Studying the Dynamics and Evolution of Large Elliptical Galaxies at High-redshift (z~0.7) using the LEGA-C spectra and UltraVista photometry

(Extension to Proposal I)

Rachel Bezanson, Department of Physics and Astronomy

Resource Type	SUs Requested	Maximum SUs allowed
MPI	1,000,000	3,200,000
SMP standard (default)	80,000	1,300,000
HTC	0	500,000
GPU	0	100,000

Project Description ‘Testing the Formation of the Largest Elliptical Galaxies’

This project focuses on studying the formation and evolution of these galaxies either by slow channel (cannibalizing neighboring galaxies) or by fast channel (merging of two large spiral galaxies) using the LEGA-C dataset. We want to study the interplay between the history of star formation and the motions of stars in distant galaxies. Although we have made progress on measuring the dynamics of the stars (via the doppler effect), we must make use of more sophisticated modeling techniques to properly explore the range of stellar populations. Using BAGPIPES python package we simulate the galaxy light as a combination of star light and dust and fit them to both photometric data and high-signal-to-noise spectroscopic data using MultiNest nested sampling algorithm and generate star-formation history for each galaxy as well as joint-probability distribution of stellar population parameters like age, mass, metallicity, dust attenuation and nebular emission. Each fit relies upon a few key assumptions about the parametrization (the type of dust, analytic form of star formation history and prior probability distribution); we hope to characterize the robustness of our results to these assumptions by conducting two families of fits with different star-formation histories.

Dr. Rachel Bezanson is the survey scientist of the Large Early Galaxy Astrophysics Census (LEGA-C) survey. It is a 130-night public spectroscopic survey conducted with VIMOS on the Very Large Telescope and has taken ~3200 spectra of galaxies with typical continuum SNR of 20 per angstrom in the redshift range $0.6< z<1.0$, each observed for ~20 hours and fully reduced with a custom-built pipeline. These deep and high resolution spectra hold detailed information about the light coming from stars in these very distant galaxies. In fact, the galaxies are so far that light has been traveling for half the age of the Universe to reach our telescopes, giving a valuable glimpse into a much earlier epoch. The primary goal of Bayesian statistical modeling of these spectra is to characterize their average ages, how rapid or extended star formation has been in the past, their dependence on stellar metallicities and chemical enrichment (which is in turn sensitive to the stellar explosions and merger history) and effects of dynamical masses derived from stellar velocity dispersions on these properties. These are some key pieces of information for the study of galaxy evolution, but have so far only been available for representative galaxy samples in the nearby universe, at look-back times of less than 1 billion years. LEGA-C therefore enables for the first time several studies of these questions about galaxy evolution on long cosmological time scales (up to 8 billion years), especially for the oldest and most massive galaxies the pressing questions that could potentially be answered by this dataset are dynamical evolution, mass assembly through merging and the mechanisms that shut down the star formation.

In previous studies of ages and metal content of galaxies at half the age of the universe, only a small sample of less than 100 galaxies were observed with deep spectroscopic measurements required to obtain these properties. Relatively large and good quality data were available then for studying the same relations for present day galaxies. The state of the art today is a huge improvement in both statistical and quality of data terms. We explicitly perform (much faster now) full bayesian analysis of over 3000 galaxies in >15 dimensional parameter space and compare our findings with what was found before. This study will help us put stronger constraints on these scaling relations at that point of time in the universe and will help provide insights about the dominant processes involved in the evolution of galaxies across the cosmic time. A result from a previous study by Anna Gallazzi 2014 is shown in Fig. 1 and our results with the Proposal I resources for our LEGA-C data with Bayesian Analysis are shown in Fig 2.

We have found some promising results that give some valuable insights about the galaxy's evolution. After extensive initial testing and being a little behind on timeline due to pandemic constraints, we are close to running our final batch of modeling soon and we would like to request the above mentioned resources. We note that we are currently investigating the systematics that could be driving the unphysical cluster of about 100 galaxies at $\log(t_r) \sim 9.5$ from a sample study of 1600.

Computational Demands :

