smooth and rounded, with no sign of a disk?

- Smooth
- Features or Disk
- Star or Artifact

NEED SOME HELP WITH THIS TASK?

Done & Talk | Done

FIELD GUIDE

SWITCH TO DARK THEME

Galaxy Zoo is a web-based project that aimed to obtain morphological classifications for roughly a million objects in the Sloan Digital Sky Survey by harnessing the power of the internet and recruiting members of the public to perform these classifications by eye. The first part of this project is now complete and the morphological classifications subsequently obtained have been described in detail in Lintott et al. (2008) where these classifications have also been shown to be credible based on comparison with classifications by professional astronomers.

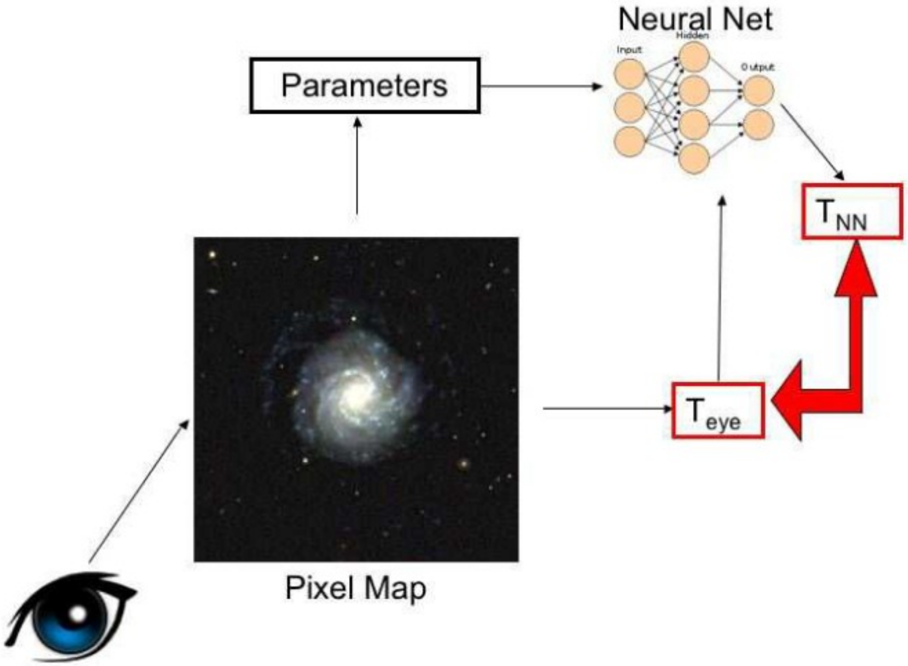
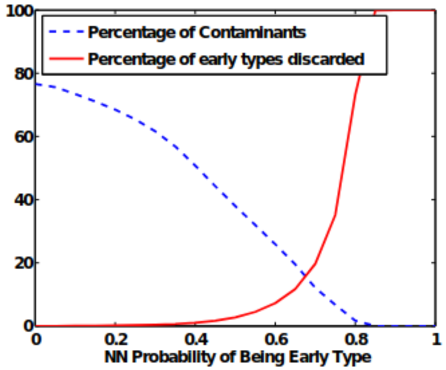
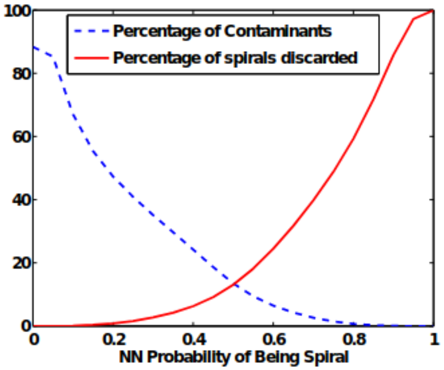
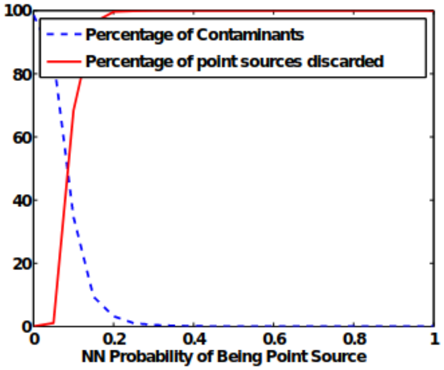


Figure 1. Cartoon schematic of how both the human eye as well as machine learning algorithms such as artificial neural networks perform morphological classification and determine parameters such as those listed in Table 1 and 2 from the galaxy images.







$$E = \sum_k (T_{NN}(w_{ij}, p_k) - T_{eye,k})^2$$

Name	Description
`petroR90_i/petroR50_i`	Concentration
`mRrCc_i`	Adaptive (+) shape measure
`aE_i`	Adaptive Ellipticity
`mCr4_i`	Adaptive fourth moment
`texture_i`	Texture parameter

Name	Description
`dered_g-dered_r`	(g-r) colour
`dered_r-dered_i`	(r-i) colour
`deVAB_i`	de Vaucouleurs fit axis ratio
`expAB_i`	Exponential fit axis ratio
`lnLexp_i`	Exponential disk fit log likelihood
`lnLdeV_i`	de Vaucouleurs fit log likelihood
`lnLstar_i`	Star log likelihood

		GALAXY ZOO		
		Early Type	Spiral	Point Source/Artifact
A	EARLY TYPE	84%	0.5%	85%
N	SPIRAL	1%	87%	0.8%
N	POINT SOURCE/ARTIFACT	32%	6.5%	32%

		GALAXY ZOO		
		Early Type	Spiral	Point Source/Artifact
A	EARLY TYPE	87%	0.3%	0.3%
N	SPIRAL	0.6%	86%	2.2%
N	POINT SOURCE/ARTIFACT	0.7%	0.5%	95%

		GALAXY ZOO		
		Early Type	Spiral	Point Source/Artifact
A	EARLY TYPE	91%	0.05%	0.8%
N	SPIRAL	0.08%	92%	0.1%
N	POINT SOURCE/ARTIFACT	2%	0.1%	95%

Summary of results for entire sample when using input parameters specified in Table 1 and Table 2 and only bright galaxies with r < 17 to train the network.

…we still have much to do…

Thanks

D. Sierra Porta
Faculty of Basic Sciences
Universidad Tecnológica de Bolivar
dporta@utb.edu.co