Optical Galaxy Cluster Detection at High Redshift

Collaborators

Michigan: Tim Mckay, Gus Evrard, Dave Gerdes, Seth Siegel, Aarons Fenyes, Kimmy Wu, Brian Nord, Anbo Chen, Carlos Cunha
Chicago: Ben Koester, Matt Becker
Fermilab: Jim Annis, Huan Lin
JPL: Dave Johnston
NYU: Erin Sheldon
OSU: Eduardo Rozo
Stanford: Risa Wechsler, Micheal Busha
UCSB: Eli Rykoff
UCSD: Sarah Hansen

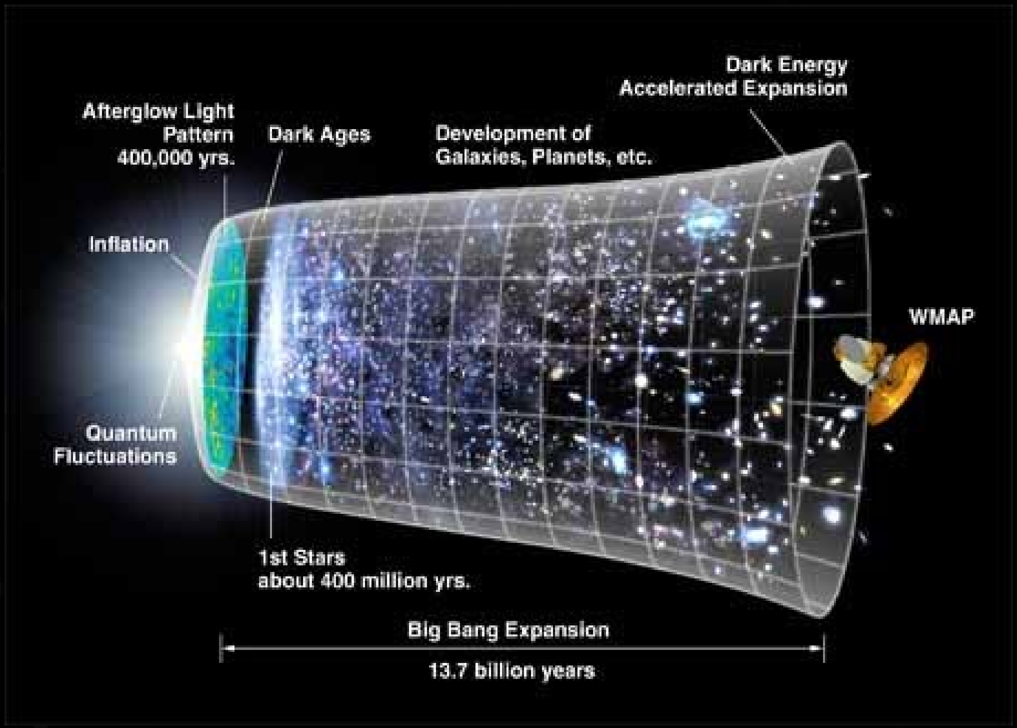
1/26/2009 @ Fermi National Accelerator Lab

Outline

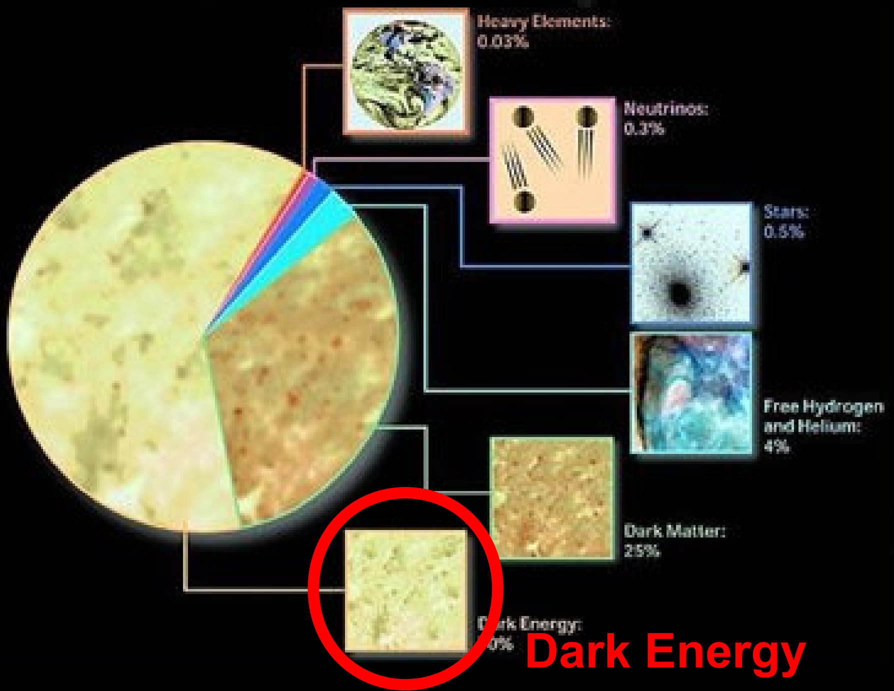
1. Cosmological motivation for cluster detection?
2. Challenge of optical galaxy cluster detection and GMBCG algorithm for high redshift cluster detection
3. Getting ready for Dark Energy Survey (DES)
4. Precision measurements of cluster properties

"Standard" Cosmological Model

History



Contents



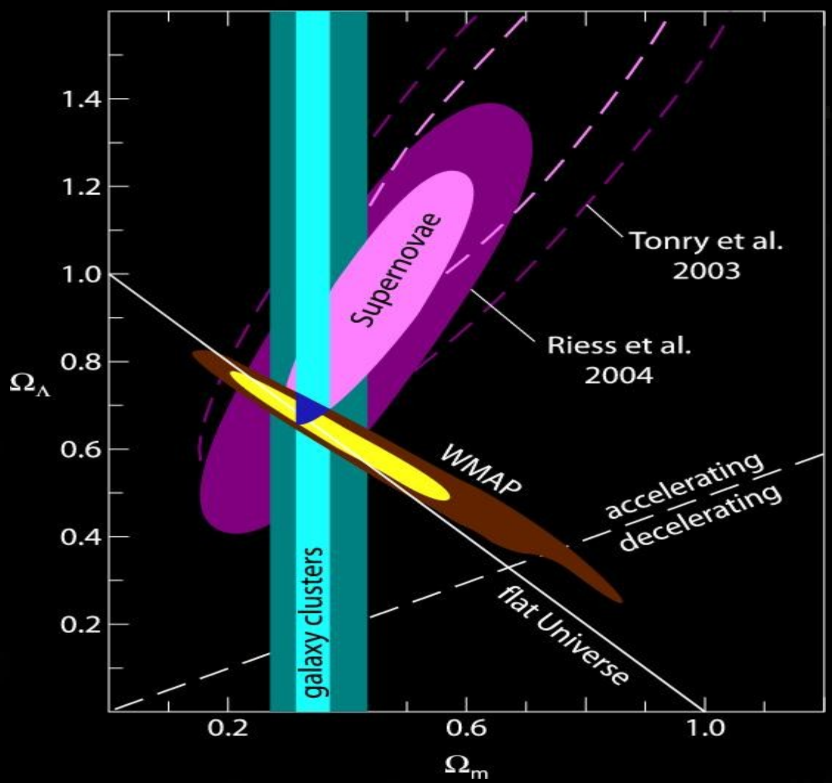
What is the nature of Dark Energy ?

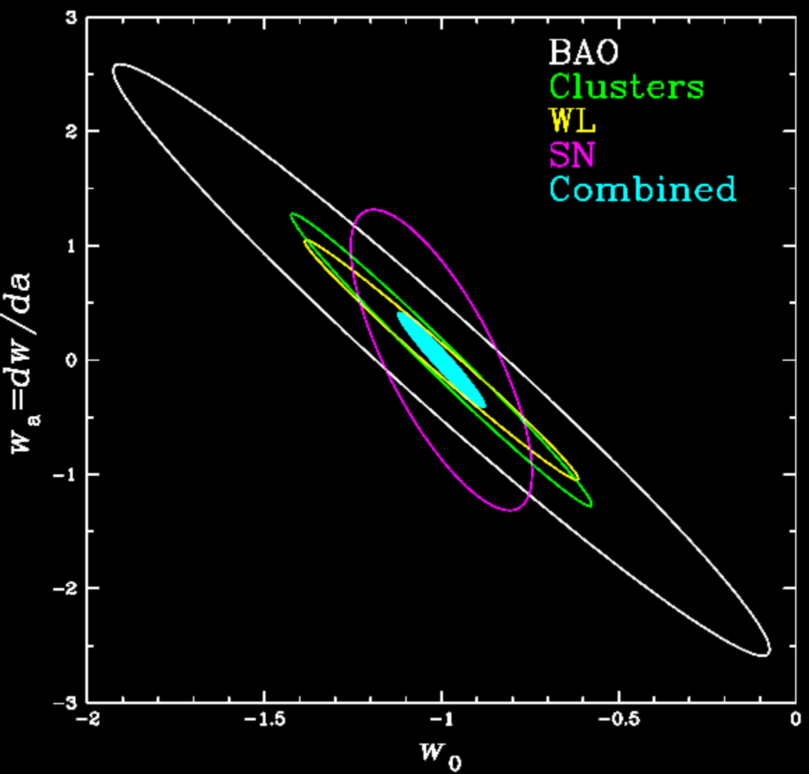
A simple question but expensive to answer!

Dark Energy Parameters

$p = w(a)\ \rho$

$w(a) = w0 + dw/da * (1-a) + o(a^2)$



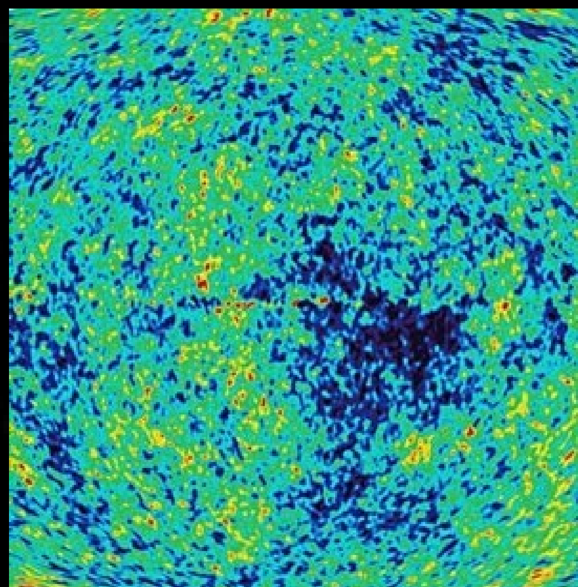


From MPE WebPage

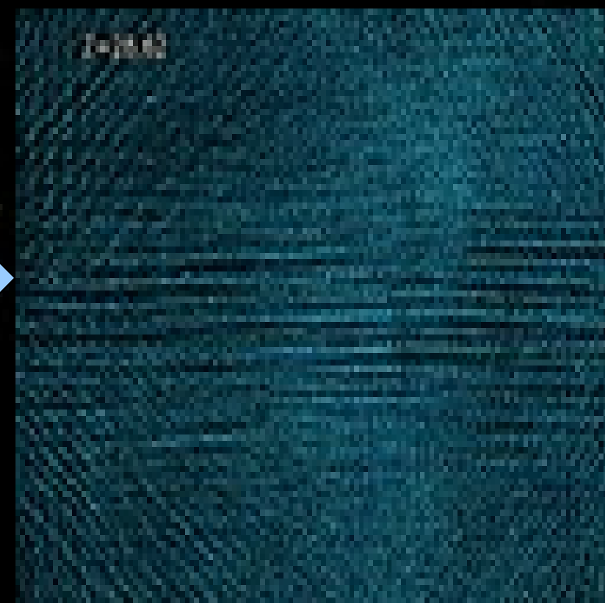
From DES science proposal, Frieman et al

Large Scale Structures as Cosmological Probe

CMB map from WMAP team



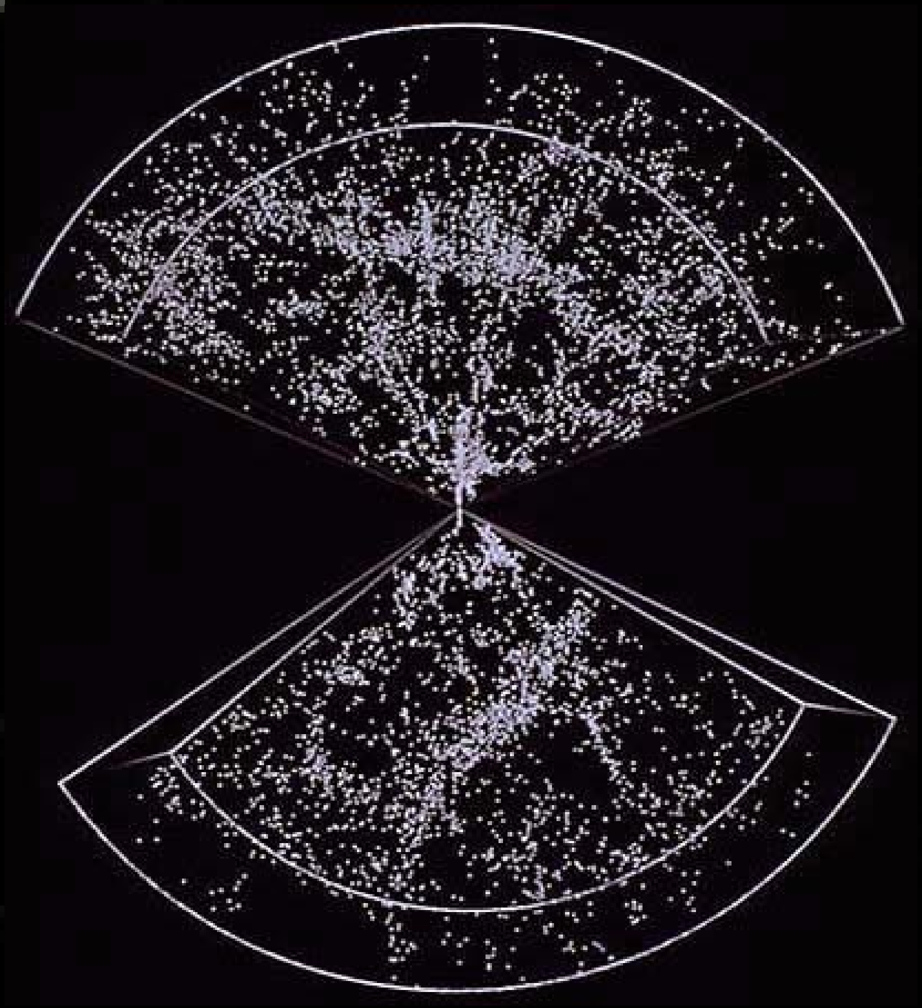
Dark matter simulation
from A. Kravtsov



Given a set of cosmological parameters, theory and simulation can predict very well the abundance and distribution of dark matter halos at given mass scales.