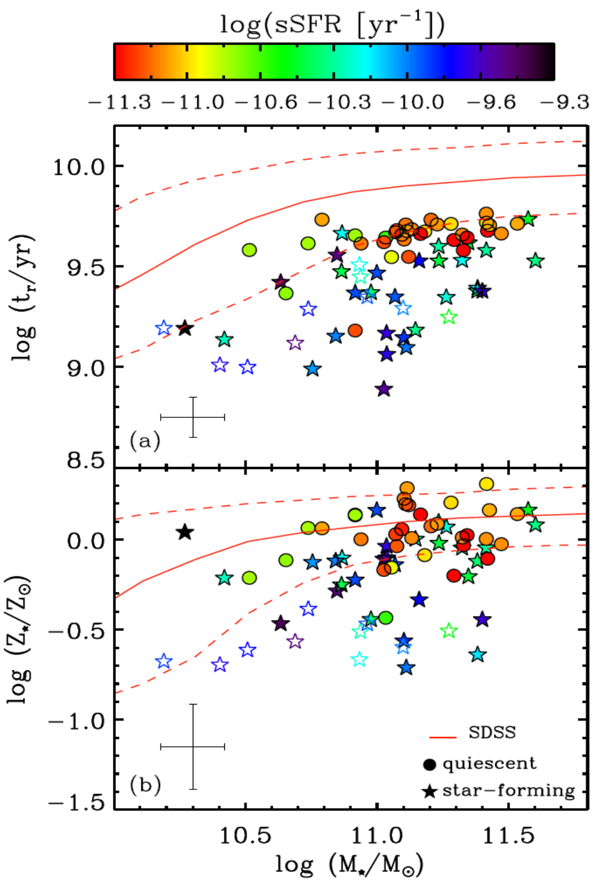
Utilizing the resource allocation we obtained in Proposal I, we performed extensive statistical analysis of modeling half (~1600 galaxies) of these high-resolution LEGA-C spectra along with their photometric measurements to obtain an optimized set of hyperparameters and fitting parameters using state-of-the-art python package called BAGPIPES (Bayesian Analysis of Galaxies for Physical Inference and Parameter Estimation). This package heavily relies on the Bayesian inference tool and multimodal nested sampling algorithm ‘MultiNest’ that requires significant computation time for each galaxy.

Starting with analysis of a subset of 500 galaxies, we first masked certain emission line regions in the spectra that include contamination from physical processes not coming from star-formation and that do not contribute to the properties of the galaxies we are interested in, then using only a subset of photometric bands (total 10 bands - BVrizYJHK + mips) that have high confidence interval and decreasing the confidence range of the calibration parameter that goes into the fits since now we are using a finely calibrated spectra which was obtained by members of our collaboration last year in November. With all these modifications, we found that on the MPI cluster **each galaxy model is now taking ~120 CPU hours** and we are saving about 300 CPU hours per galaxy model compared to old parameterizations. We found that on an average, a complete spectroscopic + photometric fit of 1 object on MPI cluster with 28 cores takes about ~120 CPU hours (average calculated for all 1600 runs on node=8) which is reasonable for a total likelihood evaluations of about ~2 million, 500 posterior points and 700 live sampling points from the posterior. Using the scaling factor of 1.0 to convert the CPU hours to SUs for the **MPI cluster** and 1-sigma variation in MPI process, we found that we need an upper limit of about **150 SUs per object for one analytical star-formation history (double-power law)**. The optimum allocation for 3200 objects would be 480,000 SUs on MPI that can fit all objects with one analytic star-formation history. We are also interested in analysing the results from more star-formation histories (exponential) for all of them and only some objects with lognormal star-formation history. As the parameter space is approximately the same for all of them, we expect to require 2 times of the above i.e. 2 x 480,000 = 960,000 SUs on the MPI cluster. Including the margin computation resources required for test runs and reruns with different parameter spaces on selected objects in the catalog, we expect to use about **1,000,000 SUs allocation on the MPI cluster**.

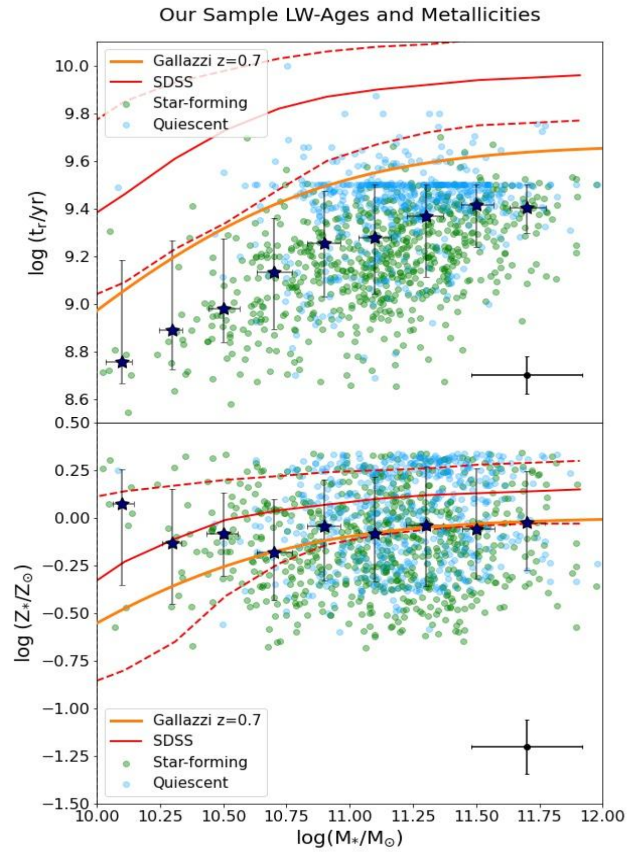
The photometry only models, that are much faster on SMP single node processors and have fewer data points to fit, require about ~0.64 CPU hours per object for exponential star formation history and ~0.95 CPU hours per object with double-power law star-formation history on the SMP cluster. For the entire catalog, we estimate about 0.64x3200=2048 CPU hours and 0.95x3200=3040 CPU hours and hence we will need about 100,000 CPU hours on the SMP cluster for 5 different star formation histories and allowing more parameters in the fit. That sums up to about **80,000 SUs on the SMP cluster with a scale factor of 0.8**.

