# Large Synoptic Survey Telescope Galaxies, Dark Matter, and Black Holes: Extragalactic Roadmap

Robertson, Brant[1], Banerji, M.[2], Cooper, Michael[3], Davies, R.[4], Ferguson, Henry C.[5], Kaviraj, S.[6], Lintott, C.[4], Lotz, J.[5], Newman, J.[7], Norman, D.[8], Padilla, N.[9], Schmidt, S.[10], Verma, A.[4], Working Group Participants, Collaboration Members

[1]*University of California, Santa Cruz,* [2]*Cambridge University,* [3]*University of California, Irvine,* [4]*Oxford University,* [5]*Space Telescope Science Institute,* [6]*University of Hertfordshire,* [7]*University of Pittsburgh,* [8]*National Optical Astronomical Observatory,* [9]*Pontifica Universidad Catolica de Chile,* [10]*University of California, Davis,*

Version March 21, 2017

The LSST Extragalactic Roadmap represents the collective efforts of more than one hundred scientists to define the critical research activities to prepare our field to maximize the science return of the LSST dataset. We want to thank the LSST Corporation for their support in developing this Roadmap and for supporting LSST-related science more broadly. We also wish to thank the LSST Galaxies Science Collaboration members for their efforts over the years in developing the case for extragalactic science with LSST. Lastly, we wish to thank Harry Ferguson for his continued efforts to organize this document.

Inquiries about this report or its content can be addressed to Brant Robertson (brant@ucsc.edu).

**Abstract**

The Large Synoptic Survey Telescope (LSST) will enable revolutionary studies of galaxies, dark matter, and black holes over cosmic time. The LSST Galaxies Science Collaboration (LSST GSC) has identified a host of preparatory research tasks required to leverage fully the LSST dataset for extragalactic science beyond the study of dark energy. This *Extragalactic Roadmap* provides a brief introduction to critical extragalactic science to be conducted ahead of LSST operations, and a detailed list of preparatory science tasks including the motivation, activities, and deliverables associated with each. The *Extragalactic Roadmap* will serve as a guiding document for researchers interested in conducting extragalactic science in anticipation of the forthcoming LSST era.

# Contents

# Chapter 1

# Introduction

The Large Synoptic Survey Telescope (LSST) is a wide-field, ground-based telescope, designed to image a substantial fraction of the sky in six optical bands every few nights. It is planned to operate for a decade allowing the stacked images to detect galaxies to redshifts well beyond unity. The LSST and the survey are designed to meet the requirements (Ivezic & the LSST Science Collaboration 2011) of a broad range of science goals in astronomy, astrophysics and cosmology. The LSST was the top-ranked large ground-based initiative in the 2010 National Academy of Sciences decadal survey in astronomy and astrophysics, and is on track to begin the survey early in the next decade.

In 2008, eleven separate quasi-independent science collaborations were formed to focus on a broad range of topics in astronomy and cosmology that the LSST could address. Members of these collaborations have been instrumental in helping to develop the science case for LSST (encapsulated in the LSST Science Book), to refine the concepts for the survey and for the data processing, and to educate other scientists and the public about the promise of this unique observatory.

The Dark Energy Science Collaboration (DESC) has taken the next logical step beyond the science book. They identified they most critical challenges that will need to be overcome to realize LSSTs potential for measuring the effects of Dark Energy. They looked at five complementary techniques for tackling dark energy, and outlined high-priority tasks for the science collaboration during construction. They designated sixteen working groups (some of which already existed) to coordinate the work. This roadmap has been documented in a 133-page white paper (arxiv.org/abs/1211.0310). The white paper provides a guide for investigators looking for ways to contribute to the overall investigation. It may help in efforts to obtain funding, because it provides clear indication of the importance of the advance work and how the pieces fit together.

The investigation of Dark Energy is only one topic for LSST. It is important to develop similarly concrete roadmaps for work in other areas. After some discussion among the collaborations, it appears useful in some cases for different science collaborations to join forces on a single whitepaper. This is particularly true for topics that involve observations of distant galaxies. With the advent of the DESC, some of the science goals of the large-scale-structure, weak-lensing, and strong-lensing collaborations have found a new home. The remaining science goals of those collaborations tend to be focused on galaxy evolution and dark matter. Two other collaborations: AGN and Galaxies, also have those topics as major themes. This roadmap identifies the major high-level science themes of these investigations, outlines how complementary techniques will contribute, and identifies areas where advance work is essential. For this advance work, the emphasis is on areas that are not adequately covered in the DESC roadmap.

Chapter 2 gives a brief summary of the science background. Many of the themes and projects are already set out in the Science Book, where more detail is provided for many of the science investigations. Chapter 3 presents preparatory science tasks for Extragalactic science with LSST. These tasks are organized by science topic. The science task list content assumes that the work plan of the DESC will be executed and that the resulting software and other data products resulting from the DESC efforts will be made available to the other science collaborations.

# Chapter 2

# Science Background

## 2.1 Galaxy Evolution Studies with LSST

Galaxies represent fundamental astronomical objects outside our own Milky Way. The large luminosities of galaxies enable their detection to extreme distances, providing abundant and far-reaching probes into the depths of the universe. At each epoch in cosmological history, the color and brightness distributions of the galaxy population reveal how stellar populations form with time and as a function of galaxy mass. The progressive mix of disk and spheroidal morphological components of galaxies communicate the relative importance of energy dissipation and collisionless processes for their formation. Correlations between internal galaxy properties and cosmic environments indicate the ways the universe nurtures galaxies as they form. The evolution of the detailed characteristics of galaxies over cosmic time reflects how fundamental astrophysics operates to generate the rich variety of astronomical structures observed today.

Study of the astrophysics of galaxy formation represents a vital science of its own, but the ready observability of galaxies critically enables a host of astronomical experiments in other fields. Galaxies act as the semaphores of the universe, encoding information about the development of large scale structures and the mass-energy budget of the universe in their spatial distribution. The mass distribution and clustering of galaxies reflect essential properties of dark matter, including potential constraints on the velocity and mass of particle candidates. Galaxies famously host supermassive black holes, and observations of active galactic nuclei provide a window into the high-energy astrophysics of black hole accretion processes. The porous interface between the astrophysics of black holes, galaxies, and dark matter structures allows for astronomers to achieve gains in each field using the same datasets.

The Large Synoptic Survey Telescope (LSST) will provide a digital image of the southern sky in six bands (*ugrizy*). The area ($\sim 18,000$ deg$^2$) and depth ($r \sim 24.5$ for a single visit, $r \sim 27.5$ coadded) of the survey will enable research of such breadth that LSST may influence essentially all extragalactic science programs that rely primarily on photometric data. For studies of galaxies, LSST provide both an unequaled catalogue of billions of extragalactic sources and high-quality multiband imaging of individual objects. This section of the *Extragalactic Roadmap* presents scientific background for studies of these galaxies with LSST to provide a context for considering how the astronomical community can best leverage the catalogue and imaging datasets and for identifying any required preparatory science tasks.

LSST will begin science operations during the next decade, more than twenty years after the start of the Sloan Digital Sky Survey (York et al., 2000) and subsequent precursor surveys including

PanSTARRS (Kaiser et al., 2010), the Subaru survey with Hyper Suprime-Cam (Miyazaki et al., 2012), and the Dark Energy Survey (Flaugher, 2005). Relative to these prior efforts, extragalactic science breakthroughs generated by LSST will likely benefit from its increased area, source counts, and statistical samples, the constraining power of the six-band imaging, and the survey depth and image quality. The following discussion of LSST efforts focusing on the astrophysics of galaxies will highlight how these features of the survey enable new science programs.

### 2.1.1 Star Formation and Stellar Populations in Galaxies

Light emitted by stellar populations will provide all the direct measurements made by LSST. This information will be filtered through the six passbands utilized by the survey (*ugrizy*), providing constraints on the rest-frame ultraviolet SEDs of galaxies to redshift $z \sim 6$ and a probe of rest-frame optical spectral breaks to $z \sim 1.5$. By using stellar population synthesis modeling, these measures of galaxy SEDS will enable estimates of the redshifts, star formation rates, stellar masses, dust content, and population ages for potentially billions of galaxies. In the context of previous extragalactic surveys, LSST will enable new advances in our understanding of stellar populations in galaxies by contributing previously unachievable statistical power and an areal coverage that samples the rarest cosmic environments.

A variety of ground- and space-based observations have constrained the star formation history of the universe over the redshift range that LSST will likely probe (for a recent review, see Madau & Dickinson, 2014). The statistical power of LSST will improve our knowledge of the evolving UV luminosity function, luminosity density, and cosmic star formation rate. The LSST observations can constrain how the astrophysics of gas cooling within dark matter halos, the efficiency of molecular cloud formation and the star formation within them, and regulatory mechanisms like supernova and radiative heating give rise to these statistical features of the galaxy population. While measurement of the evolving UV luminosity function can help quantify the role of these astrophysical processes, the ability of LSST to probe vastly different cosmic environments will also allow for the robust quantification of any changes in the UV luminosity function with environmental density, and an examination of connections between environment and the fueling of star formation.

Optical observations teach us about the established stellar content of galaxies. For stellar populations older than $\sim 100$ million years, optical observations provide sensitivity to the spectral breaks near a wavelength of $\lambda \approx 4000\AA$ in the rest-frame related to absorption in the atmospheres of mature stars. Such observations help constrain the amount of stellar mass in galaxies. For passive galaxies that lack vigorous star formation, these optical observations reveal the well-defined "red sequence" of galaxies in the color-magnitude plane that traces the succession of galaxies from recently-merged spheroids to the most massive systems at the centers of galaxy clusters. For blue, star-forming galaxies, optical light can help quantify the relative contribution of evolved stars to total galaxy luminosity, and indeed has led to the identification of a well-defined locus of galaxies in the parameter space of star formation rate and stellar mass (e.g., Noeske et al., 2007). This relation, often called the "star-forming main sequence" of galaxies, indicates that galaxies of the same stellar mass typically sustain a similar star-formation rate. Determining the physical or possibly statistical origin of the relation remains an active line of inquiry, guided by recently improved data from Hubble Space Telescope over the $\sim 0.2$ deg$^{-2}$ Cosmic Assembly Near-Infrared Deep Extragalactic Survey (Grogin et al., 2011; Koekemoer et al., 2011). While LSST will be comparably limited in redshift selection, its $30,000$ times larger area will enable a much fuller sampling of the star formation–stellar mass plane, allowing for a characterization of the distribution of galaxies that lie off the main sequence that can help discriminate between phenomenological explanations of the sequence.

### 2.1.2 Galaxies as Cosmic Structures

The structural properties of galaxies arise from an intricate combination of important astrophysical processes. The gaseous disks of galaxies require substantial energy dissipation while depositing angular momentum into a rotating structure. These gaseous disks form stars with a surface density that declines exponentially with galactic radius, populating stellar orbits that differentially rotate about the galactic center and somehow organize into spiral features. Many disk galaxies contain (pseduo-)bulges that form through a combination of violent relaxation and orbital dynamics. These disk galaxy features contrast with systems where spheroidal stellar distributions dominate the galactic structure. Massive ellipticals form through galaxy mergers and accretions, and manage to forge a regular sequence of surface density, size, and stellar velocity dispersion from the chaos of strong gravitational encounters. Since these astrophysical processes may operate with great variety as a function of galaxy mass and cosmic environment, LSST will revolutionize studies of evolving galaxy morphologies by providing enormous samples with deep imaging of exquisite quality.

The huge sample of galaxies provided by LSST will provide a definitive view of how the sizes and structural parameters of disk and spheroidal systems vary with color, stellar mass, and luminosity. Morphological studies will employ at least two complementary techniques for quantifying the structural properties of galaxies. Bayesian methods can yield multi-component parameterized models for all the galaxies in the LSST sample, including the quantified contribution of bulge, disk, and spheroid structures to the observed galaxy surface brightness profiles. The parameterized models will supplement non-parametric measures of the light distribution including the Gini and M20 metrics that quantify the surface brightness uniformity and spatial moment of dominant pixels in a galaxy image (Abraham et al., 2003; Lotz et al., 2004). Collectively, these morphological measures provided by analyzing the LSST imaging data will enable a consummate determination of the relation between structural properties and other features of galaxies over a range of galaxy mass and luminosity previously unattainable.