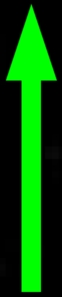
That is to say, if we know the dark matter halo distribution, we can reverse the process to constrain the cosmological parameters.

Characterizing Halo Distribution



How many ?

How Clustering ?



Determined by Cosmology

Theoretical linking I: Number Counts

Number of halos

$$\frac{dN}{d\Omega\, dz} = \frac{c\, d_A^2 (1+z)^2}{H(z)} \times \int_{M_{min}}^{\infty} dM \frac{dn}{dM}$$

Mass limit

Mass function

mass function:

$$\frac{dn}{dM} = -0.315 \frac{\rho_b}{M} \left(\frac{1}{\sigma_M} \frac{d\sigma_M}{dM} \right) \exp\{-[0.61 - \log(D_z \sigma_M)]^{3.8}\}$$

Jenkins et al. 2001

Background density

Variance of fluctuation

Growth function

Theoretical linking II: Clustering

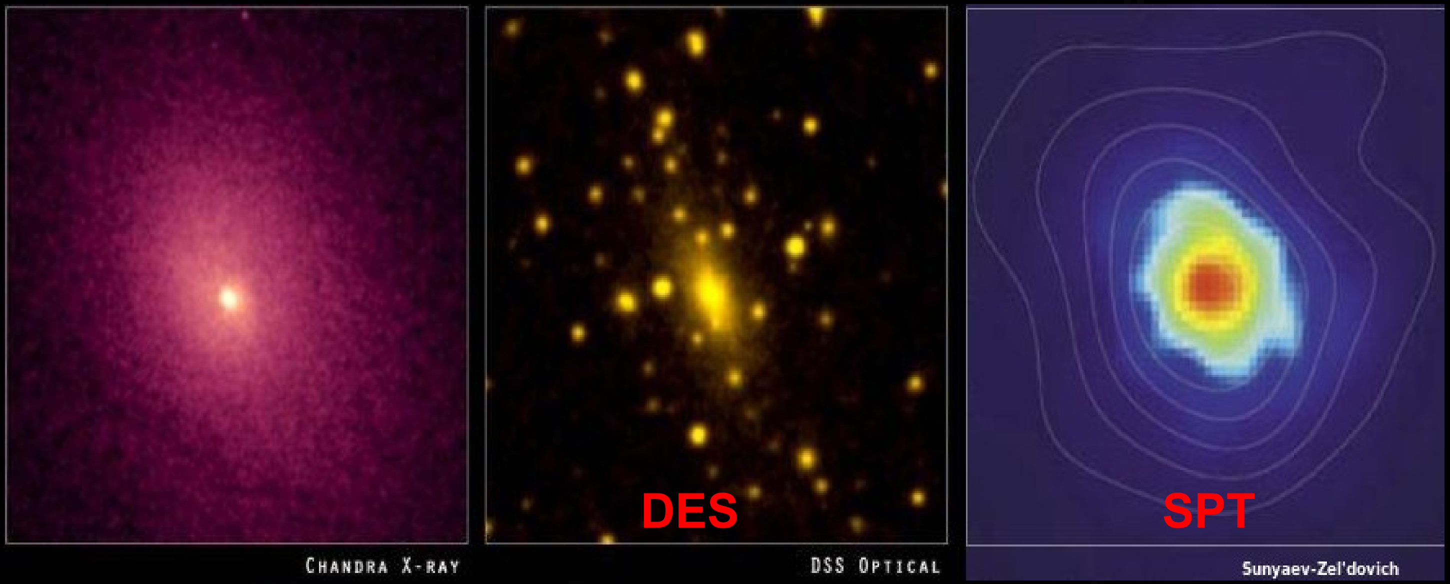
$$P_c(k) = b^2 D(z)^2 \left[1 + \beta^2\right]^2 T^2_m(k) P_{ini}(k)$$

- Halo Power spectrum
- Halo Mass Bias
- Growth function
- Redshift distortion
- Transfer function
- Initial Power spectrum

Redshift distortion: $\beta = \frac{1}{b} \frac{d \ln D(z)}{d \ln a}$ **Scales with growth function**

Tracing Halos from Galaxy Clusters

- Galaxy clusters are good tracers of dark matter halos
- Need to know the masses of clusters



A galaxy cluster catalog is the first step

Building Optical Galaxy Cluster Catalog

Pros and Cons

- Optical data are less expensive to obtain
- Large Volume Available Data (SDSS)
- High Signal/Noise

- Projection Contamination: we could not resolve the position of galaxies along the line of sight to precision we need.

Generic Data Clustering Analysis

- Supervised learning and unsupervised learning

The number of groups is known, assigning every data point into the groups.

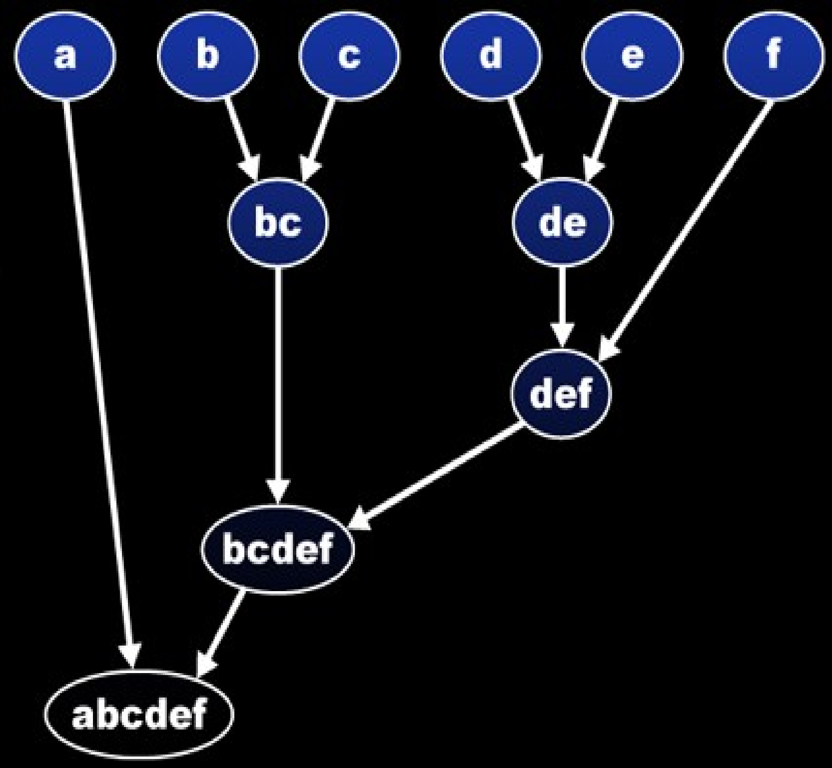
Number of groups is unknown, grouping the data point according their clustering.

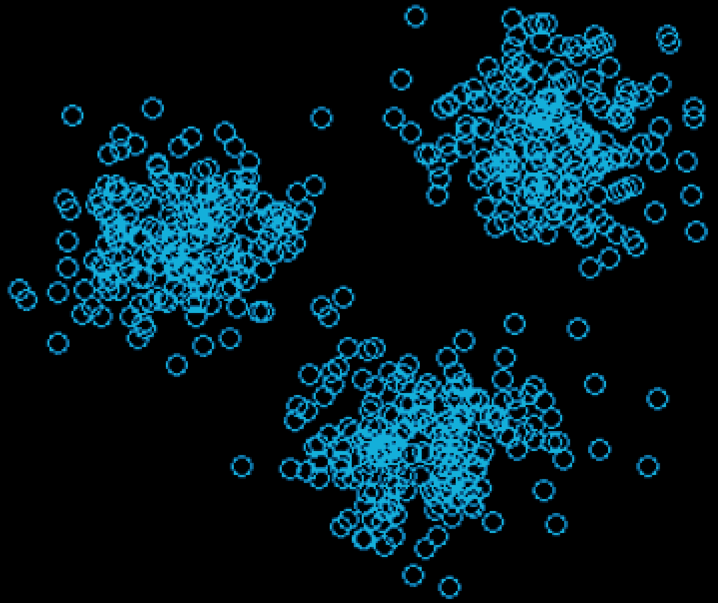
- Type of clustering analysis

  - Partition Method: objects are partitioned into non-overlapping groups and each object belongs to only one group. K-means, mixture models, Kernel Method

  - Hierarchical Method: Objects are partitioned into nested groups that are organized as a hierarchical tree. Linkage method

Generic Data Clustering Analysis





Hierarchical Method

Partition Method

All need some sort of distances being calculated!!

Galaxy Clustering Analysis

1. Two groups: clustered galaxies and non-clustered galaxies

Supervised Learning/Classification analysis

2. Unknown number of galaxy clusters

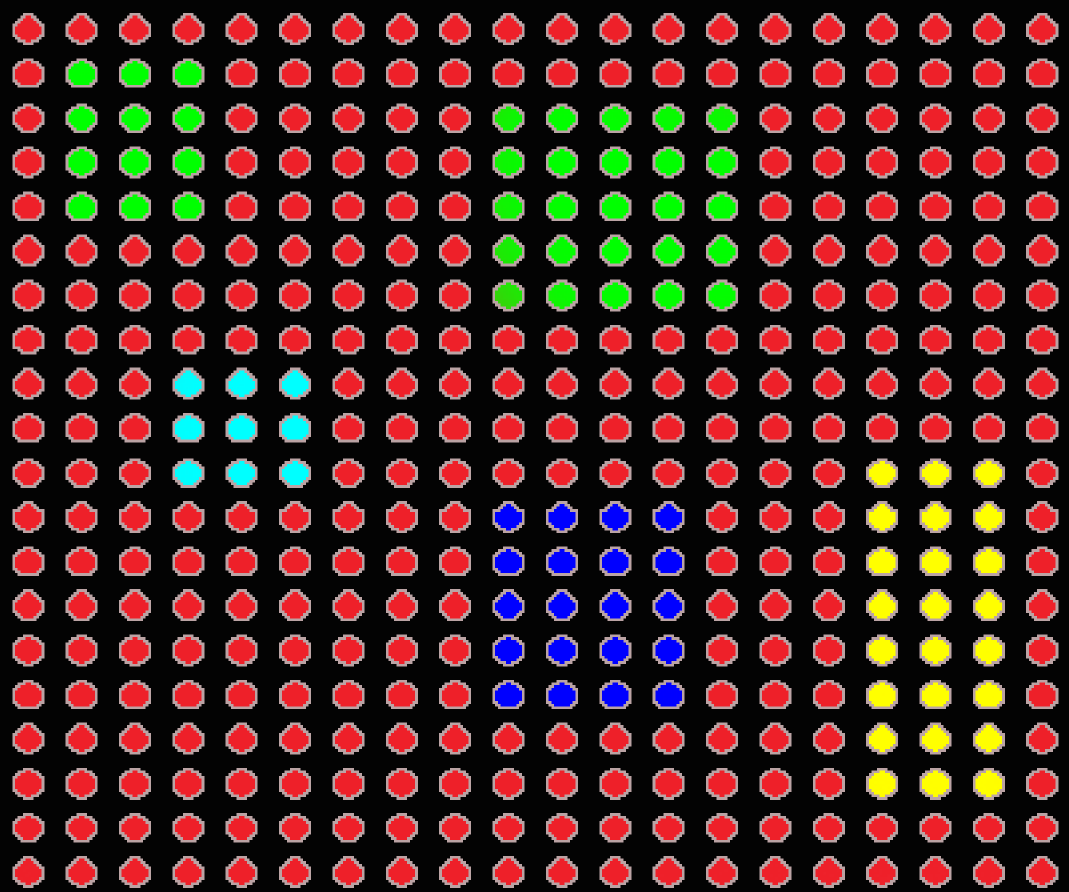
Unsupervised Learning/Clustering analysis

3. Huge uncertainties for the dimension along the line of sight

How to calculate distance? What is the metric?

Trying to look for 3D clustering with only 2D precise information!!

The Challenge due to the projections



Information about the third dimension is crucial!

De-projecting the field galaxies

- Spectroscopic redshift:
  most galaxies do not have spectra taken

REJECTED

- Photometric redshift:
  ~ 0.03 uncertainties
  In clusters, require 0.003, ~ 900km/sec

REJECTED

Fermilab-Chicago photoz group, Oyaizu et al, 2008

- Red sequence colors:
  color uncertainties ~ 0.03
  ridgeline width ~ 0.06

APPROVED

The photometric redshift is obtained from colors, how can the red sequence color do better than photozs?

Galaxy Cluster Finding Algorithms

Counts in Cell: Couch et al.1991 and Lidman & Peterson 1996

Percolation Algorithm: Davis et al. 1985, Efstathiou et al. 1988, Huchra & Geller 1982, Ramella et al. 2002

Smoothing Kernels Algorithm: Shectman et al, 1985

Adaptive Kernel Algorithm: Gal et al. 2000, 2003, 2006

Matched Filter: Postman et al. 1996

Hybrid and Adaptive Matched Filter: Kepner et al. 1999, Kim et al. 2002, Dong et al, 2007

Cut-and-Enhance: Goto et al. 2002

Voronoi Tessellation: Kim et al. 2002, Lopes et al. 2004

C4 Algorithm, Miller et al, 2005

Cluster Red Sequence Algorithm: Gladders & Yee. 2000, 2005

MaxBCG: Annis et al, 1999, Koester et al, 2007

GMBCG: Hao et al, 2009 in preparation

Red Sequence based de-projection

Gaussian Mixture BCG (GMBCG) in a nutshell

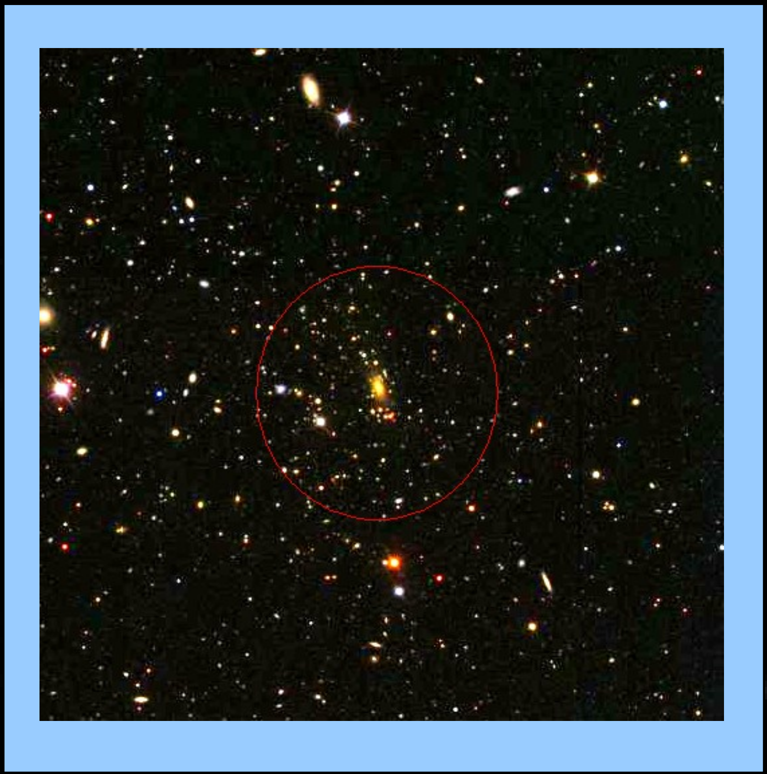
i) Detecting clusters is to look for over-densities of red sequence galaxy distribution.

ii) The center of the cluster is Brightest Cluster Galaxy(BCG)

iii) Use ECGMM algorithm to detect the peak in color space, specifying the red sequence cluster members as within +/- 2 sigma of the ridgeline and more likely from the ridgeline Gaussian component.