As the spectroscopic resolution and wavelength ranges are approximately the same for all the objects in the catalog, in computation terms, it is expected to not have drastic variation in computation time for the entire sample, but the variations in the signal to noise ratio of each spectrum and the convergence of the sampling will produce some scatter as shown in Figure 1 Proposal I.

Given limited nodes availability per research group per script and from Proposal I analysis of node configuration and ideal speedup parallelization, we adopted 7 nodes of 28 cores each (7x28 = 196 cores) from the MPI cluster OPA partition to fit models to about 100 galaxies in the normal QOS stream with default wall time of 48 hours. By submitting 4 such scripts parallely, we were using the full allocation of ~ 900 cores for normal QOS and were able to obtain models for ~350 galaxies' spectra in 2 days. Without using the long QOS resources and including the queuing period after slurm script submission, it took about 12-14 days to obtain models for half of the LEGA-C dataset (~1600) for single analytic star-formation history prescription.
Some preliminary statistical results from this **1600 objects** analysis and comparison with previous studies with a much smaller sample (**<100 objects**) is shown in figures below.



<u>Fig 1.</u> Gallazzi+2014 study of ages and metal content within galaxies as a function of the mass of stars within it.



<u>Fig 2</u>. Our analysis of ages and metal content for two types of galaxies - quiescent and star-forming and comparison with previous studies.

Funding Sources

- Title: 'Dancing of the Stars: Testing the Formation of the Largest Elliptical Galaxies'
- Duration : September, 2019 - August, 2021 (no cost extension to August, 2022 to be submitted)
- Amount : $150,000
- PI/co-PI : R. Bezanson
- One to two sentence description : Probing the star formation histories and stellar kinematics in galaxies at half the age of the Universe.
- Agency or other entity : The Kaufman Foundation (New Investigator Award)

Involvement of CRC Consultants

All consultants have been dedicatedly involved in this project by addressing issues in the computational errors via help tickets. We are especially thankful to Kim Wong, Shervin and Barry Moore for their timely responses on queries, getting codes working on the cluster and giving suggestions for the most optimized way of running a big batch of jobs efficiently.

Publications/Talks/Seminars acknowledging use of CRC (or SaM) resources over the past year

We have made great progress since our first proposal approval on June 6, 2020. I (Yasha Kaushal, III year graduate student) along with Dr. Bezanson have presented these initial findings in following places -

1. University of Iowa, Seminar Talk, Stellar Population Properties and Star formation Histories of Galaxies in 6 billion years old Universe (17 March 2021)

2. University of California, Santa Cruz, No Jargon Talk, Peeping into the Past of Galaxies (22 April 2021)

3. University of Pittsburgh, 3-Minute-Thesis, Understanding the Evolution of Galaxies from 6 billion years old Universe: Connecting Past to Present (11 April 2021)

4. University of Pittsburgh, ARC-2021 Poster Competition, Peeping Into The Past of Galaxies in 6 Billion Years Old Universe (April 6, 2021)

5. Space Telescope Science Institute Workshop, "Multi-object Spectroscopy for Statistical Measures of Galaxy Evolution", (May 2021)

- Invited Talk, *Rachel Bezanson*
- Poster Presentation, *Yasha Kaushal*

6. INAF - Arcetri Colloquium, *Rachel Bezanson*, (December 3, 2020)

7. University of California-Los Angeles, Astronomy Colloquium, *Rachel Bezanson, (*October 14, 2020)

The LEGA-C data release DR2 came out in 2018 and so far many papers have been published on it from different collaborators. I mention below the ones for which Dr. Bezanson is the first author. We are implementing this computation intensive bayesian technique of spectral analysis for the first time and hence have no previous record of using CRC resources.

1. The Fundamental Plane in the LEGA-C Survey: unraveling the M/L variations of massive star-forming and quiescent galaxies at z ~ 0.8 by Anna De Graff, Rachel Bezanson et. al. 2021 (https://arxiv.org/pdf/2103.12753.pdf)
2. Stellar Kinematics and Environment at z ~ 0.8 in the LEGA-C Survey: Massive, Slow-Rotators are Built First in Overdense Environments by Justin Cole, Rachel Bezanson et. al. 2020 (https://arxiv.org/pdf/2001.02695.pdf)
3. SPATIALLY RESOLVED STELLAR KINEMATICS FROM LEGA-C: INCREASED ROTATIONAL SUPPORT IN Z~0.8 QUIESCENT GALAXIES by Bezanson et. al. 2018 (https://arxiv.org/pdf/1804.02402.pdf)
4. 1D Kinematics from stars and ionized gas at z ~ 0.8 from the LEGA-C spectroscopic survey of massive galaxies by Bezanson et. al. 2018 (https://arxiv.org/pdf/1811.07900.pdf)