While the size of the LSST sample supplies the statistical power for definitive morphological studies, the sample size also enables the identification of rare objects. This capability will benefit our efforts for connecting the distribution of galaxy morphologies to their evolutionary origin during the structure formation process, including the formation of disk galaxies. The emergence of ordered disk galaxies remains a hallmark event in cosmic history, with so-called "grand design" spirals like the Milky Way forming dynamically cold, thin disks in the last $\sim 10$ Gyr. Before thin disks emerged, rotating systems featured "clumpy" mass distributions with density enhancements that may originate from large scale gravitational instability. Whether the ground-based LSST can effectively probe the exact timing and duration of the transition from clumpy to well-ordered disks remains unknown, but LSST can undoubtedly contribute studying the variation in forming disk structures at the present day. Unusual objects, such as the UV luminous local galaxies identified by Heckman et al. (2005) that display physical features analogous to Lyman break galaxies at higher redshifts, may provide a means to study the formation of disks in the present day under rare conditions only well-probed by the sheer size of the LSST survey.

Similarly, the characterizing the extremes of the massive spheroid population can critically inform theoretical models for their formation. For instance, the most massive galaxies at the centers of galaxy clusters contain vast numbers of stars within enormous stellar envelopes. The definitive LSST sample can capture enough of the most massive, rare clusters to quantify the spatial extent of these galaxies at low surface brightnesses, where the bound stellar structures blend with the intracluster light of their hosts. Another research area the LSST data can help address regards the central densities of local ellipticals that have seemingly decreased compared with field ellipticals at higher redshifts. The transformation of these dense, early ellipticals to the spheroids in the present day may involve galaxy mergers and environmental effects, two astrophysical processes that LSST can characterize through unparalleled statistics and environmental probes. By measuring the surface brightness profiles of billions of ellipticals LSST can determine whether any such dense early ellipticals survive to the present day, whatever their rarity.

Beyond the statistical advances enabled by LSST and the wide variation in environments probed by a survey of half the sky, the image quality of LSST will permit studies of galaxy structures in the very low surface brightness regime. Observational measures of the outer most regions of thin disks can constrain how such disks "end", how dynamical effects might truncate disks, and whether some disks smoothly transition into stellar halos. LSST will provide such measures and help quantify the relative importance the physical effects that influence the low surface brightness regions in disks. Other galaxies have low surface brightnesses throughout their stellar structures, and the image quality and sensitivity of LSST will enable the most complete census of low surface brightness galaxies to date. LSST will provide the best available constraints on the extremes of disk surface brightness, which relates to the extremes of star formation in low surface density environments.

The ability of LSST to probe low surface brightnesses also allows for characterization of stellar halos that surround nearby galaxies. Structures in stellar halos, from streams to density inhomogeneities, originate from the hierarchical formation process and their morphology provides clues to the formation history on a galaxy-by-galaxy basis (Bullock & Johnston, 2005; Johnston et al., 2008). Observations with small telescopes (Martínez-Delgado et al., 2008; Abraham & van Dokkum, 2014) have already demonstrated that stellar halo structures display interesting variety (e.g., van Dokkum et al., 2014). LSST, with its unrivaled entendue, can help build a statistical sample of stellar halos and cross-reference their morphologies with the observed properties of their central galaxies. Such studies may determine whether the formation histories reflected in the structures of halos also influence galaxy colors or morphological type. The examination of stellar halos around external galaxies may also result in the identification of small mass satellites whose sizes, luminosities, and abundances can constrain models of the galaxy formation process on the extreme low-mass end of the mass function.

### 2.1.3 Probing the Extremes of Galaxy Formation

The deep, multiband imaging LSST provides over an enormous area will enable the search for galaxies that form in the rarest environments, under the most unusual conditions, and at very early times. By probing the extremes of galaxy formation, the LSST data will push our theoretical understanding of the structure formation process.

The rarest, most massive early galaxies may form in conjunction with the supermassive black holes that power distant quasars. LSST can use the same types of color-color selections to identify extremely luminosity galaxies out to redshift $z \sim 6$, and monitor whether the stellar mass build-up in these galaxies tracks the accretion history of the most massive supermassive black holes. If stellar mass builds proportionally to black hole mass in quasars, then very rare luminous star forming galaxies at early times may immediately proceed the formation of bright quasars. LSST has all the requisite survey properties (area, mutliband imaging, and depth) to investigate this long-standing problem.

The creation of LSST Deep Drilling fields will enable a measurement of the very bright end of the high-redshift galaxy luminosity function. Independent determinations of the distribution of galaxy luminosities at $z \sim 6$ show substantial variations at the bright end. The origin of the discrepancies between various groups remains unclear, but the substantial cosmic variance expected for the limited volumes probed and the intrinsic rarity of the bright objects may conspire to introduce large potential differences between the abundance of massive galaxies in different areas of the sky. Reducing this uncertainty requires deep imaging over a wide area, and the LSST Deep Drilling fields satisfy this need by achieving sensitivities beyond the rest of the survey.

Lastly, the spatial rarity of extreme objects discovered in the wide LSST area may reflect an intrinsically small volumetric density of objects or the short duration of an event that gives rise to the observed properties of the rare objects. Mergers represent a critical class of short-lived epochs in the formation histories of individual galaxies. Current determinations of the evolving numbers of close galaxy pairs or morphological indicators of mergers provide varying estimates for the redshift dependence of the galaxy merger rate (e.g., Conselice et al., 2003; Kartaltepe et al., 2007; Lotz et al., 2008; Lin et al., 2008; Robotham et al., 2014). The identification of merging galaxy pairs as a function of separation, merger mass ratio, and environment in the LSST data will enable a full accounting of how galaxy mergers influence the observed properties of galaxies as a function of cosmic time.

### 2.1.4 Science Book

The LSST Science Book (LSST Science Collaboration et al. 2009) provided detailed descriptions of foundationl science enabled by LSST. The LSST Galaxies Science Collaboration authored the Chapter 9 "Galaxies" of the Science Book, and the a table of contents of that chapter follow below to provide an example list of topics in extragalactic science that LSST data will help revolutionize. The interested reader is referred to the LSST Science Book for more details.

1. Measurements, Detection, Photometry, Morphology
2. Demographics of Galaxy Populations
    - Passively evolving galaxies
    - High-redshift star forming galaxies
    - Dwarf galaxies
    - Mergers and interactions
3. Distribution Functions and Scaling Relations
    - Luminosity and size evolution
    - Relations between observables
    - Quantifying the Biases and Uncertainties
4. Galaxies in their Dark-Matter Context
    - Measuring Galaxy Environments with LSST
    - The Galaxy-Halo Connection
    - Clusters and Cluster Galaxy Evolution
    - Probing Galaxy Evolution with Clustering Measurements
    - Measuring Angular Correlations with LSST, Cross-correlations
5. Galaxies at Extremely Low Surface Brightness
    - Spiral Galaxies with LSB Disks
    - Dwarf Galaxies
    - Tidal Tails and Streams
    - Intracluster Light
6. Wide Area, Multiband Searches for High-Redshift Galaxies
7. Deep Drilling Fields
8. Galaxy Mergers and Merger Rates

9. Special Populations of Galaxies
10. Public Involvement

# Chapter 3

# Task Lists by Science Area

## 3.1 Active Galactic Nuclei

AGN are phenomena that enable us to understand the growth of BHs, understand aspects of galaxy evolution, probe the high redshift universe and study other physical activity, including accretion physics, jets, magnetic fields, etc. There are distinct aspects of the study of AGN that can best be explored by considering AGN as an evolutionary stage of galaxies rather than a distinct type of source. The tasks listed here explore aspects of AGN study that are particularly important AGN as a stage in galaxy evolution.

### 3.1.1 AGN feedback in clusters

**AGN-1.** *Motivation:* Brightest Cluster/Group Galaxies (hereafter BCGs) are the most massive galaxies in the local Universe residing at/near the centres of galaxy clusters/groups. They will therefore contain the largest supermassive black holes. These black holes can influence their host BCG, the cluster gas and other cluster members via the mechanical energy produced by their 100s kpc scale jets (AGN feedback).

For low redshift galaxy clusters it is possible to perform detailed studies of the star, gas and AGN jets to analyse the details of AGN feedback. LSST will provide a large sample of moderate to high redshift clusters in which we can measure AGN feedback statistically. By combining X-ray, radio and optical observations we can assess the average influence of the BCG's AGN on the hot Intra-cluster medium (ICM) for different sub-populations [e.g. Stott et al. 2012].

*Activities:* By assembling a multi-wavelength dataset (optical, X-ray, Radio) we can obtain the BCG mass, cluster mass and ICM temperature, and the mechanical power injected into the ICM. We can use this to study the interplay between the BCG, its black hole and the cluster gas, to assess the balance of energies involved and for direct comparison with theoretical models of AGN feedback. This has been done with a few hundred clusters at z¡0.3 using SDSS but we may well be able to reach z=1 and therefore look for an evolution in their interplay and therefore AGN feedback. There are also implications for cosmology too as this will help with the selection of clusters for which the X-ray properties better represent the mass of the cluster rather than the complex interplay of baryonic physics.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Investigate the number of BCGs and the mass range of their clusters with redshift that LSST is likely to be able to observe.
(b) Assess radio and X-ray data available for AGN Feedback studies (XCS, eROSITA, SKA-pathfinders, SUMSS etc).
(c) Assess the theoretical predictions expected for the above (e.g. cosmological simulations such as EAGLE or more detailed single cluster studies).

### 3.1.2 AGN Selection from LSST Data

**AGN-2.** *Motivation:* Active Galactic Nuclei are selected using a variety of different methods. At optical and infrared wavelengths, photometric selection of AGN candidates is driven by their distinctive colors at particular redshifts. X-ray and radio observations can also be efficient selectors of candidates for additional follow-up. With spectral data, AGN can be selected using the ratios of their emission lines. LSST will also open up, in a more practical way, the identification of AGN based on their variability. Each of these samples probes aspects of the AGN phenomena and a better understanding of the AGN role in galaxy evolution requires that we understand how and why each of these selection methods includes or excludes particular sources. Furthermore, currently each of these methods for identifying AGN candidates requires spectral follow-up to cull these samples to positively identify the most reliably clean AGN sample.

*Activities:* For us to use LSST as a single way to identify the diversity of AGN, we must develop selection criteria that take advantage of the source parameters available with just LSST imaging, that is, color, morphology and variability. Already there are a number of AGN surveys with input from multiple wavelength observation and spectra. Precursor work needs to be done using these surveys to determine if AGN not easily identified using optical color selection can be selected using the additional parameters of morphology, variability and/or the additional filter that LSST provides.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Cross-matched catalog of known AGN selected and verified using different methods
(b) Development of morphology parameters beyond just star/galaxy separation and an understanding of the morphology parameters to be provided by LSST level 2 products.
(c) Development of color selection criteria that takes into account the morphology of the source
(d) Understanding of how AGN variability looks given the nominal LSST cadence
(e) Development of algorithms for color selection that take into account the variability of an AGN source

### 3.1.3 AGN Host Galaxy Properties from LSST Data

**AGN-3.** *Motivation:* We are requesting that basic morphological parameters (e.g., CAS, G-M20, etc.) be measured in the pipeline and made available as products to help in the identification of merging galaxies in LSST data. The issue here is how well this can be done when the host galaxies contain AGN that are likely identified via their variability. In other words, how well can we determine the host morphology of galaxies with variable AGN? This would be interesting for models of AGN fueling during mergers.

*Activities:* Simulations of the accuracy by which the pipeline (deblender) can measure the defined morphology parameters in host galaxies as a function of AGN brightness and wavelength. We could then "vary" the central source by expected levels in certain filters to see the effect on the morphological params. To constrain this it would be helpful to add in central sources with reasonable SEDs across the LSST bands, and a limited set of frequencies/amplitudes (based on real data - perhaps Pan-STARRS?).

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Plots of the accuracy of the measured basic morphology parameters as a function of AGN brightness and wavelength.

(b) Effect of AGN brightness on classification diagrams.

### 3.1.4 AGN Variability Selection in LSST Data

**AGN-4.** *Motivation:* Most AGN exhibit broad-band aperiodic, stochastic variability across the entire EM spectrum on timescales ranging from minutes to years. Continuum variability arises in the accretion disk of the AGN, making it a powerful probe of accretion physics. The main LSST WFD survey will obtain $\sim 10^8$ AGN light curves (i.e. flux as a function of time) with $\sim$ 1000 observations ($\sim$ 200 per filter band) over 10 years. The deep drilling fields will give us AGN lightcurves with much denser sampling for a small subset of the objects in the WFD survey. The science content of the lightcurves will critically depend on the exact sampling strategy used to obtain the light curves. For example, the observational uncertainty in determining the color variability of AGN will critically depend on the interval between observations in individual filter bands. It is of crucial importance to determine guidelines for an optimal survey strategy (from an AGN variability perspective) and determine what biases and uncertainties are introduced into AGN variability science as a result of the chosen survey strategy.