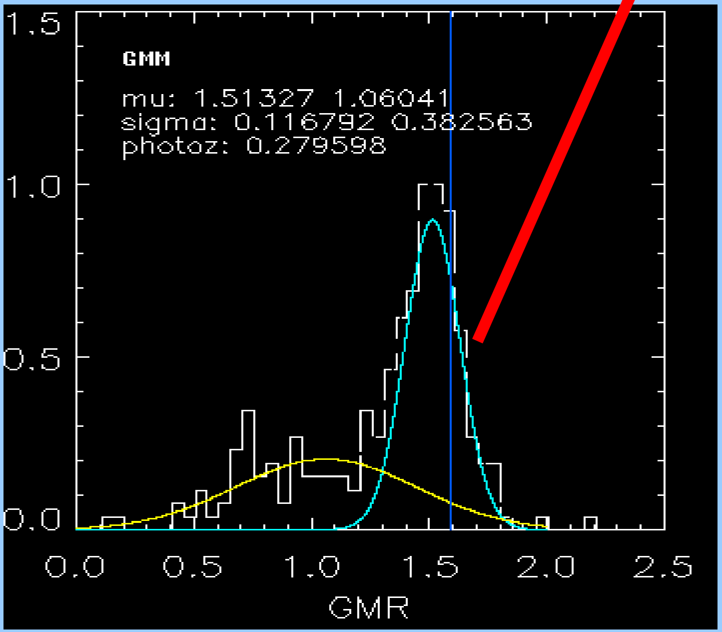
iv) Use NFW kernel(or other kernel) to convolve the projected distribution of the red sequence member galaxies.

Clusters are over-densities

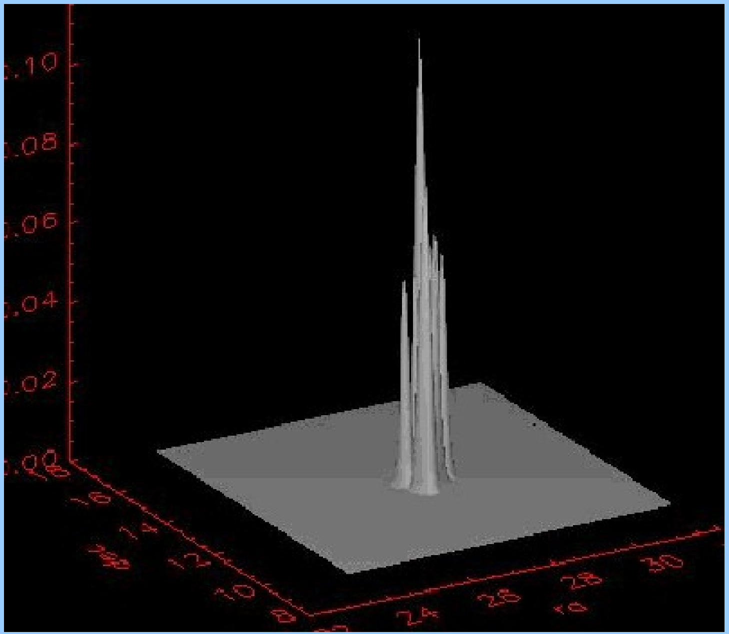


color

Red Sequence

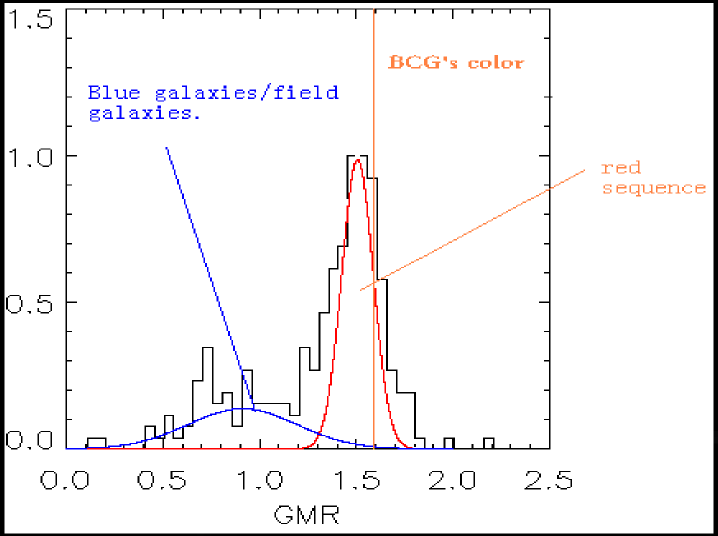


RA/DEC Plane



Detecting Color Clustering---Red Sequence

Color Distribution Around Cluster



Error Corrected Gaussian Mixture Model (ECGMM)

$$\mathcal{L}(\theta|y) = \prod_{j=1}^{M} \left[\sum_{i=1}^{N} \frac{w_i}{\sqrt{2\pi(\sigma_i^2 + \delta_j^2)}} \exp\left(- \frac{(y_j - \mu_i)^2}{2(\sigma_i^2 + \delta_j^2)} \right) \right]$$

More Math & Statistics

EM algorithm

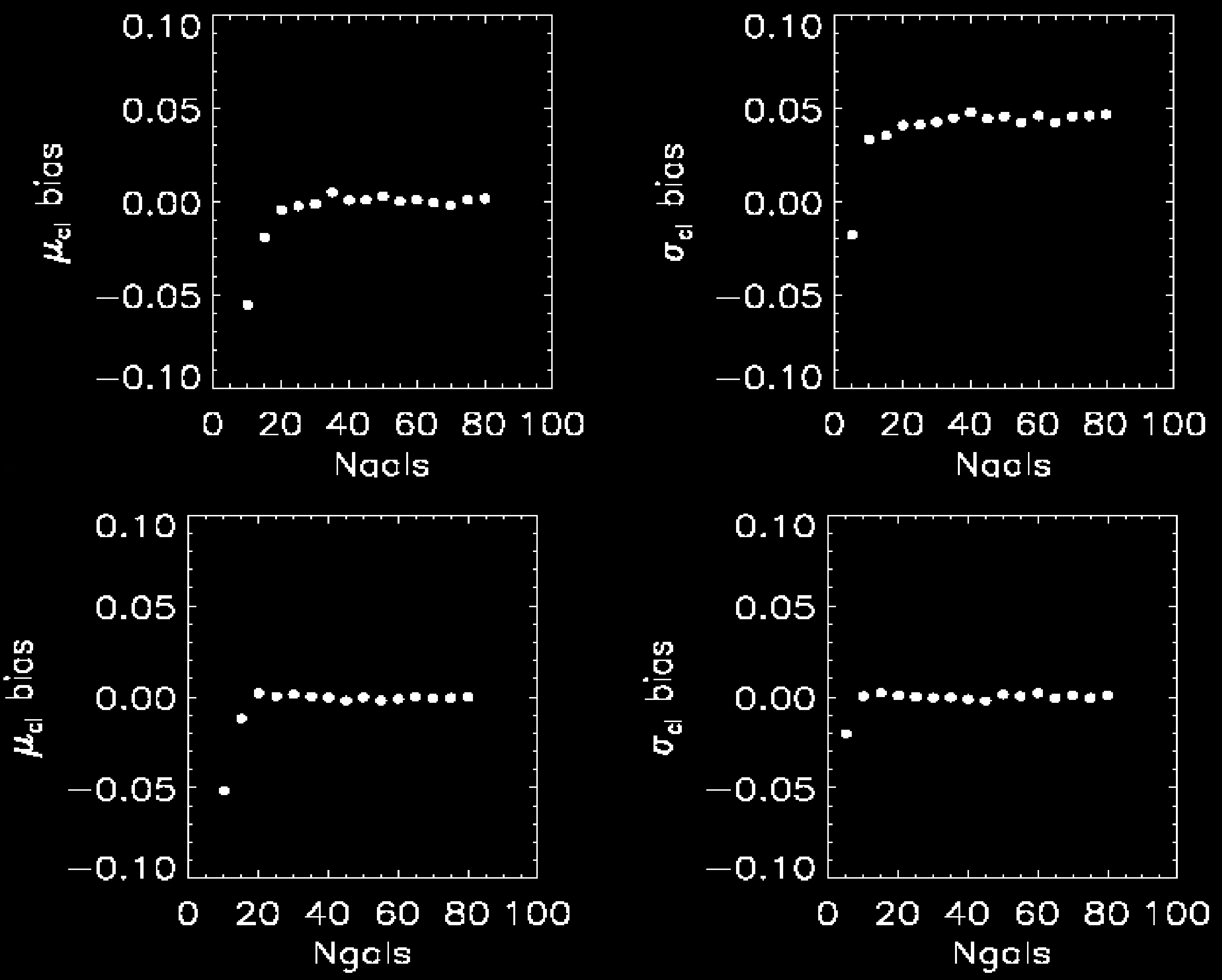
$$w_i^{(t+1)} = \frac{1}{M}\sum_{i=1}^{M} p(z_j = i | y_j, \theta_i^{(t)})$$

$$\mu_i^{(t+1)} = \frac{\sum_{j=1}^{M} y_j p(z_j = i | y_j, \theta_i^{(t)}) / (1 + \delta_j^{(t)2} / \sigma_i^{(t)2})}{\sum_{i=1}^{M} p(z_j = i | y_j, \theta_i^{(t)}) / (1 + \delta_j^{(t)2} / \sigma_i^{(t)2})}$$

$$\sigma_i^{(t+1)} = \left[\frac{\sum_{j=1}^{M} (y_j - \mu_i)^2 p(z_j = i | y_j, \theta_i^{(t)}) / (1 + \delta_j^{(t)2} / \sigma_i^{(t)2})}{\sum_{j=1}^{M} p(z_j = i | y_j, \theta_i^{(t)}) / (1 + \delta_j^{(t)2} / \sigma_i^{(t)2})}\right]^{1/2}$$

$$BIC = -2 \log \mathcal{L}_{max} + \ k \log(M)$$

Monte Carlo Test of ECGMM



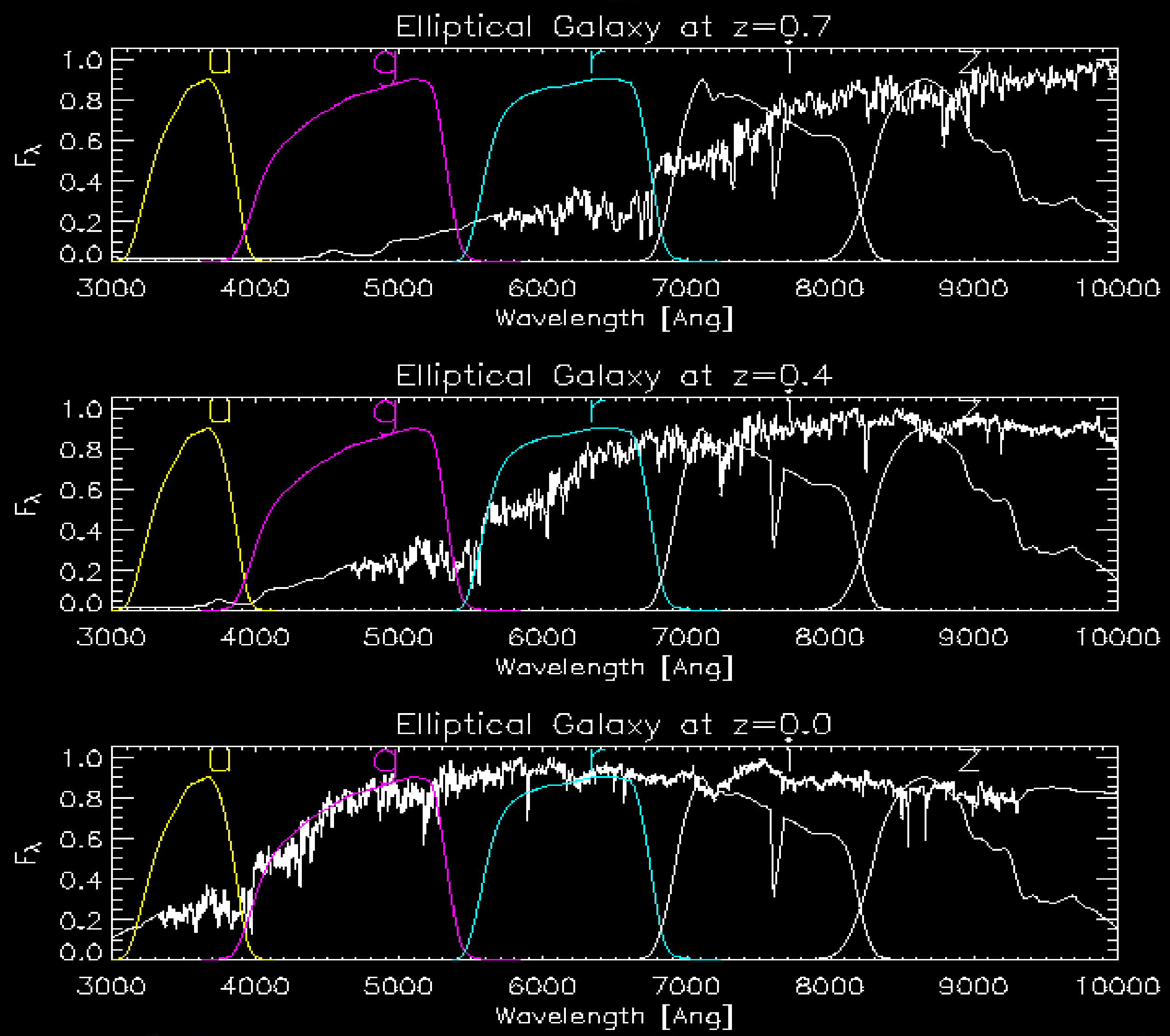
GMM

ECGMM

Bias = <estimated value> - true value
For 200 replications

Hao et al, 2009, in prep

High Redshift Challenge



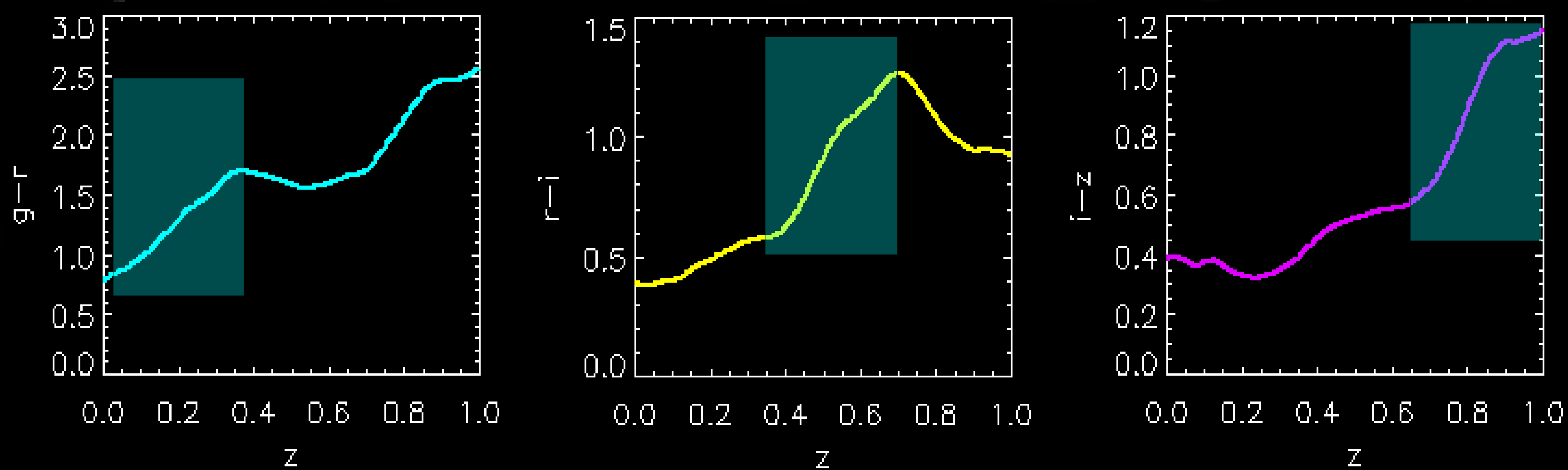
Bruzal & Charlot 2003

Ridgeline Color Selection

Photometric redshift:

Neural network (photoz2 table,Oyaizu et al, 2008), Nearest neighbor polynomial, Nearest Neighbor, Template method (photoz table), Boosted Decision Tree (Gerdes et al 08)

$$P(z_{min} \leq z \leq z_{max}) = \frac{1}{\sqrt{2\pi\sigma^2_{photoz}}} \int_{z_{min}}^{z_{max}} dz \exp\left[-\frac{(z - photoz)^2}{2\sigma^2_{photoz}}\right]$$



color:	g-r	r-i	i-z
redshift:	$0.0 \sim 0.35$	$0.35 \sim 0.70$	$0.70 \sim 1.0$

Likelihood

$$L_{tot} = L_{cluster}^{BCG} \times L_{profile}$$

$$L_{cluster}^{BCG} = \frac{1}{\sqrt{2\pi\sigma_{GM}^2}} \exp\left[-\frac{(C - C_{GM})^2}{2\sigma_{GM}^2}\right]$$

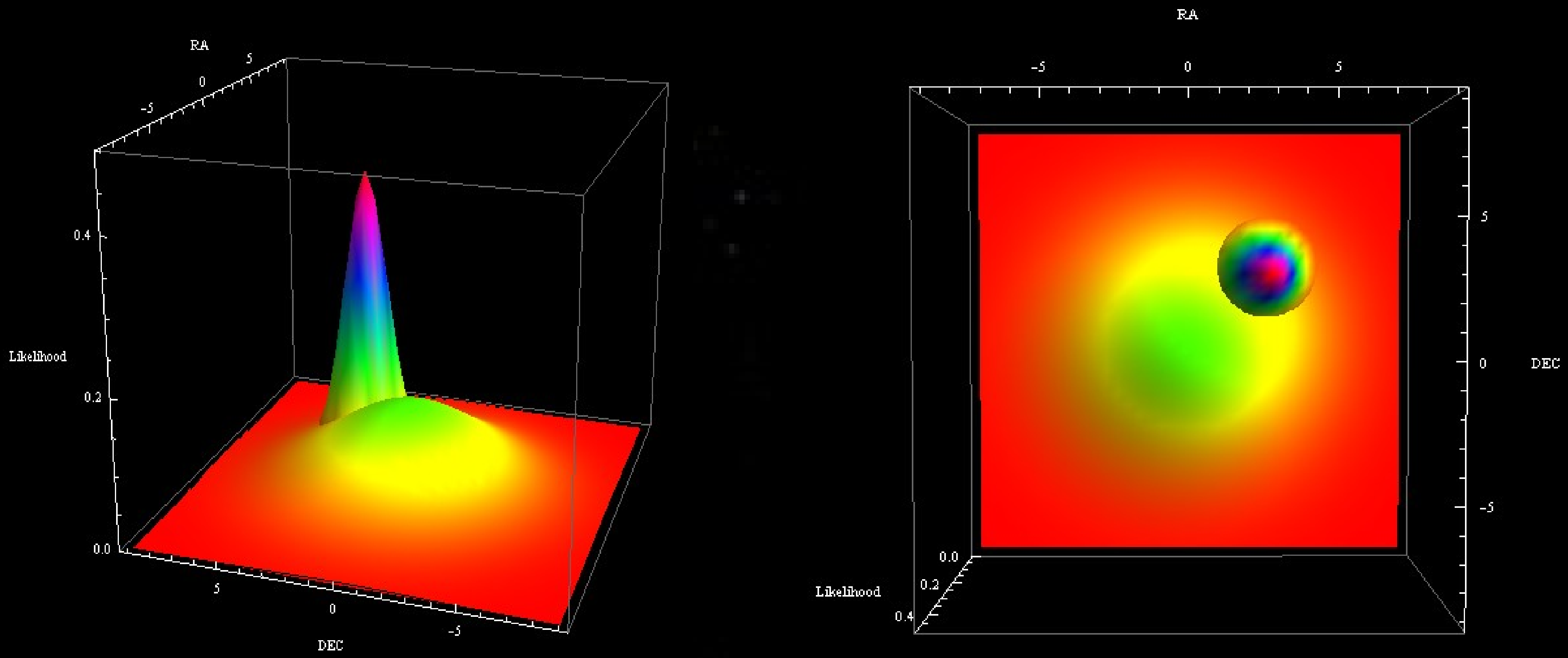
likelihood of the BCG's color belonging to the ridgeline

$$L_{profile} = \sum_{k=1}^{N_g} \Sigma(x_k)$$

Radial profile

No color filter is applied because of the scatter of ridgeline is comparable to the color uncertainty

Interpretation of the Likelihood



Height of the convolved density peak modulated by the closeness of the BCG's color to the ridgeline color.

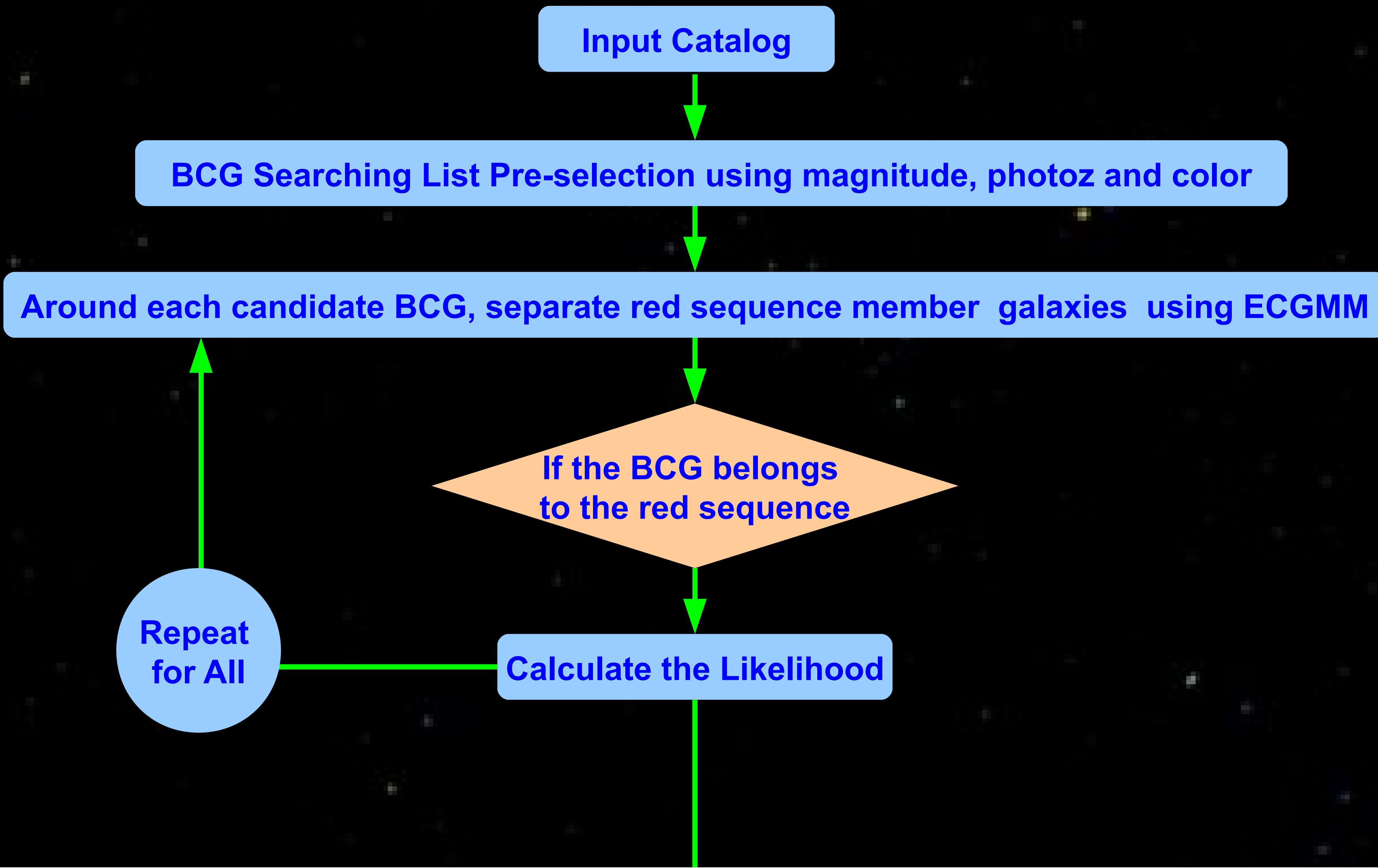
Why Use Brightest Cluster Galaxy as Cluster Centre?

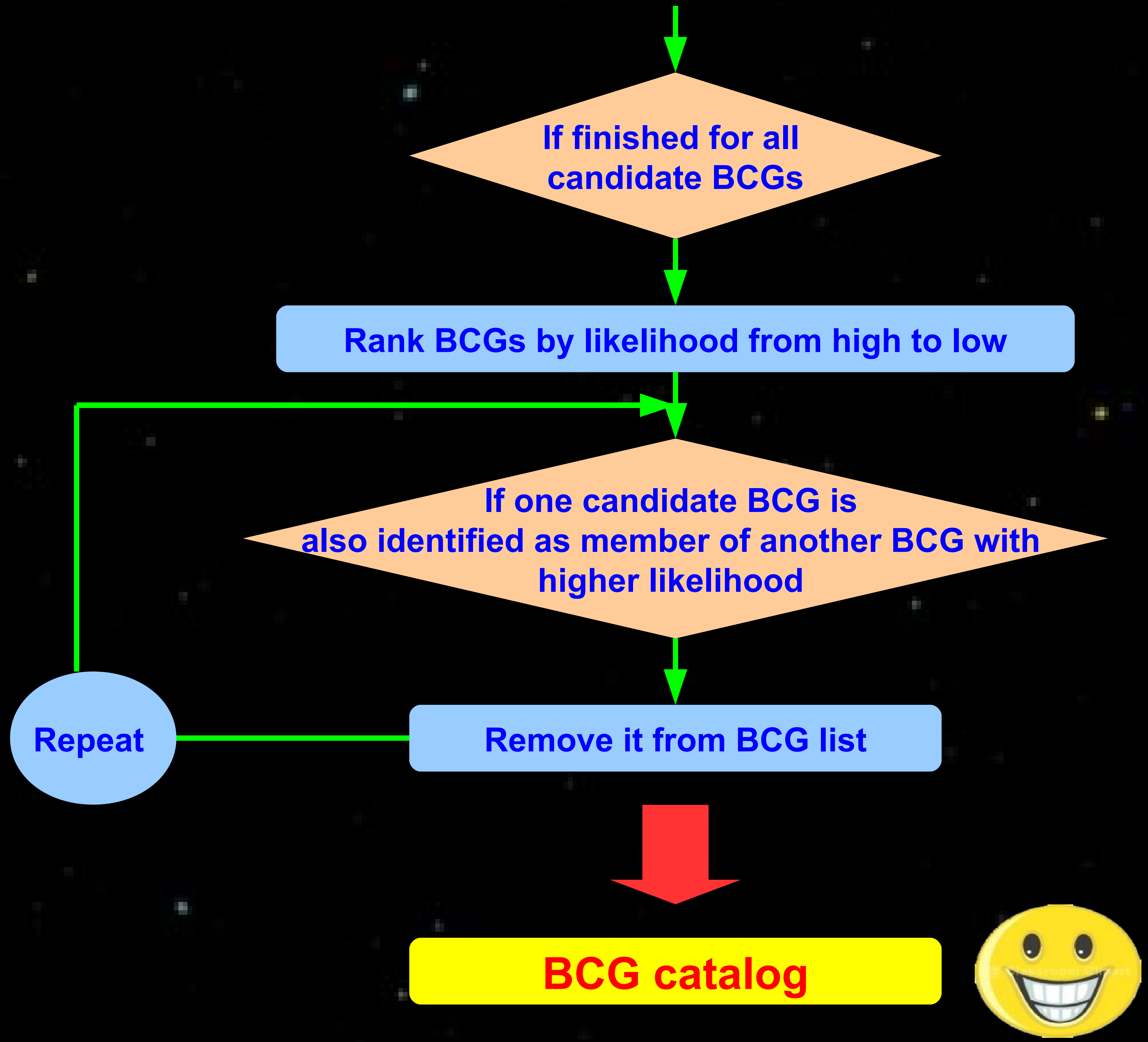
1. Physics motivation: gas is dragged to the bottom of the cluster's gravitational well, producing the BCG.

2. Algorithmic motivation: alleviate effects due to the chance projected galaxies.

3. Computational motivation: boost the efficiency of cluster finding.

Implementation flowchart




If finished for all candidate BCGs
Rank BCGs by likelihood from high to low
If one candidate BCG is also identified as member of another BCG with higher likelihood
Remove it from BCG list
Repeat
BCG catalog

Cartoon Illustration

Comparison with MaxBCG

Koester et al, 2007

GMBCG is using local ridgeline measured using the ECGMM. MaxBCG assume an average/constant ridgeline for all clusters.

GMBCG does not match to a model filters for cluster. It is therefore less biased

GMBCG does not “optimize” the photoz w.r.t to a specific filter. Photoz is from colors only using other algorithm. A photoz can be estimated from the resulting ridgeline as a by product.

Efficient !!!

40 Minutes for a stripe of SDSS DR6 data (about 300 square degree) using Dell PC with 2.8 GHz Intel CPU (1 core) and only 1G RAM.

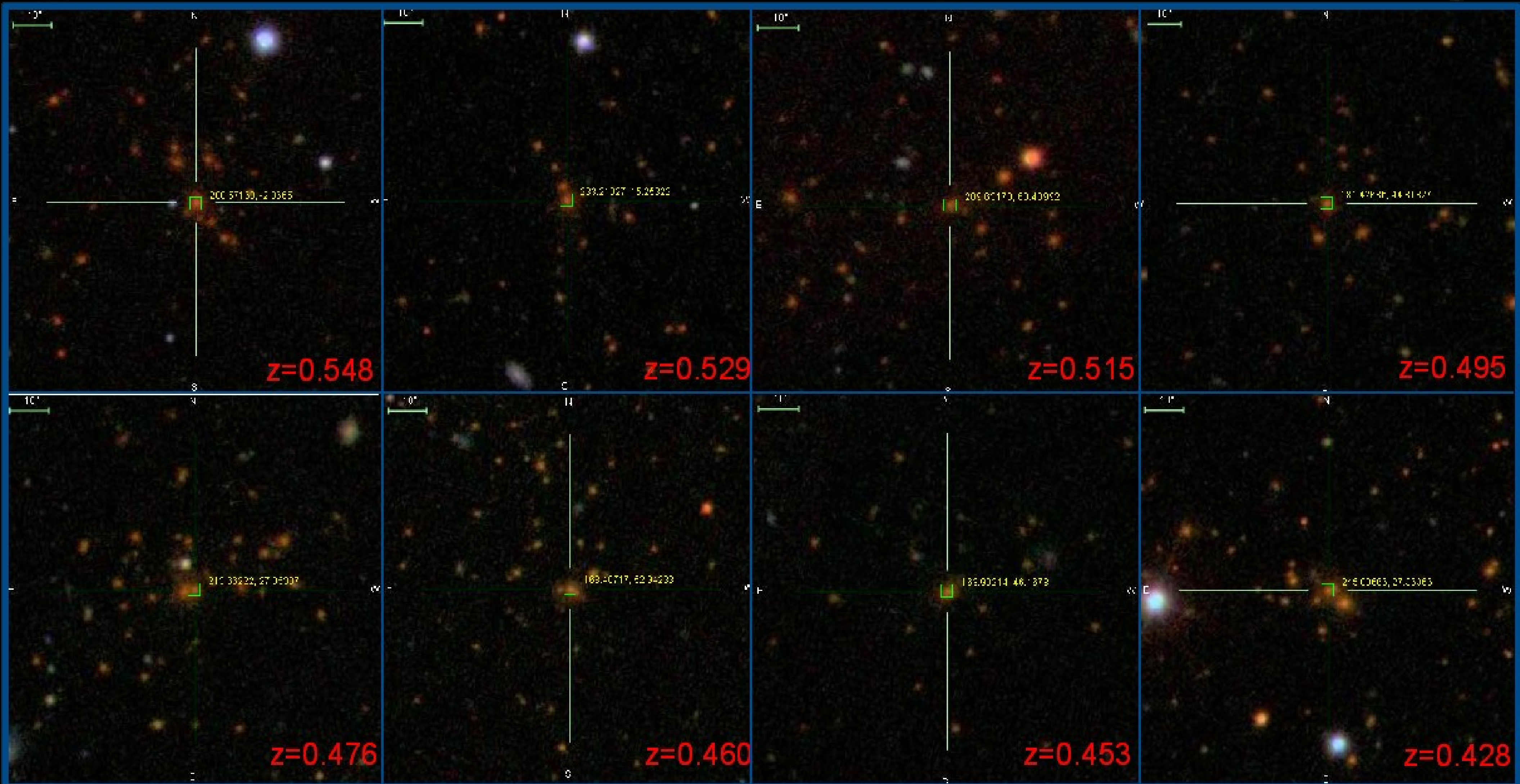
Compared to

30 Hours for a similar stripe using one dual-processor node in a Linux Beowulf Cluster with 3.06 GHz clock speed each. (Dong et al, 2007)

FULL SDSS DR7 Catalog (about 9000 deg^2) within **30 hours** on a Computer less than **$1500**

GMBCG Catalog for SDSS DR7

Hao et al, 2009, in prep



GoogleEarth

Evaluating the Cluster Finder

Completeness: does the cluster finder find all clusters?
Purity: are those found clusters true clusters?

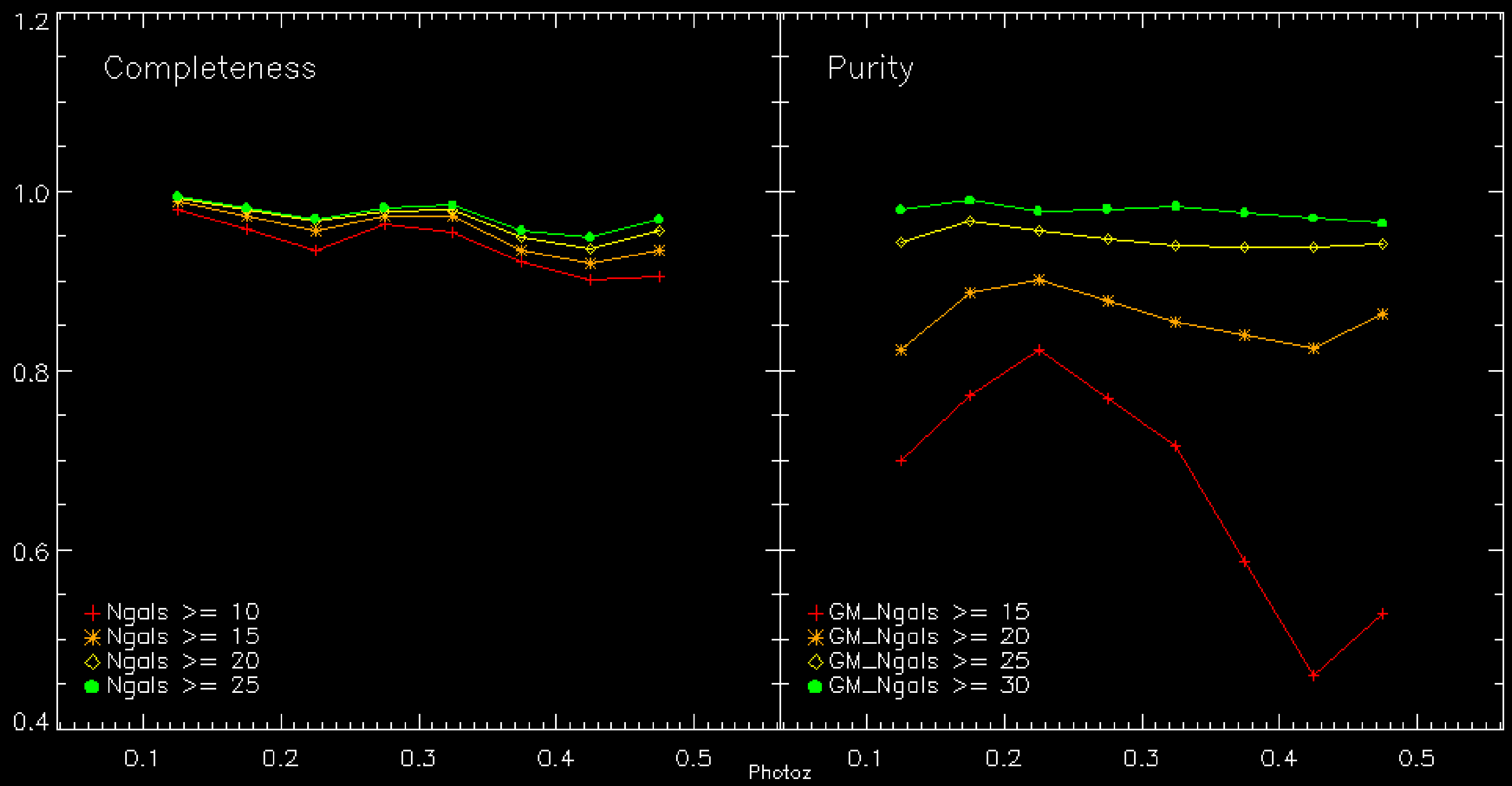
Challenge: How do we know the true clusters?

Mock Catalog

N-Body Simulation: how to add galaxies and make it realistic?

Monte Carlo Catalog: as realistic as possible

Completeness and Purity Based on Monte Carlo Mock Catalog



Getting ready for DES

Dark Energy Survey

1. DES Filter and red sequence ridgeline

Filters: g, r, i, Z, Y

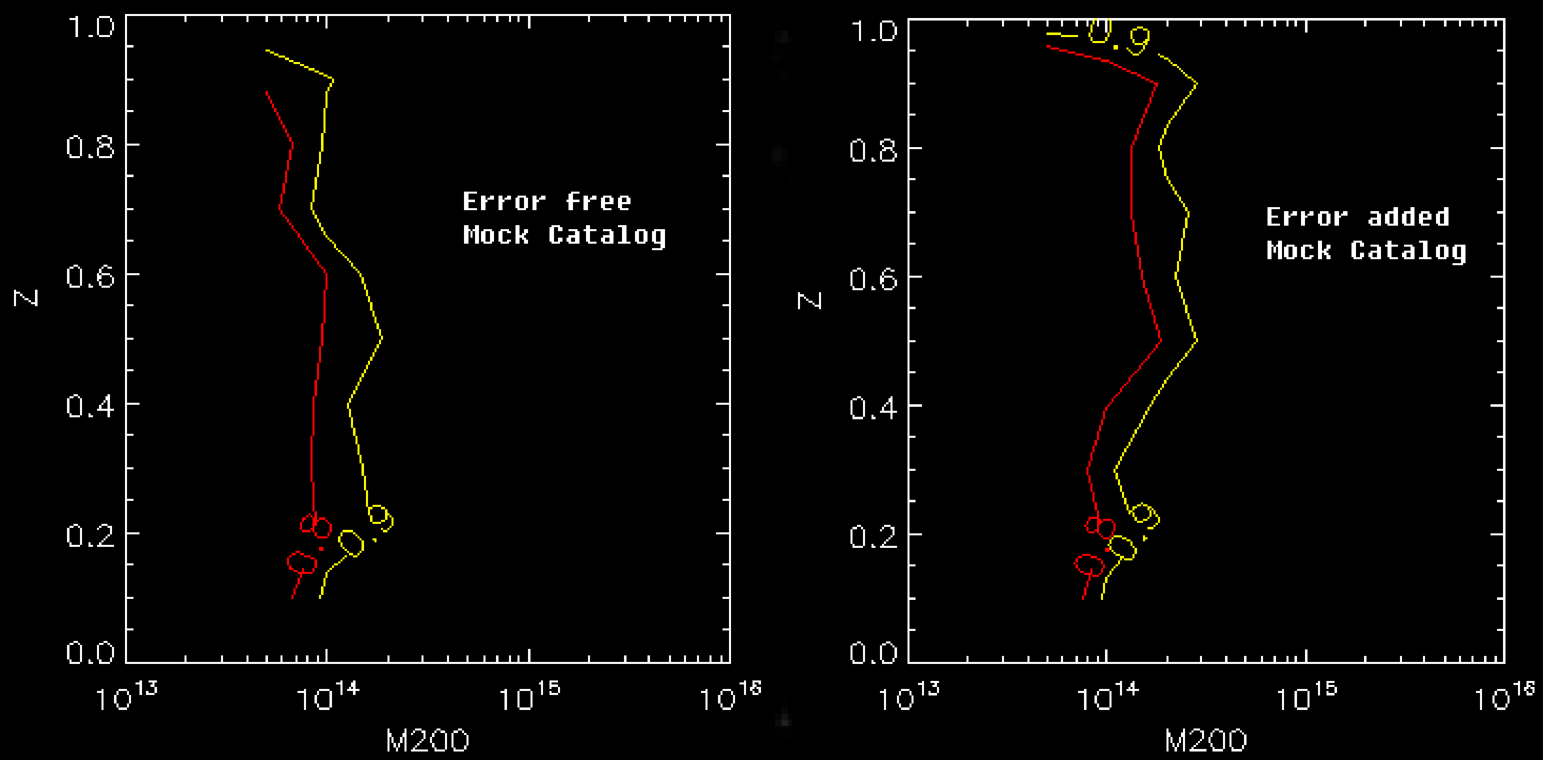
Ridgeline color: g-r, r-i, i- Z, Z-Y

So, to apply GMBCG to DES data, we just need to extend the previous prescriptions to include i-Z and Z-Y ridgelines

Getting ready for DES

Dark Energy Survey

2. Applying GMBCG to DES Mock Catalog



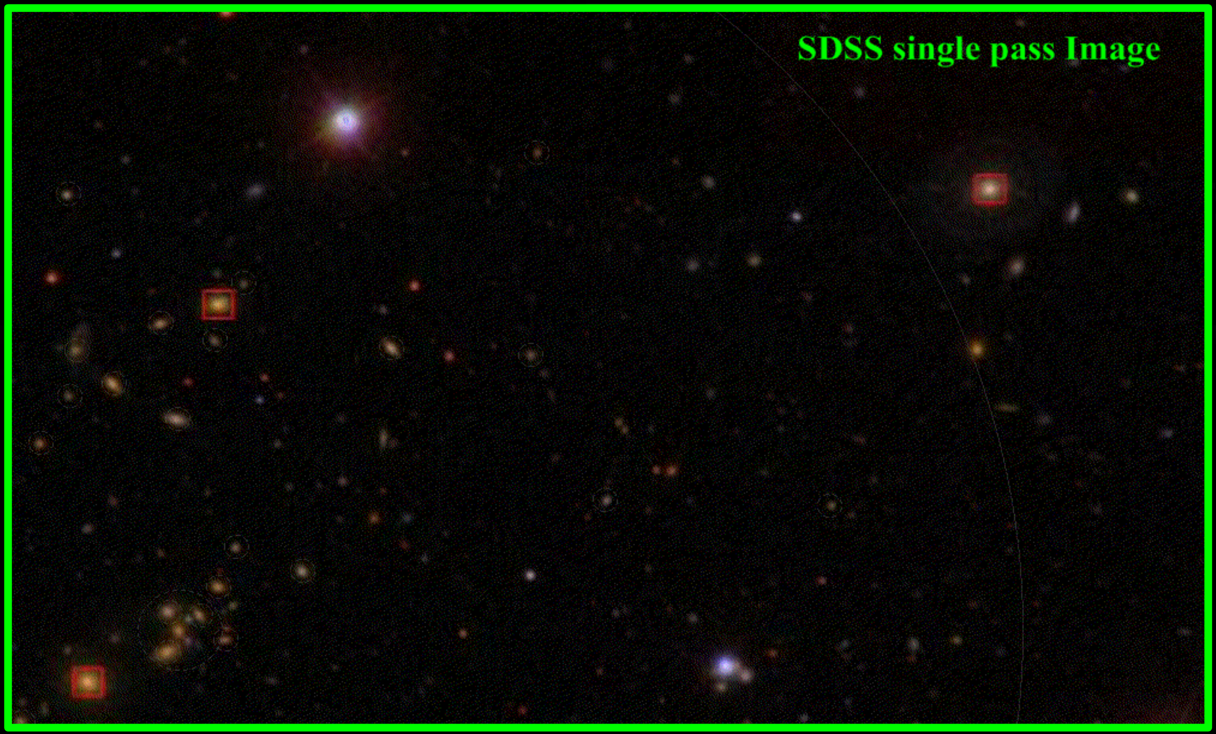
Completeness

(DES Mock catalog from Michael Busha and Risa Wechsler, v1.04)

Getting ready for DES

Dark Energy Survey

3. SDSS coadd and SDSS single pass



Applying GMBCG to latest SDSS Coadd is undergoing. It can help us to understand the transition from first year DES and final DES.