*Activities:* Study existing AGN variability datasets (SDSS Stripe 82, OGLE, PanSTARRS, CRTS, PTF + iPTF, Kepler, & K2) to constrain a comprehensive set of AGN variability models. Generate & study simulations using parameters selected from these models with the observationally determined constraints to determine goodness of simulations for carrying out various types of AGN variability science - PSD models, QPO searches, binary AGN models, etc.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Observational constraints on AGN variability models.

(b) MAF metrics quantifying the goodness of different survey strategies for AGN variability science.

### 3.1.5 AGN Photometric Redshifts from LSST Data

**AGN-5.** *Motivation:* Given the large number of AGN that will be observed with LSST, many of these will not be followed up with spectral observations. However, understanding the large scale structure of the universe, requires a 3-D understanding of the distribution of these galaxies in the universe. Photometric redshifts can provide relatively accurate redshifts for large numbers of galaxies. However, it is harder to obtain accurate photometric redshifts for galaxies that contain AGN compared to those that do not. We must understand how to get accurate photometric redshifts of galaxies with AGN.

*Activities:* An initial activity for this need to include comprehensive review of the state of the art in obtaining photo-zs for AGN host galaxy populations and how those compare to non-AGN galaxies. A comparison of model and/or observed AGN host SEDs with a matched set of non-host galaxies at a variety of redshifts will be used to determine color selection criteria for identifying AGN hosts. Explore whether variability can be used to break degeneracies.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Plots that show AGN host color selection criteria and where that color selection might become ambiguous (be degenerate) for non-host galaxies with different parameters.
(b) Plots that show if other parameters might break degeneracies.

### 3.1.6 AGN Merger Signature from LSST Data

**AGN-6.** *Motivation:* Understanding the role AGN play in galaxy evolution requires identifying the phenomenon at all stages and in all types of galaxies. AGN host galaxies are often found to be disturbed suggesting that the galaxy merger process is an important trigger of AGN activity. While the trainwrecks may be easier to find, galaxies in other merger stages can be difficult to identify and those experiencing pre-merger harassment may be particularly hard to recognize. Preliminary work needs to be done to understand how to identify mergers from the LSST data products and whether galaxy deblending and segmentation methods and procedures are adequate or mask galaxy mergers.

*Activities:* Create or Identify simulated and real images that contain known galaxy mergers, these images should contain mergers with and without AGN. Run LSST detection and identification software on these images. Identify metrics that describe/quantify the accurate detection of galaxy mergers (with and without AGN).

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Give feedback to LSST software teams about metrics and detection of galaxy mergers
(b) Give feedback on structure or galaxy type that do and do not work well with current versions of LSST software

## 3.2 Clusters and Large Scale Structure

The cosmological process of galaxy formation inextricably links the together, environment and large scale structure with the detailed properties of galaxy populations. The extent of this connection ranges from the scales of superclusters down to small groups. The following preparatory science tasks focus on this critical connection between galaxy formation, clusters, and large scale structure.

### 3.2.1 Cluster/LSS Sample Emulator

**CLSS-1.** *Motivation:* To prepare for galaxies and galaxy group/cluster science with LSST, we need to know how many galaxies will be detected in a given range of redshift, brightness, color, etc., and likewise how many groups and clusters will be detected in given ranges of redshift, richness, mass, and other physical parameters.

*Activities:* LSST has advanced simulations of its 10-year Wide Fast Deep survey available from the Operations Simulator. The output databases can be analyzed to determine the depth LSST is expected to reach in its final detection image at each sky location, and Awan et al. 2016 (http://adsabs.harvard.edu/abs/2016ApJ...82950A) turns these depths into predicted numbers of galaxies as a function of redshift and brightness.

To predict galaxy sample sizes as a function of physical parameters, the "raw" predicted galaxy numbers from Awan et al. (2016) will be interfaced with semi-analytical models painted on large N-body simulations by Risa Wechsler and collaborators. This will extend the predictions to include observed properties of color, size, morphology and physical properties of halo mass, stellar mass, and star formation rate.

To predict group/cluster sample sizes as a function of physical parameters, the properties such as temperature, richness, etc., will be painted on to dark matter halos drawn from a numerical simulation. The properties will be based on simple scaling laws, with the user allowed freedom to choose the parameters of the scaling laws, including how they evolve. This will then be interfaced with the "raw" predicted galaxy numbers from Awan et al. (2016), to determine which of the groups and clusters should be detectable in the LSST data.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Create a public LSST Extragalactic Sample Emulator with a simple GUI. Enable user input of a range of redshift, and physical parameters (e.g. galaxy magnitudes, colors, size, morphology, cluster richness, mass, temperature, etc.) to estimate the size of a given sample detected by LSST.

### 3.2.2 Identifying and Characterizing Clusters

**CLSS-2.** *Motivation:* LSST photometry will make it possible to search for and study the galaxy populations of distant clusters and proto-clusters over huge volumes of the high-$z$ Universe. These clusters are testbeds for cosmology, hierarchical structure formation, intergalactic medium heating and metal enrichment, as well as laboratories for studying galaxy evolution. However, standard approaches for identifying clusters, such as the red sequence method, will be hampered by the limited wavelength coverage of LSST. For example, at $z \gtrsim 1.5$, near-IR photometry is required to identify systems with Balmer/4000Å breaks. To maximize cluster science with LSST, we must devise new techniques for cluster identification as well as incorporate complementary data from projects such as *Euclid*, *eROSITA*, etc.

*Activities:* Using existing imaging datasets and simulations, algorithms need to be developed and optimised to identify clusters at intermediate and high redshift within the LSST footprint. Specifically, this work should characterize the selection function, completeness, and contamination rate for different cluster identification algorithms. This requires realistic light-cone simulations spanning extremely large volumes, so as to capture significant numbers ($\gg 10,000$) of simulated galaxy clusters at high $z$. Potential algorithms to be tested include adaptations of RedMaPPer (Rykoff et al. 2014) as well as methods that search for galaxy overdensities over a range of scales (e.g. Chiang et al. 2014; Wang et al. 2016). In parallel, a comprehensive search for multiwavelength data (specifically IR and X-ray imaging) is needed to aid in the search for high-$z$ clusters and in the confirmation and characterization of systems at all redshifts.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) The primary product of this analysis will be improved cluster identification algorithms that can be applied to LSST data once science operations commence.

(b) In addition, this work will produce a compilation of ancillary data that will be helpful in cluster identification and characterization, such as X-ray (e.g. XCS, eROSITA, etc.), SZ (Planck, SPT, ACT) and radio (SKA and its pathfinders, SUMSS), within the LSST footprint.

### 3.2.3 Developing and Optimizing Measurements of Galaxy Environment

**CLSS-3.** *Motivation:* Over the past decade, many studies have shown that “environment” plays a important role in shaping galaxy properties. For example, satellite galaxies in the local Universe exhibit lower star formation rates, more bulge-dominated morphologies, as well as older and more metal-rich stellar populations when compared to isolated (or “field”) systems of equivalent stellar mass (Baldry et al. 2006; Cooper et al. 2010; Pasquali et al. 2010). Unlike spectroscopic surveys, LSST will lack the precise line-of-sight velocity measurements to robustly identify satellite galaxies in lower-mass groups, where the expected photo-$z$ precision will greatly exceed the velocity dispersion of the host halo. Instead, LSST will likely be better suited to measuring environment by tracing the local galaxy density (and identifying filaments). However, LSST is unlike any previous photometric survey and may require new approaches to measuring environment. The challenge remains to find the measure(s) of local galaxy density with the greatest sensitivity to the true underlying density field (or to host halo mass, etc.), so as to enable analyses of environment’s role in galaxy evolution with LSST.

*Activities:* Using mock galaxy catalogs created via semi-analytic techniques, we will compare different tracers of local galaxy density (i.e. ”environment”) measured on mock LSST photometric samples to the underlying real-space density of galaxies (or to host halo mass). In addition to testing existing density measures, such as $N^{\rm th}$-nearest-neighbor distance and counts in a fixed aperture, we will explore new measures that may be better suited to LSST. For each measure, we will examine the impact of increasing survey depth and photo-$z$ precision over the course of the survey.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) With an improved understanding of the strengths and weaknesses of different environment measures as applied to LSST, this effort will yield code to measure local galaxy density (likely in multiple ways) within the LSST dataset.

(b) Create a Level 3 data product for use by the entire project.

### 3.2.4 Enabling and Optimizing Measurements of Galaxy Clustering

**CLSS-4.** *Motivation:* Contemporary galaxy surveys have transformed the study of large-scale structure, enabling high precision measurements of clustering statistics. The correlation function provides the most fundamental way to characterize the galaxy distribution. The dependence of clustering on galaxy properties and the evolution of clustering provide fundamental constraints on theories of galaxy formation and evolution. Interpreting these measurements provides crucial insight into the relation between galaxies and dark matter halos. Understanding how galaxies relate to the underlying dark matter is also essential for optimally utilizing the large-scale distribution of galaxies as a cosmological probe.

*Activities:* Preparatory work will be along two main tracks. The first one will be support work to define and characterize the upcoming galaxy samples from LSST to enable clustering measurements from them. Several distinct sets of information need to be made available or be calculable from pipeline data. Such requirements include a detailed understanding of any selection effects impacting the observed galaxies, the angular and radial completeness of the samples, and the detailed geometry of the survey (typically provided in terms of random catalogs that cover the full survey area).

The second track will be the development, testing, and optimization of algorithms for measuring galaxy clustering using LSST data. One aspect to address is how best to handle the large data sets involved (e.g. the "gold" galaxy sample will include about 4 billion galaxies over 20,000 square degrees). Another is to develop the methodology to optimally incorporate the LSST photo-$z$ estimates with the angular data to obtain "2.5-dimensions" for pristine clustering measurements.

These algorithms will be tested on realistic LSST mock catalogs, which will also later serve as a tool for obtaining error estimates on the measurements. This endeavor overlaps with DESC-LSS working group efforts, and requires cooperation of the DESC-PhotoZ working group and the Galaxies Theory and Mock Catalogs working group.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Ensuring LSST galaxy pipelines include all the necessary information for measurements of the correlation function and related statistics to take place once data is available

(b) Developing and refining techniques for measuring galaxy clustering of large LSST galaxy samples. Together these will enable realizing the full potential of LSST data for large-scale structure studies and galaxy formation inferences thereof.

### 3.2.5 Disentangling Complicated Lines of Sight

**CLSS-5.** *Motivation:* Lines of sight through galaxy clusters and groups are the most challenging lines of sight along which to measure reliable photometric redshifts because crowding of galaxies complicates the basic process of galaxy photometry, and the presence of significant correlated large-scale structure (LSS) complicates interpretation of the P(z) of the galaxies that has been computed by an algorithm that ignores the presence of the LSS. Numerous science goals require the most robust probabilistic statements possible as to the location of galaxies along lines of sight through clusters, for example, identification of background galaxies for weak-lensing, identification of faint cluster members to study the evolution of the luminosity function in clusters, identification of star-forming galaxies in clusters and their infall regions to probe the physics of quenching of star formation.

*Activities:* The LSST will deliver the most information rich dataset ever in relation to the masses and internal structures of clusters and their infall regions. Moreover, the dataset can be enhanced significantly via the addition of data at other wavelengths, including X-ray, millimeter, and near-infrared.

A tool is therefore envisaged, that can take an input catalogue of cluster centres that has been obtained from LSST or any other dataset (e.g. *Planck*, *eROSITA*). The tool will pull out the basic L2 LSST photometry of objects within a cone centred on the cluster centre, and compute the $p(z)$ of each galaxy based on a cluster-specific algorithm. This algorithm will take account of the following where they are available: brightness and extent of X-ray emission, over-density of galaxies as a function of magnitude and colour, any available spectroscopic redshifts, amplitude and extent of any SZ decrement/increment. The algorithm will likely adopt a Bayesian hierarchical modelling approach to forward model the problem. The algorithm can be tested on existing datasets from surveys such as the Local Cluster Substructure Survey (LoCuSS), XXL, HSC data processed by DM Stack within LSST, and any others that would like to join in.

This work has links with the work on deblending/ICL, forward modelling of cluster and groups, environmental measures, cluster detection, complementary data, and also work in the DESC Clusters WG via $p(z)$ of background galaxies.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) A new cluster-specific photometric redshift algorithm that can be applied to a list of cluster detections that is itself based on LSST or external data.

### 3.2.6 Forward Modeling LSST Clusters and Groups

**CLSS-6.** *Motivation:* Most of the interesting cluster and group physics from LSST and its union with complementary surveys will be derived from studies that explore the full range of halo mass relevant to groups and clusters: $M_{200} \simeq 10^{13} - 10^{15} M_{\odot}$. This is a wider range than cluster cosmologists (e.g. colleagues in DESC, with whom we collaborate) aim to incorporate into their cosmological inference – they restrict attention to $M_{200} > 10^{14} M_{\odot}$.