Refined Measurements on Clusters

Ongoing

Motivation: narrow down the mass – observable scatter → improving the cosmological constraints

1. **Better Centers: Improve the weak lensing analysis**

   Sheldon et al, 2007, Johnston et al, 2007

2. **Better understanding of the cluster ridgeline—better richness**

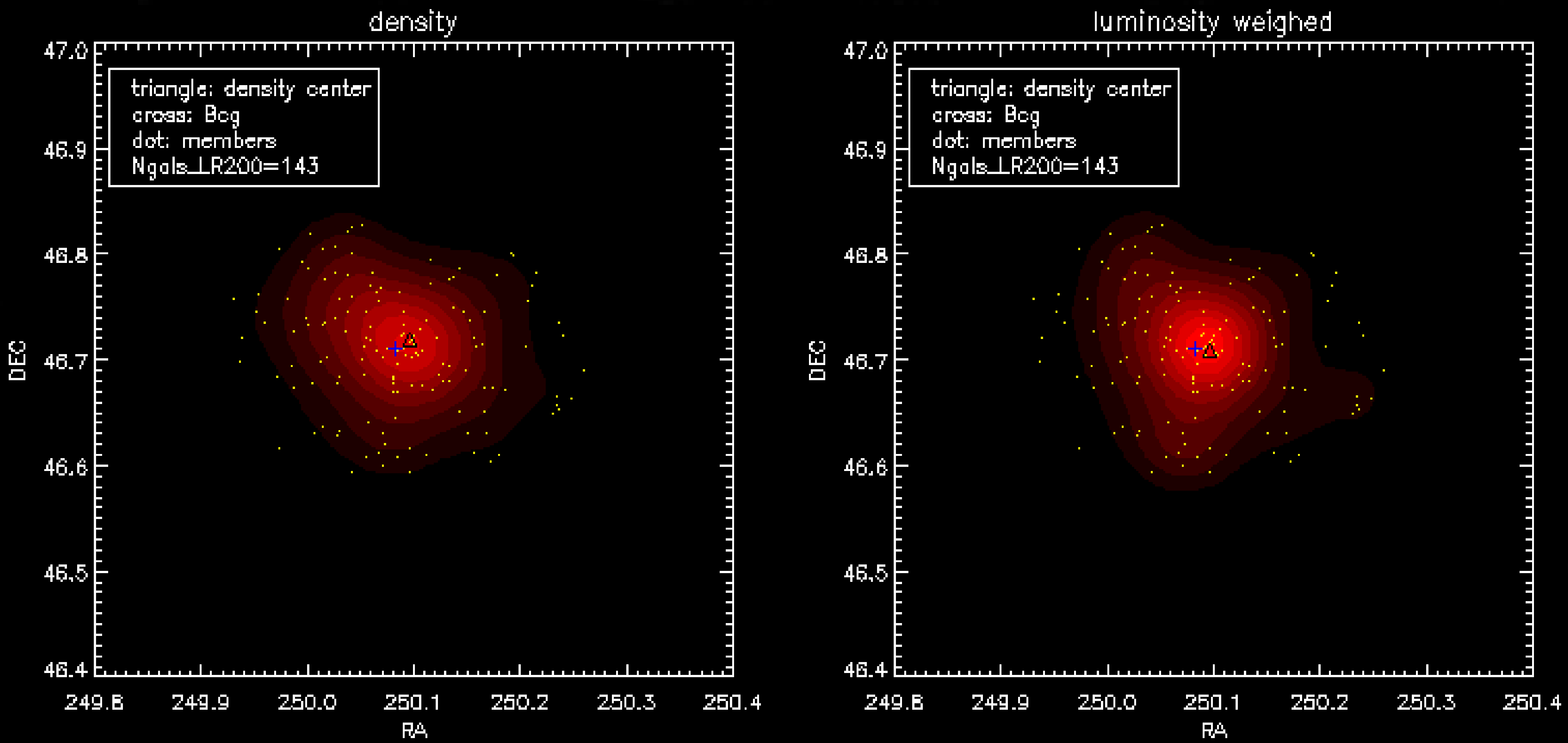
Cluster Centering

Is BCG always the best center of cluster?

Is the center of the cluster member distribution a better center?

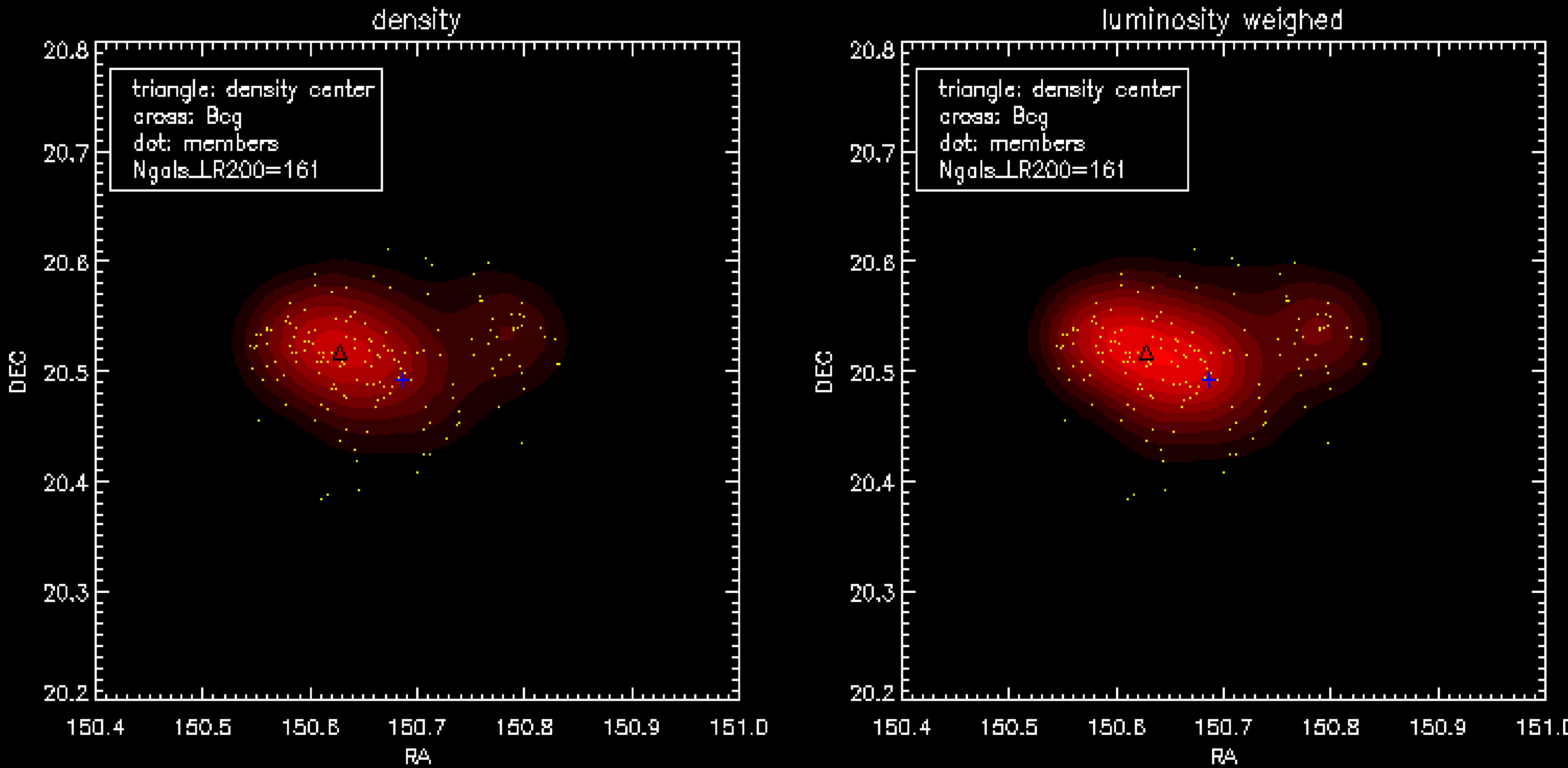
Based on MaxBCG clusters

Cluster Centering



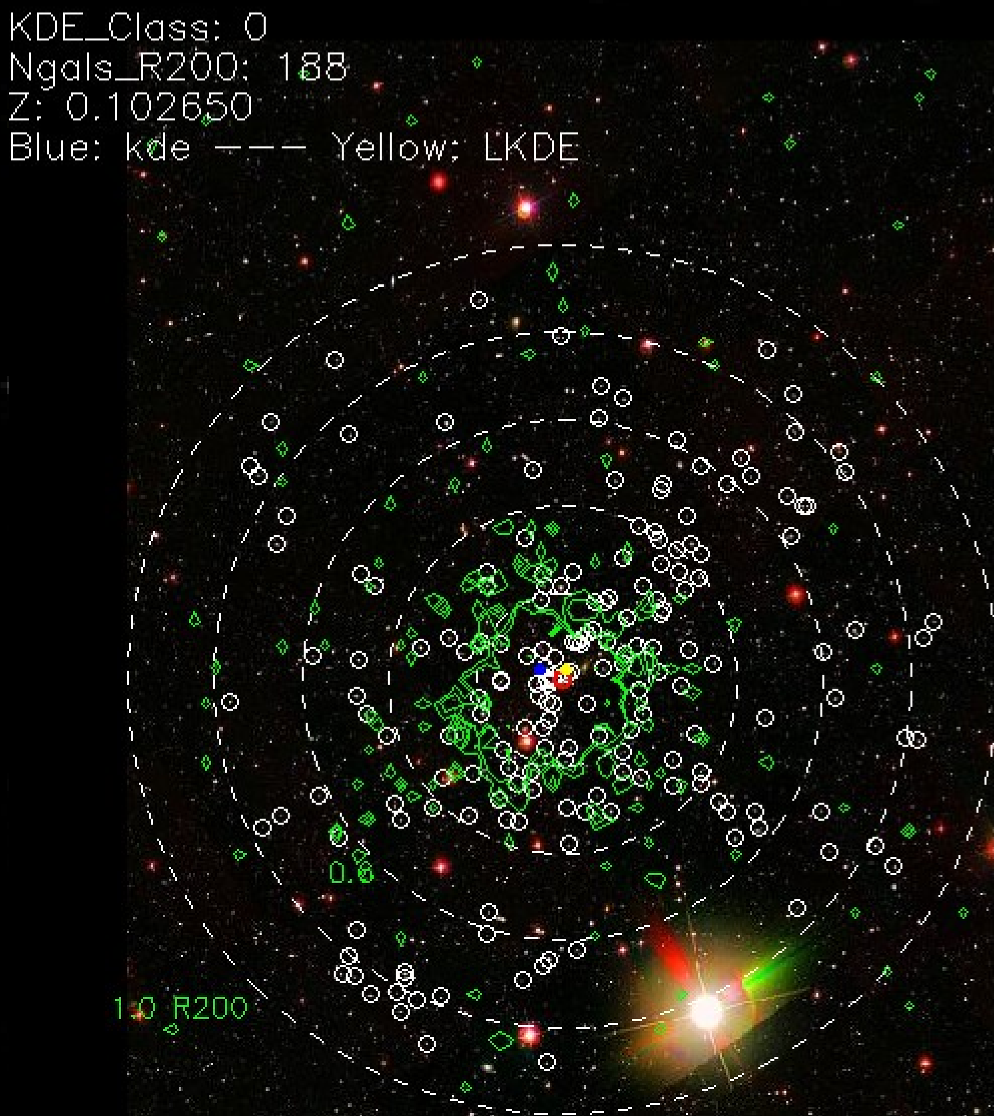
Case 1: BCG is close to the density center

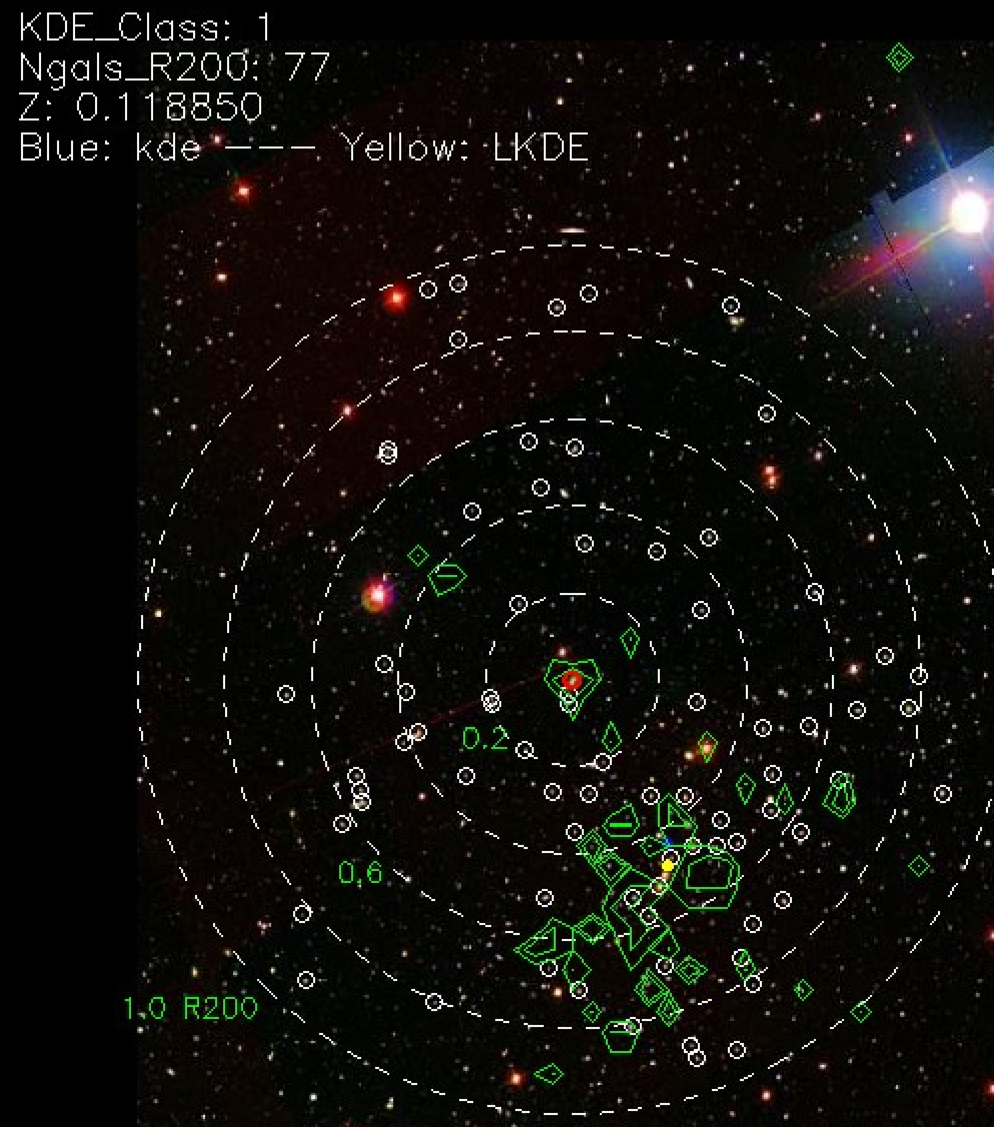
Cluster Centering



Case 2: BCG is apart from the density center

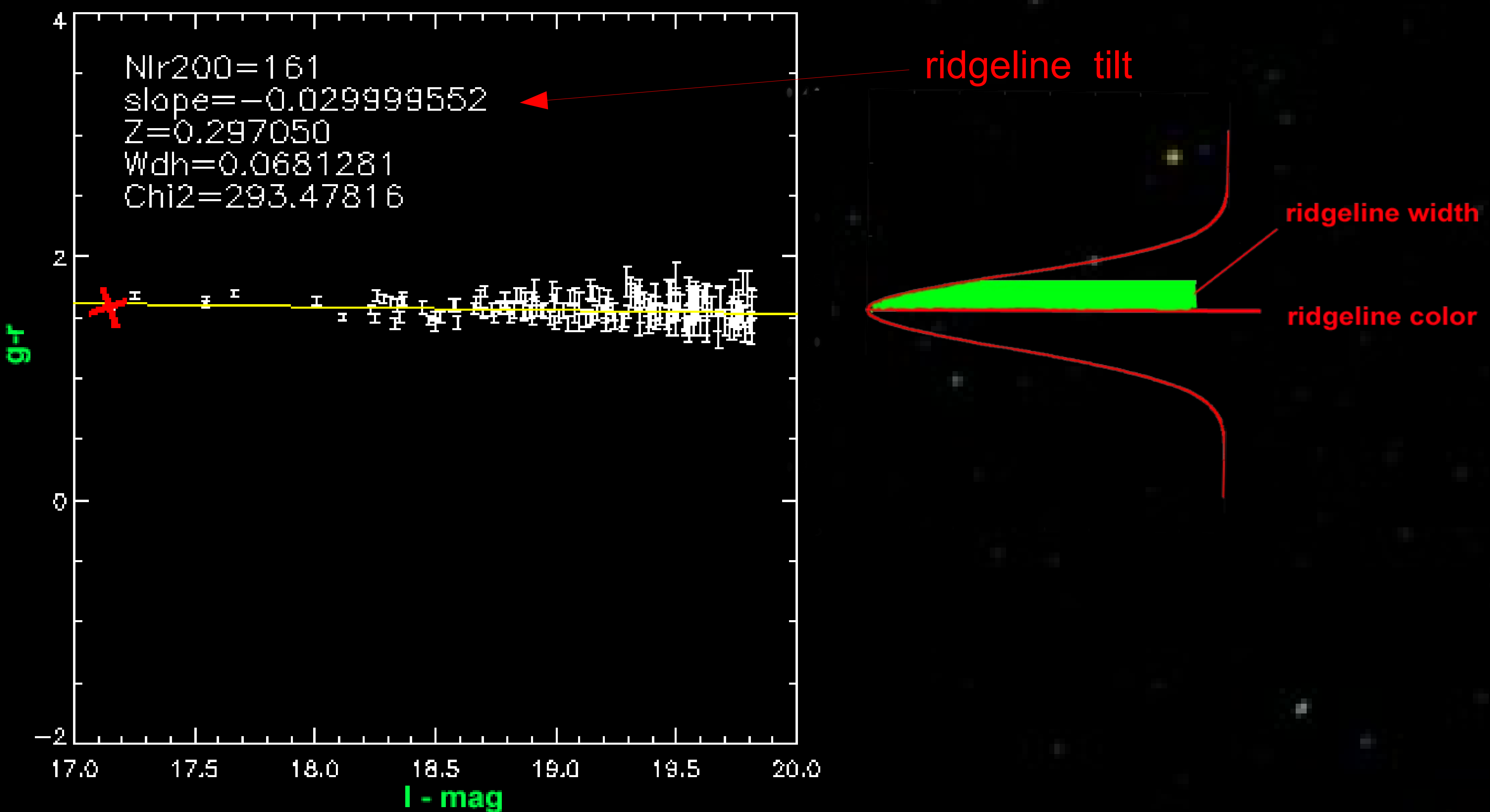
Cluster Centering



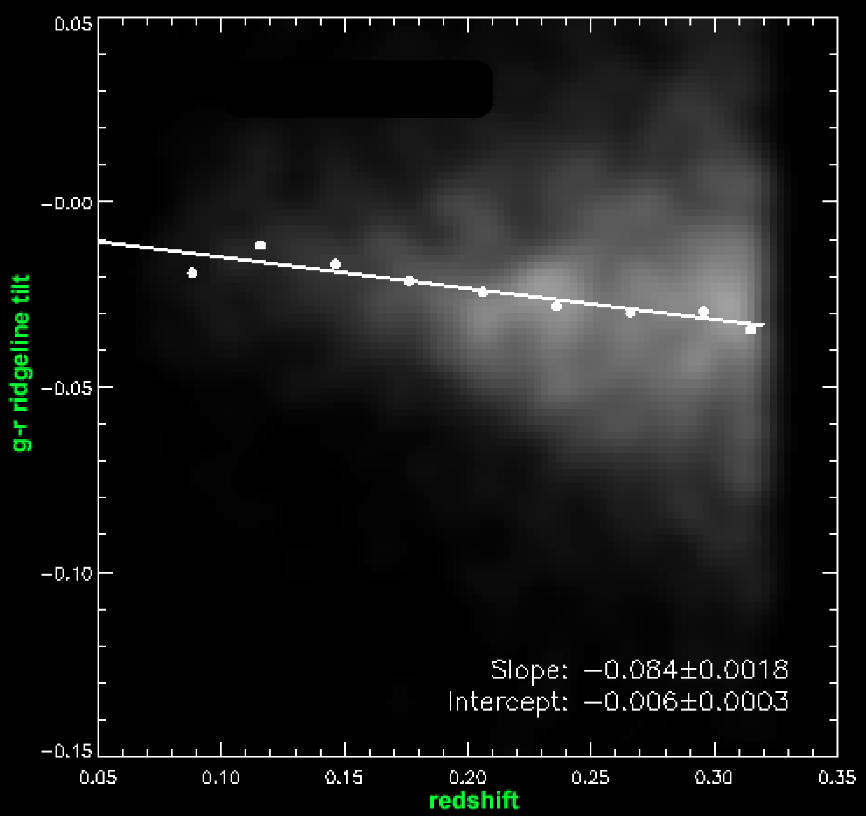


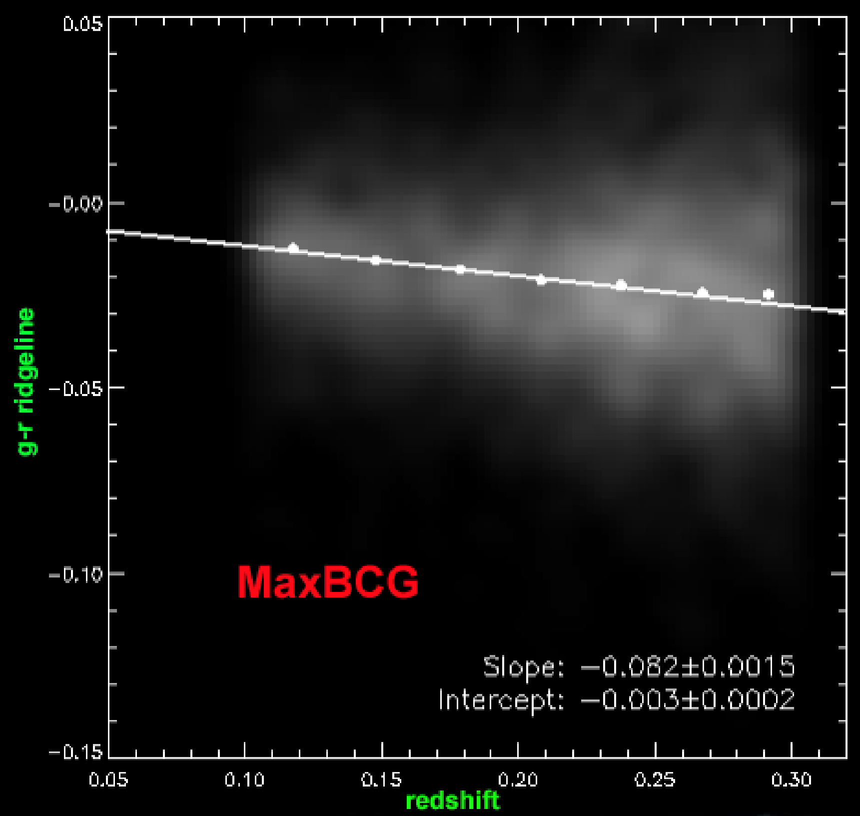
X-ray Center

Precision Measurements of E/S0 Ridgeline and its Tilt



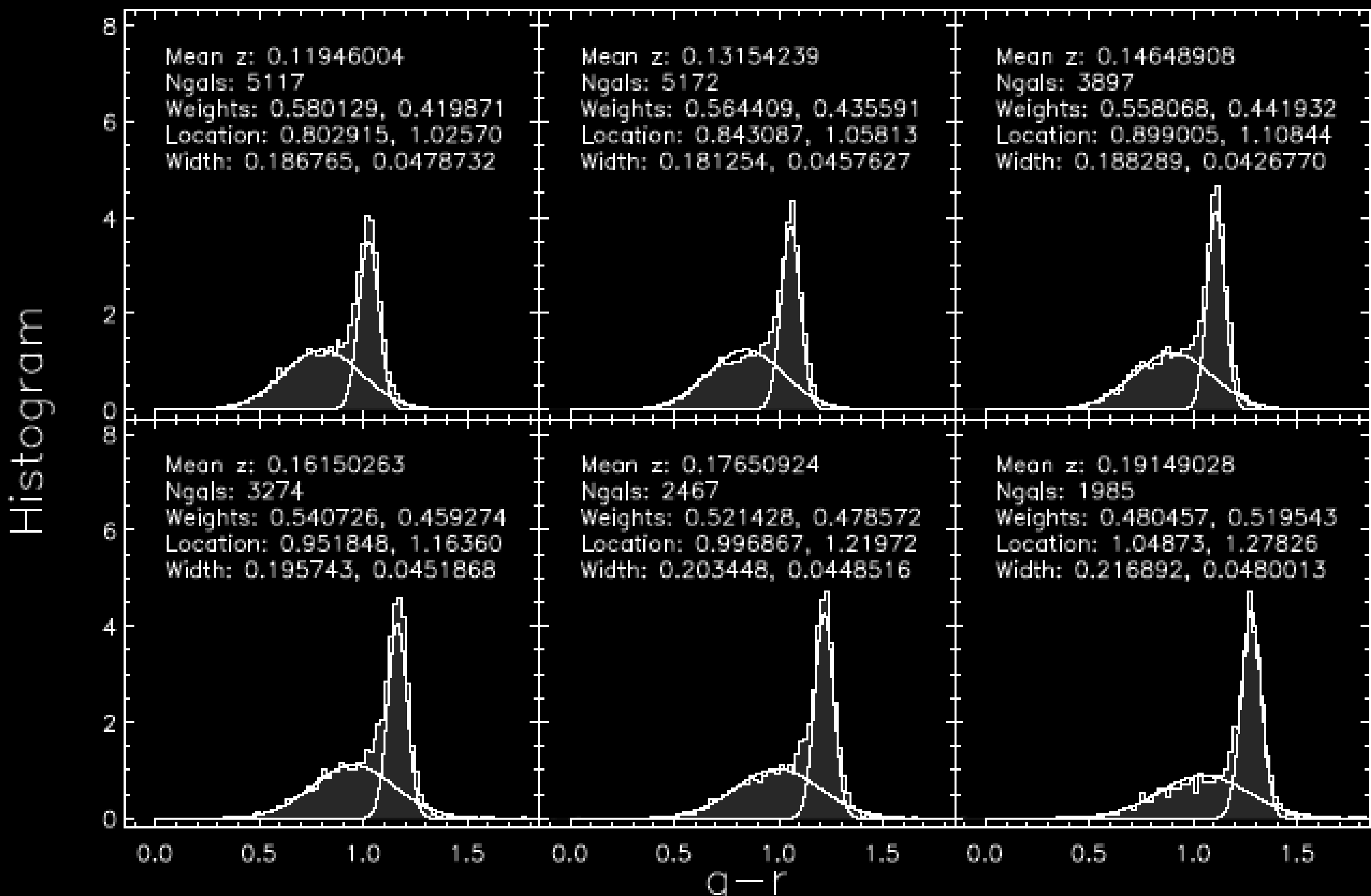
Evolution of Ridgeline and its Tilt





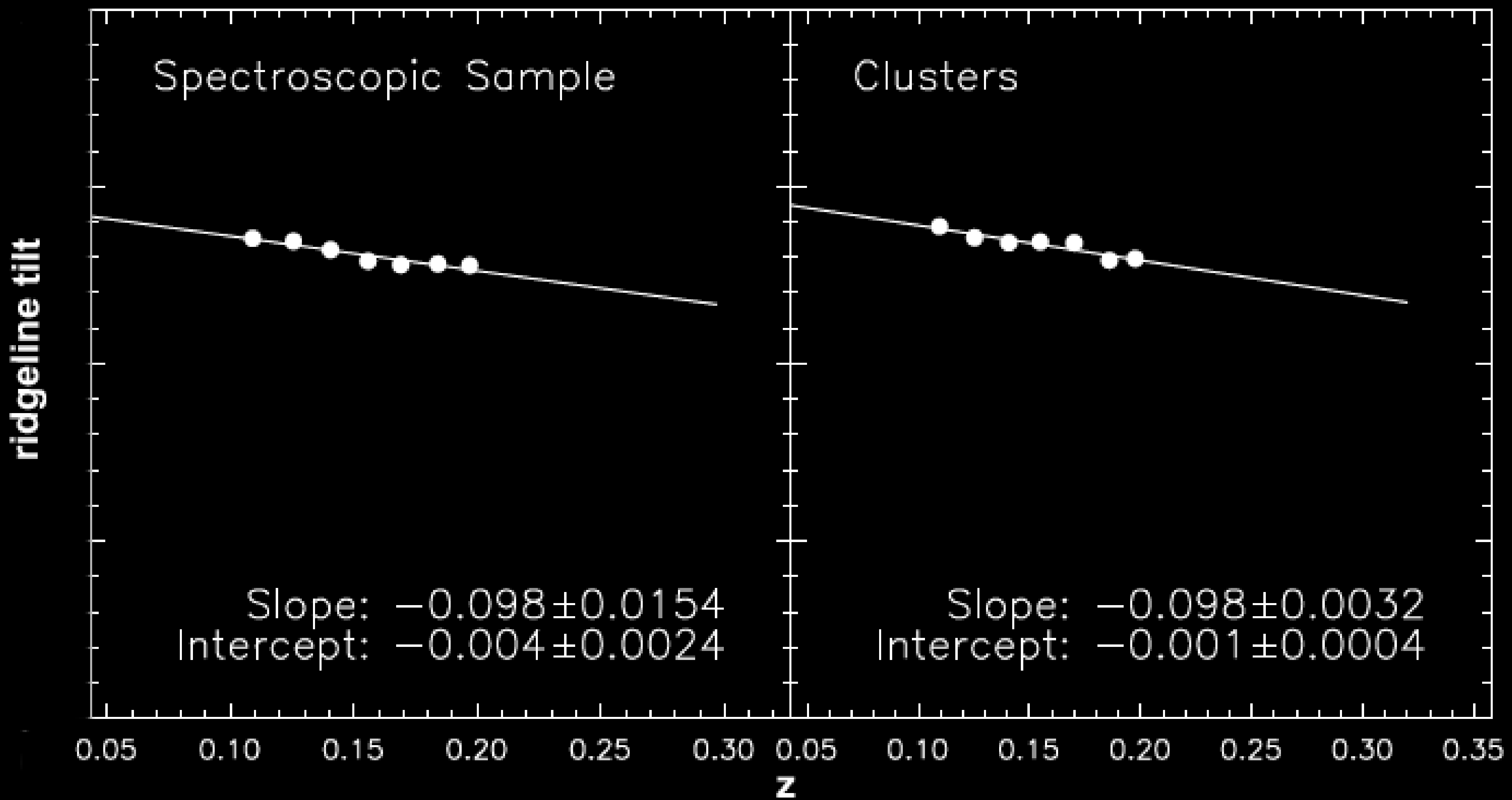
Hao et al, 2009, in prep

Cluster Environment Matters?



Non-clustered spectroscopic data, binned into z slice of 0.003

Cluster Environment Matters?



Cluster environment is not important for CMD, in agreement with the result from Hogg et al, 2004

Better Richness Estimate

Richness Measure	Scatter of X-ray ~ richness relation Rykoff *et al*, 2007
Ngals_R200 Koester *et al*, 2007	0.95
Ngals_R200*LBCG^0.79 Reyes *et al*, 2008	0.84
Ngals_GMM Hao *et al*, 2009	0.87
Ngals_ECGMM Hao *et al*, 2009	0.84
Lambda Rozo *et al*, 2008	0.79

Based on MaxBCG clusters

Legacy from Fermi

A meeting with Enrico Fermi

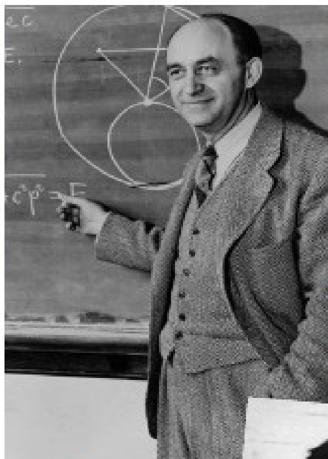
How one intuitive physicist rescued a team from fruitless research.

Freeman Dyson

One of the big turning points in my life was a meeting with Enrico Fermi in the spring of 1953. In a few minutes, Fermi politely but ruthlessly demolished a programme of research that my students and I had been pursuing for several years. He probably saved us from several more years of fruitless wandering along a road that was leading nowhere. I am eternally grateful to him for destroying our illusions and telling us the bitter truth.

Fermi was one of the great physicists of our time, outstanding both as a theorist and as an experimenter. He led the team that built the first nuclear reactor in Chicago in 1942. By 1953 he was head of the team that built the Chicago cyclotron, and was using it to explore the strong forces that hold nuclei together. He made the first accurate measurements of the scattering of mesons by protons, an experiment that gave the most direct evidence then available of the nature of the strong forces.

At that time I was a young professor of theoretical physics at Cornell University, responsible for directing the research of a small army of graduate students and postdocs. I had put them to work calculating meson–proton scattering, so that their theoretical calculations could be compared with Fermi's measurements. In 1948 and 1949 we had made similar calculations of atomic processes, using the theory of quantum electrodynamics, and found spectacular agreement between experiment and theory. Quantum electrodynamics is the theory of electrons and photons interacting through electromagnetic forces. Because the electromagnetic forces are weak, we could calculate the atomic processes precisely. By 1951, we had triumphantly finished the atomic calculations and were looking for fresh fields to conquer. We decided to use the same techniques of calculation to explore the strong nuclear forces. We began by calculating meson–proton scattering, using a theory of the strong forces known as pseudoscalar meson theory. By the spring of 1953, after heroic efforts, we had plotted theoretical graphs of meson–proton scattering. We joyfully observed that our calculated numbers agreed pretty well with Fermi's measured numbers. So I made an appointment to meet with Fermi and show him our results. Proudly, I rode the Greyhound bus from Ithaca to Chicago with a package of our theoretical graphs to show to Fermi.