Another important difference between the cluster/group physics explored here, and the dark energy-motivated DESC work, is that the requirement on controlling systematic biases is roughly an order of magnitude less stringent here than in DESC. Arguably, $\sim 10\%$ control of systematic biases in weak-lensing measurements of low redshift clusters ($\gtrsim 2 \times 10^{14} M_{\odot}$)

has already been achieved (Okabe et al. 2013; Applegate et al. 2014; Hoekstra et al. 2015; Okabe & Smith 2016). Therefore in this Science Collaboration we have the challenge of maintaining that level of control down to smaller masses and out to higher redshifts.
A growing number of studies are adopting an approach of forward modelling the cluster population simultaneously with the cosmological model to obtain constraints on scaling relations and cosmological parameters. Here, the idea is to borrow this same approach, but adopt a fixed cosmological model, broaden the mass range of systems considered, and expand the forward modeling to include additional relationships of interest. For example, simultaneously fitting density profile models to the shear profiles, the mass-concentration relation, and the star-formation rates of clusters and groups. Overall, this will provide a robust Bayesian inference code with which to constrain the physics of galaxies and hot gas in groups and clusters, tied directly to the halo mass function via weak-lensing.

*Activities:* Key activities include:

- Select the elements of the cluster population to include in the model
- Write the first version of the code, and test on simulated (toy model and n-body) data
- Improve code and consider extending range of physics explored by adding more relations
- Test code on existing datasets from pointed surveys (e.g. LoCuSS, others) and wide area surveys (e.g. LoCuSS, DES, others)
- Combine this development work other work packages within Galaxies and DESC

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Bayesian inference code to simultaneously model cluster shear profiles, scaling relations (including and beyond cosmological scaling relations, across the full range of halo mass of groups and clusters, to $\sim 10\%$ control on systematics.

## 3.3 Deep Drilling Fields

The LSST Drilling-Fields (DDF) are areas that have a higher cadence and deeper observations than the Deep-Wide survey. Many of the details of the observing strategy have yet to be finalized. Four Deep-Drilling fields have been selected. Whether to include any others will be part of a complex trade involving other special projects that depart from the Deep-Wide survey strategy. The details of the observing cadence, final depth in each band, and dithering strategy are all still under study, and the Project needs input from the science collaborations to inform these decisions. The tasks outlined in this section are intended to help optimize the LSST observing strategy, gather supporting data, and ensure that the data processing and measurements meet the needs for galaxy-evolution science.

### 3.3.1 Coordinating Ancillary Observations

**DDF-1.** *Motivation:* It is crucial that the LSST deep-drilling fields be supported by observations from other facilities. While the LSST data by themselves will be unique in having deep and accurate photometry, good image quality, and time-series sampling, the amount of information in six bands of relatively broad optical imaging is quite limited. Estimates of photometric redshifts and stellar-population parameters (e.g. mass and star-formation rate) are greatly improved with long-wavelength data. Combining these quantities with information on dust and gas from far-IR, mm and radio observations allows one to build and test models that track the flow of gas in and out of galaxies. Deep and dense spectroscopy is essential both providing precise redshifts, calibrating photometric redshifts, and measuring physical properties of galaxies. Properly supported by this additional data, the LSST DDFs will become the most valuable areas of the sky for galaxy-evolution science. The central regions of the four fields already selected are already in this category; the main challenge is filling out the much larger area subtended by the LSST field of view.

*Activities:* The major challenge in supporting the Deep Drilling Fields is the huge investment of telescope time. There is a need for coordination across facilities and collaborations to make the most efficient use of this time. Coordination is certainly happening somewhat haphazardly, but there has not to date been a dedicated effort to get all the potential stakeholders involved in developing a coherent plan. The LSST science collaborations can and should be taking the lead here. The SERVS program to observe the already-designated DDFs with Spitzer is a good example of where this has happened (Manduit et al. 2012), but there is much more to be done. Activities include:

- Workshops to discuss LSST DDF coordination
- Proposals for major surveys or even new instrumentation to provide supporting data
- Executing those supporting programs
- Working to integrate the data from those programs with the LSST data
- Working to enable DDF support through policies and strategic planning at major observatories

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Workshops on LSST DDF supporting observations
(b) Annually updated roadmap of supporting observations (conceived, planned or executed)
(c) Public Release of data from supporting observations

(d) Level 3 software to enable use of LSST data with supporting data

### 3.3.2 Observing Strategy Cadence

**DDF-2.** *Motivation:* The LSST DDF observing strategy will need to serve diverse needs. For galaxy-evolution science, the time series aspect of the observation is less important than the depth, image quality, and mix of filters. Optimizing the observing strategy (including timing) is influenced by non-LSST factors like the availability of supporting data from other facilities, or the timing of the availability of such data. For example, for many science goals, completing the observations of one DDF to the final 10-year depth in the first year could be very beneficial. But there is work to be done to justify that, select the field, and find synergies with other science areas (e.g. DESC, AGN, transients).

*Activities:* The LSST observing strategy is optimized using the Operations Simulator (OpsSim). The Project works with the community to develop both strawman observing strategies and figures of merit for comparing different strategies. The figures of merit are implemented programmatically via the Metrics Analysis Framework (MAF) so that they can be easily applied to any candidate LSST cadence. The LSST project has called on the Science Collaborations to develop these metrics to codify their science priorities. The major activity here is involvement in the optimization of the DDF strategy through participation in Cadence workshops, training on the MAF and OpsSim, developing metrics and coding them in MAF, and proposing and helping to evaluate DDF cadences.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Figures of Merit via MAF for use by OpSim
(b) Proposed observing strategies for DDFs with rationale
(c) Proposing and/or helping to assess selection of additional DDFs

### 3.3.3 Data Processing

**DDF-3.** *Motivation:* Getting the most out of the DDFs may require data processing beyond that required for the Deep-Wide Survey. There are a variety of issues that ought to be considered in trying to optimize the science output. These include different strategies for making co-adds, determining sky levels, treating scattered light, detecting and characterizing faint or low-surface brightness features , deblending overlapping objects, or estimating photometric redshifts. The fields are small enough that it is conceivable to process or reprocess them making use of data from supporting observations. It will clearly be advantageous to have one official LSST-released catalog, but defining such a catalog to support a very broad range of science is challenging. This does not preclude having additional special-purpose catalogs, but it is clearly beneficial to the advancement of extragalactic research to have a high-quality official catalog that has buy in from the LSST Science Collaborations. This requires time and effort both in the Project and in the Collaborations.

*Activities:* A major activity here is to identify the most important DDF-specific science drivers and identify any processing requirements that are distinct from the Deep-Wide survey. This ought to be coordinated with the Project and the other Science Collaborations to provide a coherent set of specifications and priorities.

Another major activity is to develop the machinery to test and validate the data-processing on the DDFs (via pure simulations and artificial-source injection) This may stress the inputs to the image simulator, requiring more realistic inputs for low-mass galaxies, galaxy morphologies, and low-surface brightness features. Use of the supporting data sets in level 2 or level 3 processing requires careful thought. For example, source identification and photometry can be improved using pixel-level information for either Euclid or WFIRST. However, this will not be available for all the DDFs and is not in the baseline plan for any of the projects, and the timing of the various projects and associated data rights create their own set of challenges. The collaborations need to work with the various projects to identify a clear path forward.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Science drivers and input to the development of level 2 processing of the DDFs.
(b) Specifications for galaxy-evolution oriented level-3 DDF processing
(c) Specifications for data-processing using supporting data from other facilities
(d) Simulations tailored to the DDFs
(e) Level 3 processing code

## 3.4 Galaxy Evolution Task Lists

The LSST design, and to a certain extent the design of the data-management system, is optimized to carry out the core science mission. For measurements of dark-energy, that generally means treating galaxies as "tracer particles" – using statistical measures of ellipticity and position provide statistical constraints on large-scale structure and cosmic geometry. While many of the DESC tasks are directly relevant to studying galaxy evolution, they are incomplete. In particular, studies of galaxy evolution require more attention to optimizing multi-wavelength supporting data, different kinds of spectroscopy, different kinds of simulations and theoretical support, and greater attention to detection and characterization of low-surface brightness features or unusual morphologies.

The task list presented here highlights the preparation work needed in the next 3-4 years. Of primary importance are tasks that might influence the detailed survey design or the algorithms used in the DM to construct catalogs. These are the most urgent. Also included are activities that can be reasonably independent of the LSST survey design and DM optimization, but which will ensure good support for LSST galaxy studies.

### 3.4.1 Techniques, Algorithms, or Software Development

**G-TAS-1.** *Motivation:* A huge benefit of LSST relative to prior large-area surveys will be its ability to detect low-surface-brightness (LSB) features associated with galaxies. This includes tidal streams and other features associated with past and ongoing mergers, it includes intra-cluster and intra-group light, and it includes relatively nearby, extended low-surface-brightness galaxies. Prior to LSST, typical studies of the low-surface-brightness universe have focused on relatively small samples, often selected by criteria that are difficult to quantify or reproduce in theoretical models. Measurements of the LSB features themselves themselves are challenging, often requiring hand-tuning and interactive scientific judgment. This is important for accurately quantifying what we observe, but such interactive tuning of the measurements (a) is not something that can be applied on the LSST scale and (b) is difficult to apply to theoretical models. For LSST it is crucial that we automate the detection and characterization of LSB features, at least to the point where samples for further study can be selected via database queries, and where the completeness of samples returned from such queries can be quantified.

*Activities:* Several activities are of crucial importance: (1) simulating realistic LSB features, (2) using the simulations to optimize detection and measurement, (3) ensuring that LSST level-2 processing strategies and observing strategies are at least cognizant of needs of LSB science and (4) developing a strategy for finding and measuring LSB features through some combination of level 2 measurements, database queries, and level 3 processing.

It is important to insert realistic low-surface-brightness features into LSST simulated images and try to extract and measure them, exploring different techniques or algorithms for doing the detection and measurement. Because the LSB objects are sparse on the sky, making realistic LSST sky images is probably not the most efficient way to accomplish this; more targeted simulations with a higher density of LSB objects are needed. The simulated observations need to be realistic in their treatment of scattered light, particularly scattering from bright stars which may or may not be in the actual field of view of the telescope. Scattering from bright stars is likely to be the primary source of contamination when searching for extended LSB features. Ideally, the LSST scattered-light model, tuned by repeated observations, will be sufficiently good that these contaminants can be removed or at least flagged at level 2. Defining the metrics for "sufficiently good," based on analysis of simulations, is an important activity that needs early work to help inform LSST development. Including Galactic cirrus in the simulations is important for very large-scale LSB features. Including a cirrus model as part of the LSST background estimation is worth considering, but it is unclear yet whether the science benefit can justify the extra effort.

Because the LSST source extraction is primarily optimized for finding faint, barely-resolved galaxies, it is going to be challenging to optimize simultaneously for finding large LSB structures and cataloging them as one entity in the database. For very large structures, analysis of the LSST "sky background" map, might be the most productive approach. We need to work with the LSST project to make sure the background map is stored in a useful form, and that background measurements from repeated observations can be combined to separate the fluctuating foreground and scattered light from the astrophysically interesting signal from extended LSB structures. Then, we need strategies for measuring these background maps, characterizing structures, and developing value-added catalogs to supplement the level 2 database.

For smaller structures, it is likely that the database will contain pieces of the structure, either as portions of a hierachical family of deblended objects, or cataloged as separate objects. Therefore, we need to develop strategies for querying the database to find such structures and either extract the appropriate data for customized processing, or develop ways to put back together the separate entries in the database. A possible value-added catalog, for example, from the galaxies collaboration might be an extra set of fields for the database to indicate which separate objects are probably part of the same physical entity. This would be sparsely populated in the first year or two of LSST, but by the end of the survey could be a useful resource for a wide variety of investigations.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) realistic inputs of LSB galaxies or LSB features for the LSST image simulations;
(b) custom simulations;
(c) algorithms for finding and measuring LSB features;
(d) input to the Project on scattered-light mitigation and modeling strategies;
(e) input to the project on photometric and morphological parameters to measure/store at level 2;
(f) query strategies and sample queries for finding LSB structures; and
(g) a baseline concept for a value-added database of LSB structures

### 3.4.2 Techniques for identifying and deblending overlapping galaxies

**G-TAS-2.** *Motivation:* The Level 2 data products are the most relevant starting point for galaxy-evolution science. In the LSST nomenclature, `Objects` represent astrophysical entities (stars, galaxies, quasars, etc.), while `Sources` represent their single-epoch observations. The master list of Objects in Level 2 will be generated by associating and deblending the list of single-epoch source detections and the lists of sources detected on coadds. The exact strategies for doing this are still under active development by the LSST project, and engagement with the science community is essential. While each data release will have unique object IDs, it will be a huge impediment for LSST science if the first few generations of catalogs turn out severely the limit the science that can be done via database queries.

For galaxies science, the issue of deblending is of critical importance. For example, searches for high-redshift galaxies via color selection or photometric redshifts involve model or template spectra that make the prior assumption that the object in question is a single object at one redshift, not a blend of two objects at two different redshifts. Therefore to get a reliable estimate of the evolution of classes of galaxies over redshift, we need to (a) have reasonably clean catalogs to start with and (b) be able to model the effects of blending on the sample selection and derivation of redshift and other parameters. This is critical not just for galaxy-evolution science, but for lensing and large-scale structure studies. This is just one example. Another is the evolution of galaxy morphologies, where the effects of blending and confusion may well be the dominant source of uncertainty.

The plan for the level-2 catalogs is that sources are hierarchically deblended and that this hierarchy is maintained in the catalog. Scientifically important decisions are still to be made about whether and how to use color information in the deblending, and how to divide the flux between overlapping components. Even if the Project is doing the development work, engagement with the community is important for developing tests and figures of merit to optimize the science return.

*Activities:* Preparations for LSST in this area involve working both with simulations and real data. The current LSST image simulations already have realistic source densities, redshift distributions, sizes, and color distributions. However, the input galaxies do not have realistic morphologies. At least some simulations with realistic morphologies are needed, especially for the Deep Drilling Fields. Inputs should come both from hydrodynamical simulations (where "truth" is known), *Hubble* images, and images from precursor surveys such as CFHTLS, DES and HyperSuprimeCam. The science collaborations should help provide and vet inputs.

More challenging is to come up with techniques and algorithms to improve the deblending. When two galaxies at different redshifts overlap, using observations from all the LSST filters and perhaps even EUCLID and WFIRST might help to disentangle them. Some attempts have been made over the past few years to incorporate color information into the deblending algorithm, but this needs much more attention, not only for developing and testing algorithms, but for deciding on figures-of-merit for their performance.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) providing realistic galaxy image inputs to the ImSim team;
(b) developing tests and figures of merit to quantify the effects on several science objectives;
(c) assessing the current baseline plan for level-2 deblending and for parameter estimation for blended objects; and
(d) developing prototype implementations of deblending algorithms that take advantage of the LSST color information.

### 3.4.3 Optimizing Galaxy Morphology Measurements

**G-TAS-3.** *Motivation:* Measurements of galaxy morphologies are an important tool for constraining models of galaxy evolution. While fairly simple measures of galaxy ellipticity and position angles may be sufficient for the Dark Energy science goals, other kinds of measurements are needed for galaxy-evolution science. The “multifit” approach of fitting simple parametric models to galaxy profiles has been the baseline plan. This will be useful but insufficient. For well-resolved galaxies it is desirable to have separate measures of bulge and disk, and spiral-arm structure, measures of concentration, asymmetry, and clumpiness. These ought to be measured as part of the level 2 processing, to enable database queries to extract subclasses of galaxies. Both parametric and non-parametric measures are desirable. While there will no doubt be optimization in level 3 processing, it is important to have enough information in the level 2 output products to pick reasonable subsets of galaxies.

*Activities:* The preparation work, therefore, focuses on defining measures to enable these queries. Two aspects of LSST data make this a significant research project: the fact that LSST provides multi-band data with a high degree of uniformity, and the fact that the individual observations will have varying point-spread functions. The former offers the opportunity to use much more information than has been generally possible. The latter means that it will take some effort to optimize and calibrate the traditional non-parametric measure of morphology (e.g. the CAS, GINI and M20 parameters), develop new LSST-optimized parameters, and optimize their computation to avoid taxing the level-2 pipeline.

Given the very large data set and the uncertainty in how to use specific morphological parameters to choose galaxies in certain physical classes (e.g. different merger stages or stages of disk growth), it is important to have extensive training both from hydrodynamical simulations with dust (where physical truth is known, even if the models are imperfect) and from observations where kinematics or other information provide a good understanding of the physical nature of the object. These training sets ought to be classified by humans (still the gold-standard for image classification) and via machine-learning techniques applied to the morphological measurements. A series of “classification challenges” prior to the LSST survey could help to refine the techniques.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) providing realistic galaxy image inputs for classification tests to the ImSim team;
(b) human classification of the images;
(c) machine-learning algorithms to be tested and developed into suitable SQL queries;
(d) developing a menu of candidate morphological measurements for level 2 and level 3 processing; and
(e) developing tests and figures of merit to quantify the effects on several science objectives.

### 3.4.4 Optimizing Galaxy Photometry

**G-TAS-4.** *Motivation:* Systematic uncertainties will dominate over random uncertainties for almost any research question one can imagine addressing with LSST. The most basic measurement of a galaxy is its flux in each band, but this is a remarkably subtle measurement for a variety of reasons: galaxies do not have well-defined edges, their shapes vary, they have close neighbors, they cluster together, and lensing affects both their brightness and clustering. These factors all affect photometry in systematic ways, potentially creating spurious correlations that can obscure or masquerade as astrophysical effects. For example, efforts to measure the effect of neighbors on galaxy star-formation rates can be thrown off if the presence of a neighbor affects the basic photometry. Measurements of galaxy magnification or measurements of intergalactic dust can be similarly affected by systematic photometric biases. It is thus important to hone the photometry techniques prior to the survey to minimize and characterize the biases. Furthermore, there are science topics that require not just photometry for the entire galaxy, but well-characterized photometry for sub-components, such as a central point-source or a central bulge.

*Activities:* The core photometry algorithms will end up being applied in level 2 processing, so it is important that photometry be vetted for a large number of potential science projects before finalizing the software. Issues include the following. (1) Background estimation, which, for example, can greatly affect the photometry for galaxies in clusters or dwarfs around giant galaxies. (2) Quantifying the biases of different flux estimators vs. (for example) distances to and fluxes of their neighbors. (3) Defining optimal strategies to deal with the varying image quality. (4) Defining a strategy for forced photometry of a central point source. For time-varying point-sources, the image subtractions will give a precise center, but will only measure the AC component of the flux. Additional measurements will be needed to give the static component. (5) Making use of high-resolution priors from either Euclid or WFIRST, when available. Because photometry is so central to much of LSST science, there will need to be close collaboration between the LSST Project and the community.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) developing metrics for various science cases to help evaluate the level 2 photometry;
(b) providing realistic inputs for difficult photometry cases (e.g. close neighbors, clusters of galaxies, AGN in galaxies of various morphologies);

Deliverables over the longer term include develping optimal techniques for forced photometry using priors from the space missions.

### 3.4.5 Optimizing Measurements of Stellar Population Parameters

**G-TAS-5.** *Motivation:* The colors of galaxies carry information about their star-formation histories, each interval of redshift being a snapshot of star-formation up until that time. Unfortunately, estimates of star-formation rates and star-formation histories for a single galaxy based on only the LSST bands will be highly uncertain, due largely to degeneracies between age, dust extinction and metallicity. Strategies for overcoming the degeneracies include hierarchical modeling – using ensembles of galaxies to constrain the hyper-parameters that govern the star-formation histories of sets of galaxies rather than individuals, and using ancillary data from other wavelengths.

*Activities:* Activities in this area include developing scalable techniques for hierarchical Bayesian inference on very large data sets. These can be tested on semi-analytical or hydrodynamical models, where the answer is known, even if it does not correctly represent galaxy evolution. The models should also be analyzed to find simple analytical expressions for star-formation histories, chemical evolution and the evolution and behavior of dust to make the Bayesian inference practical.

Another important activity is to identify the ancillary data sets and observing opportunities, especially for the deep fields.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) developing and refining techniques for constraining star-formation histories of large ensembles of galaxies;
(b) providing model inputs to guide in developing these techniques;
(c) refining the science requirements for ancillary multi-wavelength data to support LSST.

## 3.5 High-Redshift Galaxies

Observations of distant galaxies provide critical information into the efficiency of the galaxy formation process, the end of the reionization era, the early enrichment of the intergalactic medium, and the initial conditions for the formation of modern galaxies at later times. Through its wide area and sensitivity in $zy$ LSST will probe galaxies out to $z \sim 7$, and will probe yet further in conjuction with future wide-area infrared surveys. The following science tasks address outstanding preparatory work for maximizing high-redshift science with LSST.

### 3.5.1 Optimizing Galaxy Photometry for High-Redshift Sources

**HZ-1.** *Motivation:* The identification and study of high-redshift galaxies with LSST hinges on reliable, accurate and optimal measurements of the galaxy flux in all LSST passbands. Galaxies at redshifts above 7 will only be detected in the LSST y-band and will be non-detections or "drop-outs" in the other LSST filters. Galaxies at redshifts above 8 will not be detected at all in the LSST filters but combining LSST with infrared surveys such as Euclid and WFIRST would enable this population to be identified. It is particularly important to have robust flux measurements and robust flux limits for the undetected high-redshift galaxies in the blue LSST filters so this information can be utilized in the high-redshift galaxy selection. Since Euclid and WFIRST are space-based missions with very different spatial resolutions and point spread functions (PSFs) compared to LSST, algorithms also need to be devised to provide homogenous flux measurements for sources across the different surveys.

It is not clear if the current Level 2 data products package will meet all the requirements for high-redshift science with LSST and this therefore needs to be investigated before the start of the survey.

*Activities:* Firstly, we need to get a clearer picture of what constitutes the LSST Level 2 data products so we can assess whether these will be adequate for the high-redshift science. Issues that we need to understand are: 1) Will photometric catalogues be produced using the reddest LSST (e.g. y-band) images as the detection image? This is critical for high-redshift science as high-redshift galaxies will not be detected in the bluer bands. 2) When computing model galaxy fluxes, will negative fluxes be stored? Negative fluxes for undetected galaxies together with their corresponding errorbars, provide useful input into spectral energy distribution (SED) fitting codes for high-redshift galaxy selection.

The second major activity will be determining the best approach to combining LSST data with infrared data from Euclid/WFIRST for high-redshift galaxy selection. We will need to determine the optimal measure of an optical-IR colour for sources from these two datasets. There is the additional complication that sources that are resolved in the Euclid/WFIRST data could be blended in LSST and will therefore need to be accurately de-blended, perhaps using the high-resolution IR data as a prior, before a reliable flux and colour measurement can be made. Tests can be run using existing datasets e.g. from the Dark Energy Survey (DES) and HST.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Determine what constitutes LSST Level 2 data products and document what additional data products will be required for high-redshift science.
(b) Develop tools to produce optimal combined photometry from ground and space-based

surveys and test these on existing datasets.

### 3.5.2 High-Redshift Galaxies and Interlopers in LSST Simulations

**HZ-2.** *Motivation:* Before the start of LSST operations, it is important that we are able to test our selection methods for high-redshift galaxies on high-fidelity simulations. Given the wide-field coverage of LSST, it will be uniquely positioned to uncover large samples of the most luminous and massive high-redshift galaxies at the Epoch of Reionisation and beyond. The most significant obstacle to selecting clean samples of such sources from the photometric data, is the presence of significant populations of interlopers e.g. cool stars in our own Milky Way and low-redshift, dusty and/or red galaxies, both of which can mimic the colours of high-redshift sources. Using the LSST simulations, we want to be able to devise the most effective way of separating these different populations, and utilising both photometric and morphological information for the sources. Based on experience with ground-based surveys such as the Dark Energy Survey and VISTA infrared surveys, we expect at least some of the most luminous z ¿ 6 galaxies to be spatially resolved in the LSST images.

*Activities:* Liaise with the LSST simulations working group to ensure that high-redshift galaxies have been incorporated into the simulations with a representative set of physical properties (e.g. star formation histories, UV-slopes, emission line equivalent widths, dust extinction, metallicity). It is also important that the high-redshift galaxies have the correct number density and size distribution in the simulations. The latter will allow us to investigate how effectively we can use morphology to separate these galaxies from interlopers.

In addition to the high-redshift galaxies, it is equally important from a high-redshift science perspective, that interlopers have been incorporated into the simulations with the correct number densities and colours. Interlopers of particular relevance to the high-redshift searches will be cool stars in our own Milky Way (e.g. L and T-dwarf stars) as well as populations of very red, massive and/or dusty galaxies at lower redshifts of $z \sim 2$.