Crossed paths: A discussion with Enrico Fermi (above) made Freeman Dyson (right) change his career direction.

When I arrived in Fermi's office, I handed the graphs to Fermi, but he hardly glanced at them. He invited me to sit down, and asked me in a friendly way about the health of my wife and our newborn baby son, now fifty years old. Then he delivered his verdict in a quiet, even voice. "There are two ways of doing calculations in theoretical physics", he said. "One way, and this is the way I prefer, is to have a clear physical picture of the process that you are calculating. The other way is to have a precise and self-consistent mathematical formalism. You have neither." I was slightly stunned, but ventured to ask him why he did not consider the pseudoscalar meson theory to be a self-consistent mathematical formalism. He replied, "Quantum electrodynamics is a good theory because the forces are weak, and when the formalism is ambiguous we have a clear physical picture to guide us. With the pseudoscalar meson theory there is no physical picture, and the forces are so strong that nothing converges. To reach your calculated results, you had to introduce arbitrary cut-off procedures that are not based either on solid physics or on solid mathematics."

In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, "How many arbitrary parameters did you use for your calculations?" I thought for a moment about our cut-off procedures and said, "Four." He said, "I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." With that, the conversation was over. I thanked Fermi for his time and trouble, and sadly took the next bus back to Ithaca to tell the bad news to the students. Because it was important for the students to have their names on a published paper, we did not abandon our calculations immediately. We finished them and wrote a long paper that was duly published in the *Physical Review* with all our names on it. Then we dispersed to find other lines of work. I escaped to Berkeley, California, to start a new career in condensed-matter physics.

Looking back after fifty years, we can clearly see that Fermi was right. The crucial discovery that made sense of the strong forces was the quark. Mesons and protons are little bags of quarks. Before Murray Gell-Mann discovered quarks, no theory of the strong forces could possibly have been adequate. Fermi knew nothing about quarks, and died before they were discovered. But somehow he knew that something essential was missing in the meson theories of the 1950s. His physical intuition told him that the pseudoscalar meson theory could not be right. And so it was Fermi's intuition, and not any discrepancy between theory and experiment, that saved me and my students from getting stuck in a blind alley. ■

Freeman Dyson is at the Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA.

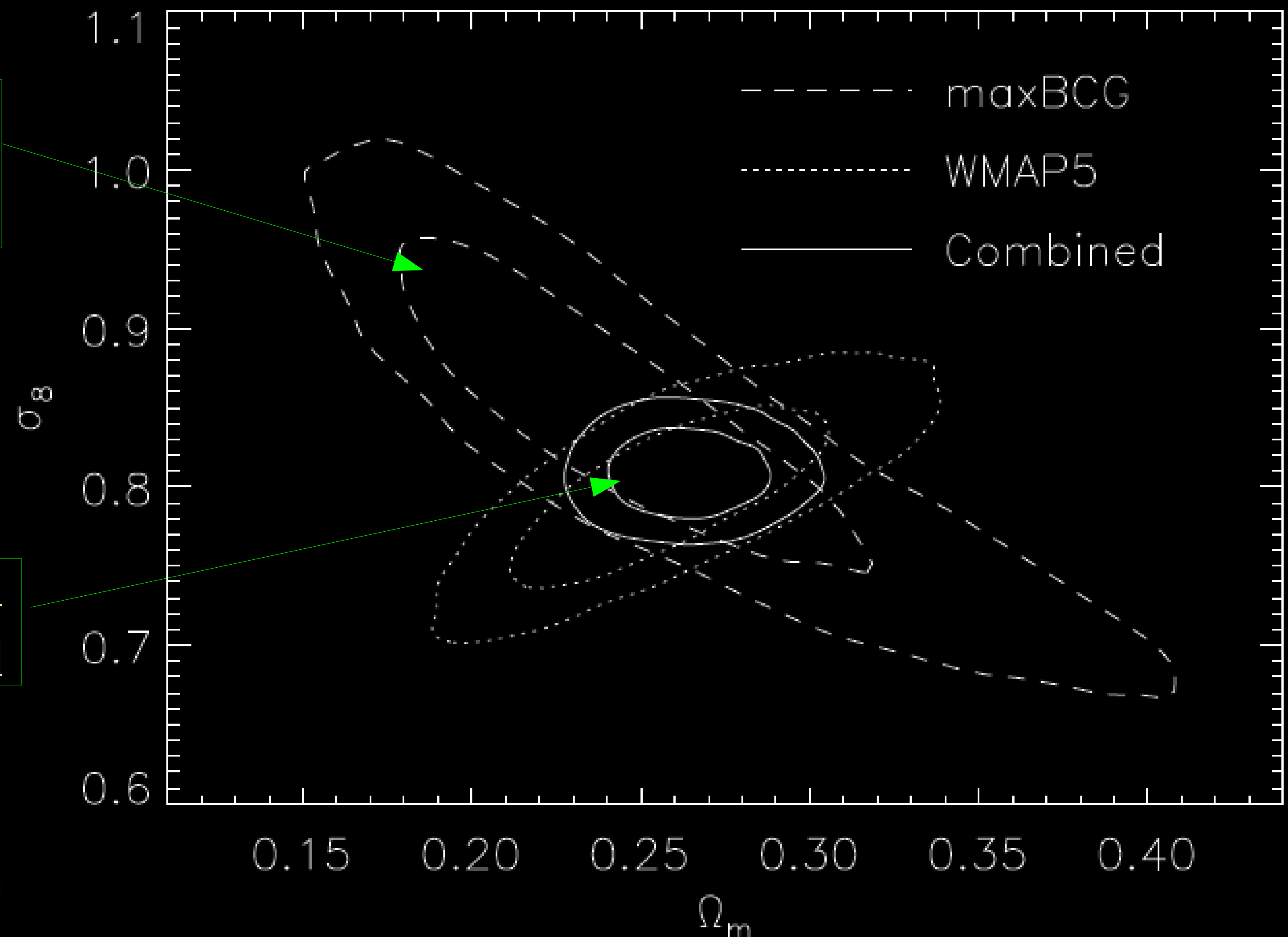
In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, "How many arbitrary parameters did you use for your calculations?" I thought for a moment about our cut-off procedures and said, "Four." He said, "I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk." With that, the conversation was over. I thanked Fermi for his

Thanks

Cosmological Constraints

Rozo et al, 2009, in prep
Based on MaxBCG clusters

$$\sigma_8 = 0.84 \pm 0.07$$
$$\Omega_m = 0.25 \pm 0.05$$

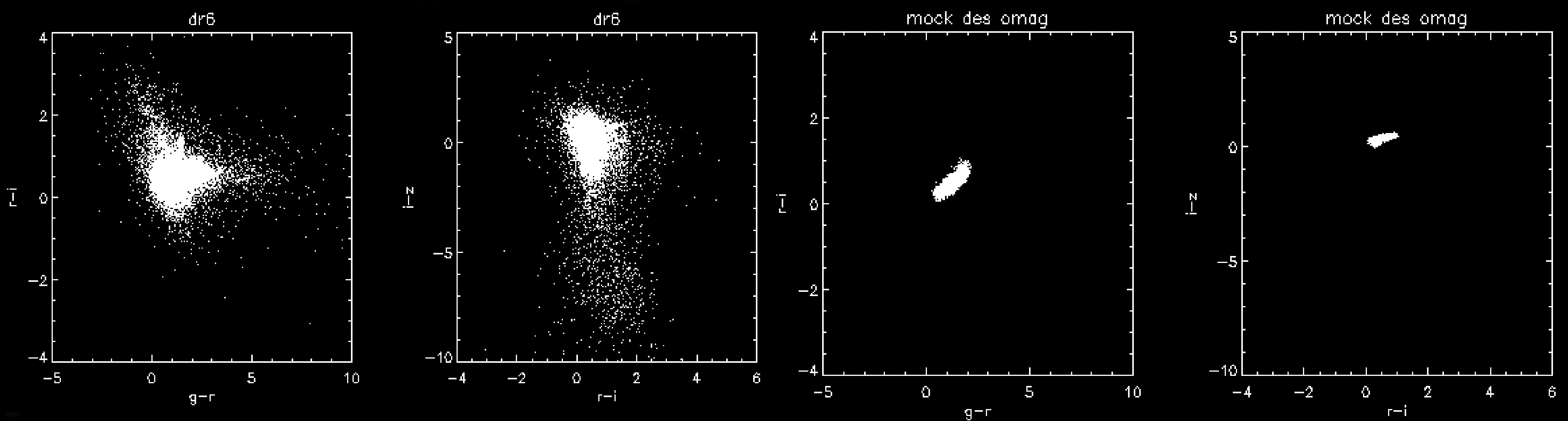


$$\sigma_8 = 0.809 \pm 0.01$$
$$\Omega_m = 0.264 \pm 0.01$$

Getting ready for DES

Dark Energy Survey

2. Applying GMBCG to DES Mock Catalog



Purity is not yet ready because of the unrealistic galaxy colors in mock catalog