Finally, we may want to consider whether to include colour information in the infrared filters (e.g. those from Euclid/WFIRST) in the simulations as this information will undoubtedly help with the high-redshift selection.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Incorporate high-z galaxies into LSST simulations with a realistic and representative set of properties.
(b) Incorporate brown dwarfs into LSST simulations
(c) Extend simulations to other datasets beyond LSST (e.g. Euclid/WFIRST filters).

## 3.6 Low-Surface Brightness Science

The exquisite data quality of LSST will enable a new regime in low-surface brightness (LSB) science over large areas of the sky. The capability to conduct unparalleled LSB science with LSST will uncover new evidence for and measures of the cosmic merger rate, reveal the signature of hierarchical structure formation in extragalactic stellar halos, and probe the LSB outskirts around other galaxies. The following science tasks provide an enumeration of preparatory research tasks for leveraging fully the LSST dataset for LSB science.

### 3.6.1 Techniques for Finding Low-Surface Brightness Tidal Features

**LSB-1.** *Motivation:* A key advantage of LSST over previous large-area surveys (e.g. the SDSS) is its ability to detect low-surface-brightness (LSB) features associated with galaxies. This includes tidal streams and other features associated with past and ongoing interactions, intra-cluster and intra-group light, and nearby, extended low- surface-brightness galaxies. Prior to LSST, typical studies of the LSB universe have focused on small galaxy samples (e.g. in the SDSS Stripe 82), often selected by criteria that are difficult to quantify (e.g. visual inspection that can somewhat subjective) or reproduce in theoretical models. Automated (algorithmic) measurements of the LSB features themselves can be challenging and many past studies have relied on visual inspection for the identification and characterization of features (which may not easily applied on the LSST scale). For LSST it is highly desirable that we automate the detection and characterization of LSB features, at least to the point where samples for further study can be selected via database queries, and where the completeness of samples returned from such queries can be quantified.

*Activities:* Several activities are of crucial importance:

(a) Simulating realistic LSST images and LSB features (using, e.g., high-resolution hydro simulations)
(b) Identifying precursor datasets that can be used as proxies for developing LSB tools for use on LSST data
(c) Using the simulations to develop algorithms for detection and measurement of LSB features
(d) Applying these algorithms to the precursor datasets to test their suitability
(e) Ensuring that LSST level-2 processing strategies and observing strategies are aligned with the needs of LSB science
(f) Developing a strategy for finding and measuring LSB features through a combination of level 2 measurements, database queries, and level 3 processing

It is important to produce realistic LSST images from e.g. the current generation of hydrodynamical cosmological simulations (which faithfully incorporate both the evolution of large-scale structure and the interplay between baryons and dark matter during interactions). Scattering from bright stars (which may or may not be in the actual field of view of the telescope) is likely to be the primary source of contamination when searching for extended LSB features. Ideally, the LSST scattered-light model, tuned by repeated observations, will be sufficiently good that these contaminants can be removed or at least flagged at level 2. Defining the metrics for sufficiently good, based on analysis of simulated images, is an important activity that needs early work to help inform LSST development.

Including Galactic cirrus in the simulations will be important when developing strategies for detecting for large-scale LSB features. Including a cirrus model as part of the LSST background estimation is worth considering, but it is unclear yet whether the science benefit can justify the extra effort.

Because the LSST source extraction is primarily optimized for finding faint, barely-resolved galaxies, it will be challenging to optimize simultaneously for finding large LSB structures and cataloging them as one entity in the database. For very large structures, analysis of the LSST sky background map, might be the most productive approach. We need to work with the LSST project to make sure the background map is stored in a useful form, and that background measurements from repeated observations can be combined to separate the fluctuating foreground and scattered light from the astrophysically interesting signal from extended LSB structures. Then, we need strategies for measuring these background maps, characterizing structures, and developing value-added catalogs to supplement the level 2 database.

For smaller structures, it is likely that the database will contain pieces of the structure, either as portions of a hierarchical family of deblended objects, or catalogued as separate objects. Therefore, we need to develop strategies for querying the database to identify galaxies which are likely to have such structures. E.g. in galaxies that have LSB tidal features around them, the main body of the galaxy is likely to be disturbed and therefore asymmetric. Measures of asymmetry will therefore be useful for flagging such systems. We then need to have a strategy for either extracting the appropriate data for customized processing, or develop ways to put back together the separate entries in the database. A possible value-added catalog, for example, from the galaxies collaboration might be an extra flag in the database to indicate that a galaxy is likely to have LSB tidal features and an extra set of fields for the database to indicate which separate objects are probably part of the same physical entity.

This would be relatively sparsely populated in the initial stages of LSST. Estimates from the Stripe 82, indicate that 15% of galaxies carry LSB tidal features (LSST will reach Stripe 82 in a single shot) but by the end of the survey will become a key resource for a wide variety of investigations.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Realistic mock LSST images from hydro cosmological simulations (including re-simulations of individual objects were necessary) with spatial resolutions of tens of parsecs.
(b) Algorithms for finding galaxies with LSB features and for measuring the properties of these features.
(c) Input to the Project on scattered-light mitigation and modeling strategies from the simulations.
(d) Input to the Project on photometric and morphological parameters (e.g. asymmetry, residual flux fractions etc) to measure and store in the level 2 database.
(e) Query strategies and sample queries for finding LSB structures.
(f) A baseline concept for a value-added database of LSB structures.

### 3.6.2 Low-Surface Brightness Galaxies

**LSB-2.** *Motivation:* Our objective is to investigate the most relevant and challenging aspects of the Low Surface Brightness (LSB) Universe. This has a direct baring on the range of galaxies initially formed, the properties that they have during and after their assembly, their connection to the cosmic web and ultimately to the nature of dark matter, which plays a large part in all of these processes.

By LSB we mean objects that have surface brightnesses much less than that of the background night sky and that which is typical of the Milky Way galaxy we live within. Many authors have previously shown how difficult it is to detect objects of LSB and, more importantly, that our current observations may be severely biased towards detecting objects that have surface brightnesses very similar to the spiral galaxy that we live within. Thus the Universe we perceive may have more to do with the position we are observing it from than its true nature - what would we see if we were able to move our telescopes away from the Sun and out to the very outer edges of the Galaxy?

The problem is that astronomical observations always include a signal from a background, a level we need to detect our sources above. For ground based observations the background arises locally from the atmosphere and our proximity to the Sun, scattered light from the solar system, diffuse star light from the Galaxy and a small contribution from other galaxies in the Universe. For an astronomical object to be detected it must stand out above the noise level in this background. If this noise was purely due to photon statistics then very simply all we would need to do is collect as many photons as possible and the signal would gradually appear out of the noise. However, we currently know that it is nowhere near this simple because of scattered light across the field of view (FOV), instrumental calibration uncertainties and real fluctuations in the cosmic background. For these reasons there has previously been little progress in making a definitive study of the extent and brightness limits of the LSB Universe.

Additionally, this LSB universe include a large percentage of galaxies representing the low-mass end of the galaxy mass function, which in turn has been a major source of tension for the LCDM cosmological model. The galaxy mass function at masses less than Mh 1010 Msun systematically departs from the halo mass function in ways that are difficult to reconcile with current models of baryonic feedback. On the observational side, a crucial step towards understanding the discrepancy is to derive a much more complete census of low-mass galaxies in the local universe. For gas-poor galaxies, which includes most dwarfs within the halos of Milky-Way like galaxies, detection via HI surveys or emission-line surveys is nearly impossible. Dwarf galaxies in the Local Group can be found by searching for overdensities of individual stars. At much larger distances, this becomes impossible. However, these galaxies are still quite easy to detect in LSST images.

The challenge is to identify them as nearby dwarfs and estimate their distances and hence luminosities. The dwarfs in question are low-surface-brightness galaxies, so many of the source-detection issues are common to the more general problem of detecting LSB features. LSST data allow us to focus on different issues. For certain distances and luminosities, typical dwarf spheroidal galaxies will be distinct from the vast majority of background galaxies in the radius vs apparent magnitude plane. However, there will often be overlapping background galaxies, so it is important that the de-blending and cataloging steps try to remove the overlaps and allow one to query for galaxies in the right portion of the color-size-brightness manifold. Once candidates are identified, it should be possible to tease out approximate distances for many dwarf galaxies via surface-brightness fluctuations (SBF). Once again this requires careful treatment of the background galaxies, but this step is now Level 3 processing, so can be customized much more than the detection step. More ambitiously, it is conceivable that machine-learning techniques could be trained to identify semi-resolved nearby dwarf galaxies given a suitable training set from LSST-precursor observations.

On the other extreme of LSB objects, the largest spiral galaxy known since 1987 (called Malin 1), has an extremely LSB disc of stars and an impressive system of spiral arms only revealed in 2015. The central bulge of the galaxy is prominent, but the stellar disc and spiral arms only revealed itself after sophisticated image processing. Malin 1 was discovered by accident and has for almost thirty years been unique. How many more galaxies with rather prominent central bulges also have extended LSB discs? This issue is very important for understanding the angular momentum distribution of galaxies and where this angular momentum comes from - for its stellar mass Malin 1 has about a factor ten higher angular momentum than typical values. The limiting SB of the LSST combined with the large FOV make this instrument unique to probe the existence of large LSBs, similar to Malin 1. There is also an existing problem relating galaxies formed in numerical simulations to those observed. Models with gas, cooling and star formation lose gas and angular momentum making disc galaxies too small. This has already been termed the angular momentum catastrophe and galaxies with giant discs like Malin 1 only make this problem worse. This is particularly important as there is increasing evidence that angular momentum plays a large part in determining the morphology of galaxies, a problem that has plagued galaxy formation studies since its inception. In addition we will be exploring the very outer regions of galaxies and so will be able to explore the connection between the decreasing surface density of baryons and the increasing significance of the dark matter component of galaxies. One reason why this subject has made little progress over the last few years is because of the limited amount of deep large area data available. Most previous deep (CCD) surveys have been specifically designed to investigate the distant Universe and so, like the Hubble Deep Field, have concentrated on long exposures over small areas of sky. The extensive sky survey that LSST will carry out will become the state-of-the-art for years to come and offers a new and enormous LSB discovery potential. As a pointer to these exciting discoveries there have recently been relatively small-scale observations that indicate that a hidden LSB galaxy population does exit. An example is the population of LSB galaxies recently detected in the Coma and Fornax clusters, galaxies not only with astonishing LSB (¿27 B mag arcsec-2), but also with some of them exhibiting effective radii similar to that of the Milky Way. This is despite both Coma and Fornax being two of the previously most studies regions of the nearby Universe.

To quantify the astronomical problem we can give some approximate numbers. The typical sky background at a good dark astronomical site is $\approx$ 22.5mag arcsec$^2$ and that from a space telescope typically an order of magnitude fainter $\approx$ 25.0mag arcsec$^2$. The mean surface brightness (averaged over the half-light radius) of a galaxy like the Milky Way is $\approx$ 23.0mag arcsec$^2$, of order the brightness of the darkest sky background seen from the ground. The mean surface brightness of the giant LSB galaxy Malin 1 is about $\approx$ 28mag arcsec$^2$, some 100 times fainter than that of the Milky Way and that of the sky background. Extreme dwarf galaxies in the Local Group have mean surface brightnesses as faint as $\approx 32$mag arcsec$^2$, $10^4$ times fainter than the background, but these have only been found because they are resolvable into high surface brightness stars - something that is not currently possible to do from the ground for distances beyond about 5 Mpc. Note that 26mag arcsec$^2$ corresponds to approximately a surface density of about one solar luminosity per sq parsec. Our intention is to explore the Universe using LSST to at least a surface brightness level of 30mag arcsec$^2$.

*Activities:* (a) Production of simulated data that can be passed through the LSST data reduction pipeline.

(b) Analysis of simulated images to ensure that LSB features can be accurately preserved and measured.

(c) The development of new object detection software specifically designed for the detection of LSB features, in particular:

- Objects with large size.
- Objects near or melted with large size, bright galaxies.
- Objects with patterns similar to galaxy streams.
- Highly irregular and distorted objects.

(d) Identification of precursor data sets that can be used to test our methods. We can use data generated using numerical simulations to look at the types of galaxies produced that have sufficient angular momentum to become LSB discs. These discs can be quantified and placed within simulated data to test the ability of the pipeline to preserve LSB features. We will develop new methods of detecting LSB objects. These will include pixel clustering methods and the labeling of pixels with certain properties i.e. surface brightness level, SED shape, proximity to other similar pixels etc. We will trial our methods on other currently available data sets (KIDS, CFHT etc).

(e) Simulate realistic LSST images of nearby dwarf galaxies.

(f) Identify nearby semi-resolved dwarf galaxies in precursor data sets to use to develop the LSST tools.

(g) Develop and test the database search queries for finding candidates of several shapes and sizes.

(h) Develop and test a measurement of semi-resolved "texture" as a candidate level 2 measurement.

The use of "texture" as a means of identifying candidate nearby dwarf galaxies is something that needs near-term attention if it is to make it into level 2 processing early in the survey. This can be developed and tested on the semi- resolved-galaxy simulations, but it is also essential to test it on precursor data sets from DES, CFHTLS or HSC.

As a natural consequence of the effort that the members of this team are going to invest on the discovery and catalogue, we can foresee a long-term group effort for continuing the research once deliverables are available. A natural strategy, will be to perform several follow ups with large aperture telescopes available in Chile, with powerful instruments capable of obtaining optical, near-IR spectra, sub-mm, mm and IFU data for LSB objects.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Realistic mock LSST images.

(b) An assessment of the influence of the PSF, scattered light and other instrument signals that may affect our ability to detect LSB features

(c) An assessment of the effect of proposed pipeline on the detection and measurement of LSB features.
(d) A baseline concept for the construction of a database of LSB features detected using LSST data.
  - Realistic inputs of dwarf galaxies for the LSST image simulations.
  - Realistic postage-stamp simulations of semi-resolved dwarf galaxies.
  - Training set of nearby dwarfs from LSST precursor data.
  - Figures of Merit for detection and selection algorithms
  - Run LSST pipeline on both simulations and precursor data and assess performance.
(e) Optimized algorithms measuring surface brightness fluctuation distances.
(f) A new LSB object detection package, friendly adapted for the user.

### 3.6.3 Probing the Faint Outskirts of Galaxies with LSST

**LSB-3.** *Motivation:* The outskirts of nearby galaxies, loosely defined as the regions below $25-26$mag arcsec$^2$ in surface brightness, have long been studied mainly in HI, and later in the UV thanks to the exquisite imaging by GALEX. Deep optical imaging of these regions has been performed on individual objects or on small samples by using extremely long exposures on small (including amateur and dedicated) telescopes, using the SDSS Stripe82 area, and using deep exposures with large telescopes (e.g., CFHT, Subaru, GTC).

The main science driver here is understanding the assembly, formation, and evolution of galaxies. This can be studied through imaging and subsequent parametrization of structural components such as outer exponential disks, thick disks, tidal streams, and stellar haloes. From numerical modelling we know that the parameters of these components can give detailed information on the early history of the galaxies. For instance, halo properties, and structure within the stellar halo, are tightly related to the accretion and merging history. This is illustrated by the imaging of the stars in the outskirts of M31 and other local group galaxies, which show detailed structure.

Ultra-deep imaging over large areas of the sky, as will be provided by LSST, can in principle be used to extend the study so far mostly limited to local group galaxies to 1000's of nearby galaxies, and even, at lower physical scales, to galaxies at higher redshifts. It is imperative, however, to understand and correct for a number of systematic effects, including but not limited to internal reflections and scattered light inside the telescope/instrument, overall PSF, including light scattered by the brighter parts of the galaxy under consideration, flat fielding, masking, residual background subtraction, and then foreground material (in particular Galactic cirrus). Many of these effects, and in particular the atmosphere part of the PSF vary with position and/or time on timescales as short as minutes, which needs to be understood before stacking. They will affect some items more than others, e.g. linear features such as tidal streams may be less affected by overall PSF, but more by foreground cirrus.

*Activities:* Most of the activities to be performed in relation to this task will be in common with other LSB tasks, in particular those related to understanding the systematics and how they vary with time and position on the sky. Good and very deep PSF models will have to be built, likely from a combination of theoretical modelling and empirical measurements, and the PSF scattering of light from the brighter parts of the galaxies will need to be de-contaminated and subtracted before we can analyse the outskirts. Dithering and rotation of individual imaging will need to be modelled before stacking multiple imaging. Commissioning data will need to be used to study the temporal and positional variations of the PSF, and how accurate theoretical predictions for the PSF are (in other words, how much a variable atmospheric PSF component complicates matters).

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Information on the stability and spatial constancy of the LSST PSF.

(b) Improved control over systematics for LSB science, and other fields including weak lensing.

### 3.6.4 Low-Surface Brightness Intracluster Light

**LSB-4.** *Motivation:* The Intra-cluster Light (ICL) is a low surface brightness stellar component that permeates galaxy clusters. It is predicted to be formed mainly of stars stripped from cluster galaxies via interactions with other members, which then become bound to the total cluster potential. The ICL is also likely to contain stars that formed in the gaseous knots torn from in-falling galaxies as they are ram-pressure stripped by the hot intra-cluster medium. Therefore, it is important to study the ICL as it has kept a record of the assembly history of the cluster. Assuming LSST and its data products are sensitive to large LSB structures (see Activities and Deliverables) then it will be possible to perform the first comprehensive survey of ICL in galaxy clusters and groups within a uniform dataset.

Some outstanding scientific questions, which LSST could solve:

- When does the ICL (to a given SB limit) first emerge i.e. at what redshift and/or halo mass?
- Does it contain significant substructure?
- What is its surface brightness profile and does it have a colour dependence, which would indicate age/metallicity gradients?
- Where does the ICL begin and the large diffuse cD halo of the Brightest Cluster Galaxy (BCG) end and do they have the same origin?

*Activities:* The preparation work for the ICL component of the LSB case involves investigating LSST specific issues for large LSB features and the known properties of the ICL itself.

The LSST specific issues fall into three categories: telescope; observation strategy; and pipeline. The faint, large radii wings of the PSF and any low-level scattered light or reflections from the telescope optics or structure will produce low surface brightness signals, which could easily mimic the ICL. The dither pattern of the observations, if smaller than the typical extent of a cluster, could mean that the ICL is treated as a variation in the background during the reduction and/or image combination process, rather than as a real object. This leads onto the pipeline itself which, regardless of the dither pattern, could remove the ICL if an aggressive background subtraction is used on either single frames or when combining images. It is therefore crucial for the LSB team to liaise with LSST strategy, telescope, instrument and data reduction teams.

The ICL specific issues are mainly the feasibility of observing the ICL given its known properties, which can be simulated from existing data. Using deep observations of the ICL in low redshift clusters we can model whether we expect to see ICL at higher redshifts (up to z=1) given dimming, stellar population evolution and the surface brightness limits of LSST. This is crucial if we want to look for an evolution in ICL properties. If we want study low mass groups or high redshift systems, we may need/want to stack populations to obtain a detection of the ICL. It is important to assess whether a genuine stacked ICL detection could be achieved by a comprehensive masking of galaxy cluster members or would faint galaxies just below the detection threshold end up combining to give a false or boosted ICL signal.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Investigate any telescope specific issues that affect the measurement of large LSB features: PSF wings; scattered light.
(b) Investigate observation specific issues that affect the measurement of large LSB features: dither pattern strategy.
(c) Investigate image pipeline specific issues that affect the measurement of large LSB features: background removal; image combining.
(d) Feasibility: given the depth/surface brightness limit of the LSST imaging, to what limits can we hope to recover ICL in clusters and to what redshifts? Can this be simulated or extrapolated from deep imaging of low-z clusters?
(e) Investigate stacking clusters to obtain faint ICL - this is difficult as will require very strong masking of even the faintest observable cluster members.

## 3.7 Photometric Redshifts

For a photometric survey like LSST, the ability to measure vast distances to galaxy samples, understand the time evolution and spatial clustering of galaxy populations, surmise the stellar mass, age, and metallicity of objects, and to identify unusual objects at various cosmic epochs relies heavily on methodologies for constructing photometric redshifts from the data. The following important preparatory science tasks address both the systematic uncertainties on photometric redshifts associated with the LSST observatory and with the requisite stellar population synthesis models.

### 3.7.1 Impact of Filter Variations on Galaxy photo-z Precision

**PZ-1.** *Motivation:* For accurate photometric redshifts, well calibrated photometry is essential. Variations in the telescope system, particularly the broad-band ugrizy filters, will need to be very well understood if we are to meet the stringent LSST calibration goals. Photometry will be impacted by multiple factors that may vary as a function of position and/or time. The position of the galaxy in the focal plane will change the effective throughput both due to the angle of the light passing through the filter, and potential variations in the filter transmission itself due to coating irregularities across the physical filter. The spatially correlated nature of these effects can induce scale-dependent systematics that could be particularly insidious for measurements of local environment and clustering. The nominal plan from LSST Data Management is to correct for variations across the focal plane. Such corrections will be SED dependent, and may leave residuals, particularly for specific populations with unusual SEDs. Tests of the amplitude of these residuals, and the impact on photo-z as a whole, and for particular object classes, is an important consideration. Beyond this, if the variations turn out to be very well calibrated, they could potentially be used to further improve, rather than degrade, photo-z performance. The variations in filter response can offer up additional a small amount of extra information on the object SED, given the slight variation in effective filter wavelength, particularly for objects with strong narrow features, i.e. emission lines. Tests of how much information is gained can inform whether or not the extra computational effort used in computing photo-zs from many slightly different filters as opposed to measurements corrected to the six fiducial filters of the survey.

*Activities:* Tests of the SED-dependent residuals in photometric redshifts induced by photometric calibration systematics.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Improved photometric calibration for LSST across all pass bands.
(b) Identification of SED classes where photometric redshift failures are prominent.

### 3.7.2 Photometric Reshifts in the LSST Deep Drilling Fields

**PZ-2.** *Motivation:* The LSST Deep Drilling Fields present different challenges than the main survey, including more confusion between sources, and the ability to use the best subsets of the images due to their being many repeat observations. These properties allow investigations of galaxies of brightness close to the noise in the main survey at higher signal to noise.

*Activities:* Assessing robustness of photo-zs with spectroscopic surveys will be difficult at the faintest fluxes, but the relationship to clustering redshifts is critically important.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) **TBD**
(b) **TBD**

### 3.7.3 Multivariate Physical Properties of Galaxies from Photometric Redshifts

**PZ-3.** *Motivation:* The knowledge of the derived physical properties underlies much of the work involving galaxies and their evolution. Derived physical properties include, among others: star formation rate (SFR), stellar mass ($M_{\star}$), specific SFR (sSFR), dust attenuation, and stellar metallicity. When it comes to scientific analysis, in recent years the derived physical properties have largely supplanted fluxes and luminosities in the UV, optical and near-IR bands. This is because derived properties require no redshift (K) corrections, are dust-corrected, and are therefore easier to relate across surveys and studies and to compare with the models. Stellar mass has emerged as a parameter of choice for selecting galaxy types and making apple-to-apple comparisons of galaxies at different redshifts. The sSFR (current SFR normalized by stellar mass) provides a rough estimate of galaxys SF history. Dust attenuation and stellar metallicity are also indicative of various processes important for understanding galaxy evolution.

*Activities:* Deriving physical properties, usually accomplished by spectral energy distribution (SED) fitting, is an involved process and the results depend on the number of factors, including: underlying population models, assumed dust attenuation law, assumed star formation histories, choice of model priors, choice of IMF, emission line corrections, choice of input fluxes, type of flux measurements, treatment of flux errors, SED fitting methodology, interpretation of the resulting probability distribution functions (PDF) (e.g., Salim et al. 2016). In the case of LSST, the additional challenge is that the redshifts are for the most part photometric, and carry a PDF (a measure of uncertainty) of their own. In principle, the redshifts could be determined as part of the SED fitting (and vice versa, physical parameters can be obtained from some photo-z codes), but it is not clear whether this joint approach is the best. Alternatives are to use empirical training sets to obtain the photo-z (some best estimate or a PDF) and then feed it into the SED fitting code.

Activities will consist of testing whether the determination of physical parameters and photo-z should be performed jointly or not, based on training sets with spectroscopic redshifts, at a range of redshifts. Furthermore, testing should be performed on mock galaxies to understand which choices of methods and assumptions (specifically related to LSST data) produces the best results in the sense of retrieving the known properties.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Pre-LSST: A set of guidelines as to optimal practices regarding the derivation of both the photo-z and properties, together with the software to be used.
(b) With LSST data: the production of catalogs of properties to be used by the collaboration.

### 3.7.4 Identifying Spectroscopic Redshift Training Sets for LSST

**PZ-4.** *Motivation:* Require deep spectroscopic redshift data in order to help train algorithms, improve algorithms with clustering etc, and also provide a basis for determining accuracy of photo-z algorithms.

*Activities:* Collate existing spectroscopic redshift data over both DDF and wider fields, along with selection biases for each spectroscopic data set. Assess robustness of existing data, determine colour space where existing surveys lack statistics. Apply for additional spectroscopy to fill in parameter space not already covered by existing surveys.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) **TBD**
(b) **TBD**

### 3.7.5 Develop Techniques to Identify Specific Sub-Populations of Galaxies

**PZ-5.** *Motivation:* Studying properties related with the star formation activity of galaxies, such as color and specific star formation rate (sSFR), as a function of mass, environment and redshift is relevant for understanding the different physical processes in galaxy formation and evolution. The aim is to develop techniques in order to identify specific sub-populations with the aforementioned properties (e.g. blue/star-forming and red/quenched galaxies) based on photometric data. Another interesting sub-population is galaxies which contain an active nucleus. The identification of AGN candidates will also be explored.

This task is potentially cross-cutting with the theory/mock catalogs, machine learning, clusters, lss, AGN, and DESC working groups and collaborations.

*Activities:* We can use simulations and mock catalogs to obtain prior estimates of the calibrations used to identify specific galaxy sub-populations. These calibrations will depend on mass and redshift (z). One technique to explore is fitting two Gaussians to the corresponding color and sSFR distributions in different mass and redshift bins to identify populations of red and blue galaxies. It is important that the mass definition assumed in the mocks be comparable to that estimated for observations. Note that the stellar mass would be used as the alpha parameter in the joint probability distribution functions, p(z,alpha).

Furthermore, we will make efforts to identify AGNs to obtain a sample of AGN candidates and, also, isolate them from normal galaxy samples without AGNs. The information of color and star formation described above can be used for this aim.

The techniques can be probed as a function of environment, which can be defined using different approaches at both small and large scales (e.g. number of neighbor galaxies, location in large-scale structures such as filaments, voids, knots, or Voronoi tessellation techniques). This would enable the characterization of galaxy sub-populations according to the environment. The resulting galaxy sub-populations can be used as training sets to be implemented on machine learning models.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Obtaining mass from mock catalogs compatible with the mass used in p(z,mass).
(b) Developing techniques that depend on this mass and redshift using mock catalogs for selecting samples with red/blue colors.

(c) Developing multiple techniques that depend on this mass and redshift using mock catalogs for selecting star-forming/quenched samples.
(d) Developing techniques that may depend on star formation, color and redshift for selecting AGN samples.
(e) Defining several environment estimators in simulated datasets.
(f) Probing techniques in b), c) and d) as a function of the environments defined in e).
(g) Obtaining training sets to be implemented on machine learning models.

### 3.7.6 Simulations with Realistic Galaxy Colors and Physical Properties

**PZ-6.** *Motivation:* As representative samples of spectroscopic redshifts will be very difficult to compile for LSST, simulations will play a key role in calibrating estimates of physical properties such as galaxy stellar mass, star formation rate, and other properties. This is particularly problematic for photometric surveys, where photometric redshift and physical property estimates must be calculated jointly. In addition, we must include prominent effects that will influence the expected photometric performance, for example the presence of an active galactic nucleus can significantly impact the color of a galaxy and the inferred values for the physical parameters, so models of AGN components of varying strength must be included in the simulations. Many current generation simulations cannot or do not simultaneously match observed color distributions and physical property characteristics for the galaxy population at high redshift. As photo-z algorithms are highly dependent on accurate photometry, realistic color distributions are required to test the bivariate redshift-physical property estimates. Working with the galaxy simulations and high redshift galaxy working groups to develop new simulations with more accurate high redshift colors is a priority. These photo-z needs are not unique, and the improved simulations will benefit the wider Collaboration as a whole.

*Activities:* The main activity for this task is to bring together the knowledge gained from observational studies of high redshift galaxies to act as input for improved simulation metrics. This will require expertise from the photo-z group, the high redshift galaxies group, the AGN group, and the simulations group. In order to test whether mock high-z populations agree with the real Universe, we must have some real data to compare against, even if it is a luminous subsample or only complete in certain redshift intervals. Once such comparison datasets are established, metrics can be developed to determine which simulations and simulation parameters most accurately reproduce the observed galaxy distributions. Assuming that the simulations are valid beyond the test intervals, we can then test bivariate photo-z/physical process determinations to develop improved algorithms.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:
(a) Determination of a list of which physical parameters are important for galaxy science.
(b) Compiling observable datasets that can be used as comparators for simulated datasets.
(c) Developing a set of metrics to compare simulations to the observational data.
(d) Use the metrics in deliverable B to create updated simulations with more realistic parameter distributions.
(e) Development of improved joint estimators for redshift and physical properties (M*, SFR, etc.).

### 3.7.7 Using Galaxy Size and Surface Brightness distributions as Photo-z Priors

**PZ-7.** *Motivation:* Photometric redshift algorithms traditionally use galaxy fluxes and/or colors alone to estimate redshifts. However, morphological information in the form of the galaxys size/shape/surface brightness (SB) profile adds additional information that can aid in constraining both the redshift and type of the galaxy, breaking potential degeneracies that using colors alone would miss. Adding type information beyond just the rest frame SED may help to constrain bivariate galaxy properties that correlate with morphological type as well. If sufficient training samples are available, a Bayesian prior on colors and SB profile, p(z—C,SB), can be constructed that should lead to improved photometric redshifts.

*Activities:* The primary activity in this task is to develop an algorithm to compute a parameterized SB profile fit (e.g. Sersic index, though other measures may be appropriate) for a large number of galaxies. The algorithm must be fast enough to compute SB profiles for large numbers of galaxies. Simulated datasets may be necessary to calibrate this code in the limits of galaxy sizes approaching the size of the PSF, and in the limit of low signal-to-noise ratios. With SB measurements in hand, the computation of a Bayesian prior on redshift given galaxy photometry and SB. This can be done with either simulated datasets, or real observations with spectroscopic redshifts. Tests will then show the performance of such a prior relative to using galaxy photometry alone.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) A fast, scalable algorithm for measuring the surface brightness profile of galaxies.
(b) A cross matched catalog with objects at known redshifts and measured surface brightness profiles.
(c) A Bayesian prior $p(z|C,SB)$ that can be used to improve photo-z measurements.

## 3.8 Theory and Mock Catalogs

A critical challenge for interpreting the vast LSST dataset in the context of a cosmological model for galaxy formation is the development of theory, both in the practical applications of realistic simulations and the engineering of new physical models for the important processes that govern the observable properties of galaxies. The following preparatory science tasks for LSST-related theoretical efforts range from understanding the detail properties of galaxies that LSST will uncover to predicting the large-scale properties of galaxy populations that LSST will probe on unprecedented scales.

### 3.8.1 Image Simulations of Galaxies with Complex Morphologies

**TMC-1.** *Motivation:* LSST images will contain significant information about the dynamical state of galaxies. In principle, this can be exploited to learn about their formation and evolutionary histories. Examples of such features include spiral arms, tidal tails, double nuclei, clumps, warps, and streams. A wide variety of analysis and modeling techniques can be applied to determine the past, present, or future states of observed galaxies with complex morphologies, and therefore improve our understanding of galaxy assembly.

*Activities:* Activities include creating synthetic LSST observations containing a wide variety of galaxies with complex morphologies, for the purpose of testing analysis algorithms such as de-blending, photometry, and morphological characterization. Supporting activities include creating databases of galaxy images from models (such as cosmological simulations) or existing optical data, analyzing them using LSST software or prototype algorithms, and distributing the findings of these studies. These analyses can be performed on small subsets of the sky and do not necessarily have to include very-large-area image simulations or match known constraints on source density. Results will include predicting the incidence of measured morphological features, optimizing level-3 measurements on galaxy images, and determining the adequacy of LSST data management processes for these science goals.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Creating synthetic LSST images of galaxies with complex morphology from simulations.
(b) Creating synthetic LSST images based on prior observations in similar filters.
(c) Making these LSST-specific complex galaxy data products widely available.
(d) Publicizing results of algorithm tests based on these simulations.
(e) Assessing level-3 measurements to propose and/or apply to maximize the return of LSST catalogs for complex galaxy morphology science.

### 3.8.2 New Theoretical Models for the Galaxy Distribution

**TMC-2.** *Motivation:* Our aim is to bring together key areas of expertise to meet the challenge of building synthetic, computer generated mock surveys which will be used in the preparation for Galaxy science with LSST. Surveys like LSST will collect more data than is contained in the current largest survey, the SDSS, every night for ten years. The analysis of such data demands a complete overhaul of traditional techniques and will require the incorporation of ideas from different disciplines. The mock catalogs we will produce offer the best means to test and constrain theoretical models using observational data. Computer mock catalogs play a well established role in modern galaxy surveys. For the first time, the scientific potential of the new surveys will be limited by systematic errors rather than sampling errors driven by the volume mapped. The signals from viable, competing cosmological models are already extremely close. Distinguishing between the models requires that we build the best possible theoretical predictions to understand the measurements and how they should be analyzed. We also need to understand the errors on the measurements.

*Activities:* Develop a new state-of-the-art in physical models of the galaxy distribution combining models of the physics of galaxy formation with high resolution N-body simulations which track the hierarchical growth of structure in the matter distribution. The key task is to take the results of calculations in moderate volume cosmological N-body simulations and to develop schemes to embed this information into very large volume simulations. The large volume simulations will be bigger than the target survey, allowing a robust assessment of the systematic errors on large-scale structure measurements.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Physically motivated mock galaxy catalogues on volumes larger than will be sampled by LSST, with a consistently evolving population of galaxies.
(b) Base catalogues of dark matter haloes and their merger trees that will be available for other theoretical models of populating these with galaxies (halo and subhalo occupation/abundance matching techniques).
(c) Small volume simulations for further tests.

### 3.8.3 Design of New Empirical Models for the Galaxy Distribution.

**TMC-3.** *Motivation:* We will explore the galaxy-halo connection, using the simulations of the galaxy formation process as encapsulated in physically motivated models to build better empirical models. Empirical models can be adjusted to reproduce observational results as closely as possible, whereas physical models are computationally expensive, so only a small number of examples can be run, and the results cannot be tuned in the same way. Empirical models also have the advantage of being extremely fast, allowing large parameter spaces to be explored.

*Activities:* There are two stages here: one is to test current models to see how well they can reproduce the predictions of physical models and the second is to use the physical models to devise new parametrizations to describe galaxy selections for which there is little or no current observational data. This is particularly relevant for upcoming surveys which will probe regimes that remain largely unmapped.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) The evolution of the clustering predicted by the physical models may allow us to model how parameters should change in the empirical models, thereby reducing the number of parameters which we need to fit and to populate catalogs on the observers past lightcone.

### 3.8.4 Estimating Uncertainties for Large-Scale Structure Statistics

**TMC-4.** *Motivation:* The ability to interpret the relation between galaxies and the matter density field will depend critically on how well we understand the errors on large-scale structure measurements. The accurate estimation of the covariance on a large-scale structure measurement such as the correlation function would require tens of thousands of simulations.

*Activities:* Devise and calibrate analytic methods for estimating the covariance matrix on large-scale structure statistics using N-body simulations and more rapid but more approximate schemes, based for example on perturbation theory. Coordinate with WGs of the Dark Energy Science Collaboration as these covariance matrices can also be applied to Cosmological parameter constraints.

*Deliverables:* Deliverables over the next several years from the activities described above include the following:

(a) Physically motivated estimates of covariance matrices for galaxy occupation (and other) parameter searches.

# Bibliography


Abraham, R. G., van den Bergh, S., & Nair, P. 2003, ApJ, 588, 218

Abraham, R. G., & van Dokkum, P. G. 2014, PASP, 126, 55

Bullock, J. S., & Johnston, K. V. 2005, ApJ, 635, 931

Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183

Flaugher, B. 2005, International Journal of Modern Physics A, 20, 3121

Grogin, N. A., et al. 2011, ApJS, 197, 35

Heckman, T. M., et al. 2005, ApJ, 619, L35

Johnston, K. V., Bullock, J. S., Sharma, S., Font, A., Robertson, B. E., & Leitner, S. N. 2008, ApJ, 689, 936

Kaiser, N., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7733, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0

Kartaltepe, J. S., et al. 2007, ApJS, 172, 320

Koekemoer, A. M., et al. 2011, ApJS, 197, 36

Lin, L., et al. 2008, ApJ, 681, 232

Lotz, J. M., et al. 2008, ApJ, 672, 177

Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163

LSST Science Collaboration, et al. 2009, ArXiv e-prints

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415

Martínez-Delgado, D., Peñarrubia, J., Gabany, R. J., Trujillo, I., Majewski, S. R., & Pohlen, M. 2008, ApJ, 689, 184

Miyazaki, S., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8446, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, 0

Noeske, K. G., et al. 2007, ApJ, 660, L43

Robotham, A. S. G., et al. 2014, MNRAS, 444, 3986 van Dokkum, P. G., Abraham, R., & Merritt, A. 2014, ApJ, 782, L24

York, D. G., et al. 2000, AJ, 120, 1579