The Physics of Galaxy Morphology

PhD Project Proposal

Matthew Wilkinson[1,2]

Supervisors: Dr Aaron Ludlow and Dr Claudia Lagos

[1] International Centre of Radio Astronomy Research (ICRAR) - University of Western Australia (UWA)

[2] ARC Centre for Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO-3D)

August 12, 2020

Abstract

A complete understanding of the physical processes that give rise to the diverse morphologies of galaxies remains an outstanding problem in astronomy. Some galaxies have clear structures, such as disks and spiral arms, while others are closer to featureless ellipticals. Several breakthroughs in numerical simulations over the past 5 years have enabled computational astrophysicists to produce realistic galaxy populations that display a similar morphological diversity to observed ones. These simulations implement general physical laws (e.g. gravity and hydrodynamics) and start from cosmologically-motivated initial conditions; the galaxies that form within them are an emergent phenomenon; they are not implemented *a priori*. During my PhD I will investigate how galaxy morphologies arise in simulations and pin-down the factors leading to the wealth of structure identified in them. This is not a trivial task: even for simulated galaxies – for which the full position and velocity information of all stellar particles is available – robustly disentangling the various components of galaxies remains an open problem. As a first step, I will improve upon past attempts to develop a verified structural decomposition method that can be used to properly identify and classify various galaxy components. I will then employ this method to analyze the `EAGLE` simulation (a current state-of-the-art cosmological simulation of galaxy formation) and use the results to quantify galaxy morphologies statistically, and to study their cosmological origins.

International Centre for Radio Astronomy Research

THE UNIVERSITY OF WESTERN AUSTRALIA

1 Research Project

1.1 Background

Galaxies are the end result of the complex interplay between physical processes that affect stars, gas, black holes and dark matter. These physical processes are ultimately what determines the shape and structure of the galaxies we observe. Gravitational interactions between the components above are a key element, as they determine how quickly features, such as spiral arms or bars in galaxies, can form and for how long they live. However, other physical processes related to how stars and black holes affect their environment via their injection of energy and momentum onto their surroundings, as well as radiative processes that can lead to the formation of new stars, can have far reaching effects in the formation of the central overdensity of galaxies (a.k.a. galaxy bulge), their spiral structure etc. The purpose of this thesis is to connect the components we observe in galaxies with the underlying physical processes, but to also determine which features are predominantly a consequence of specific events (e.g. galaxy collisions). For this, it is important to first consider what we understand of galaxy morphology from observations (addressed in Section 1.1.1), and secondly how galaxy morphology has been studied in galaxy formation simulations up to date (addressed in Section 1.1.2 and Section 1.1.3).

1.1.1 Observations

There are a two ways that starlight from galaxies is collected: through photometry or spectroscopy. Their respective strengths and weaknesses make them useful for different types of scientific questions. First, for photometry, a telescope filters out light except for a narrow wavelength range. Photometry is relatively fast so can be carried out for many galaxies simultaneously (e.g. photometry for over 1,000,000 galaxies was released at one time in Simard et al. (2011)). The technique has good spatial resolution, which is useful for studying the positions of stars in galaxies, and often exposes galaxy structural features. Massive stars are brighter and hotter but use their fuel faster and therefore have shorter lifetimes. We can estimate where star formation is occurring in a galaxy based on the location of the brightest stars and the location of the gas. We can further estimate the ages of populations of stars by the brightest (most massive) stars in that population.

The other observational technique is spectroscopy, which separates starlight across a range of wavelengths. This technique allows for the study of chemical emission and absorption lines of stars and gas. If a star is moving away from us then the emission lines will be shifted to longer wavelengths by the Doppler effect and the magnitude of the shift allows astronomers to calculate the line of sight recession velocity. With accurate chemical models of stars, we can derive both the average velocity and the range of velocities for entire stellar populations, allowing us to map the movement of stars across a galaxy (Cappellari, 2017).

These two observational techniques have helped build a functional understanding of galaxies. The first, obvious phenomenon observed was that galaxies had a range of morphologies (shapes and structures). Edwin Hubble (1926) described a classification scheme to categorise galaxies, now known as the “Hubble Tuning Fork,” depicted in Figure 1. This classification scheme is commonly used throughout astronomy and was initially thought to reflect the life cycle of a galaxy. Broadly, when galaxies are evolving without major disruptions, there are two major categories of galaxies: ellipticals and spirals.

Elliptical galaxies, also known as “Early type” or by their Hubble type “E”, are generally featureless galaxies that are thought to be highly evolved. Elliptical galaxies have few new stars and little gas to create new stars so they are sometimes referred to as “red and dead” If these galaxies were spirals at some point in their past, interactions and mergers with other galaxies caused the orbits of the stars to be randomised and dominate the galaxy dynamics. Spectroscopy shows that ellipticals are supported against gravity by velocity dispersion (disordered motion) and have relatively little rotation. They are usually best described by a single continuous component, although some galaxies classified as ellipticals contain bulges.

In contrast, spiral “late type” galaxies are evolving galaxies that contain more observable substructure (see Figure 2 and Table 1). Spirals contain gas and generally have active star formation occurring in the disk with the youngest stars observed to be in the galaxy’s spiral arms. Most spirals harbour some kind of central bulge that contains the oldest stars and almost no star formation. The relative ages of stars in each structure suggests that either bulges were the first structures to form or that as a galaxy evolves, its older stars sink from the disk to the bulge. Spectroscopy shows that disks are rotationally supported (ordered motion) and that there are different types of bulges. Classical bulges are almost completely dispersion supported similar to ellipticals, but some galaxies contain a pseudo-bulge which

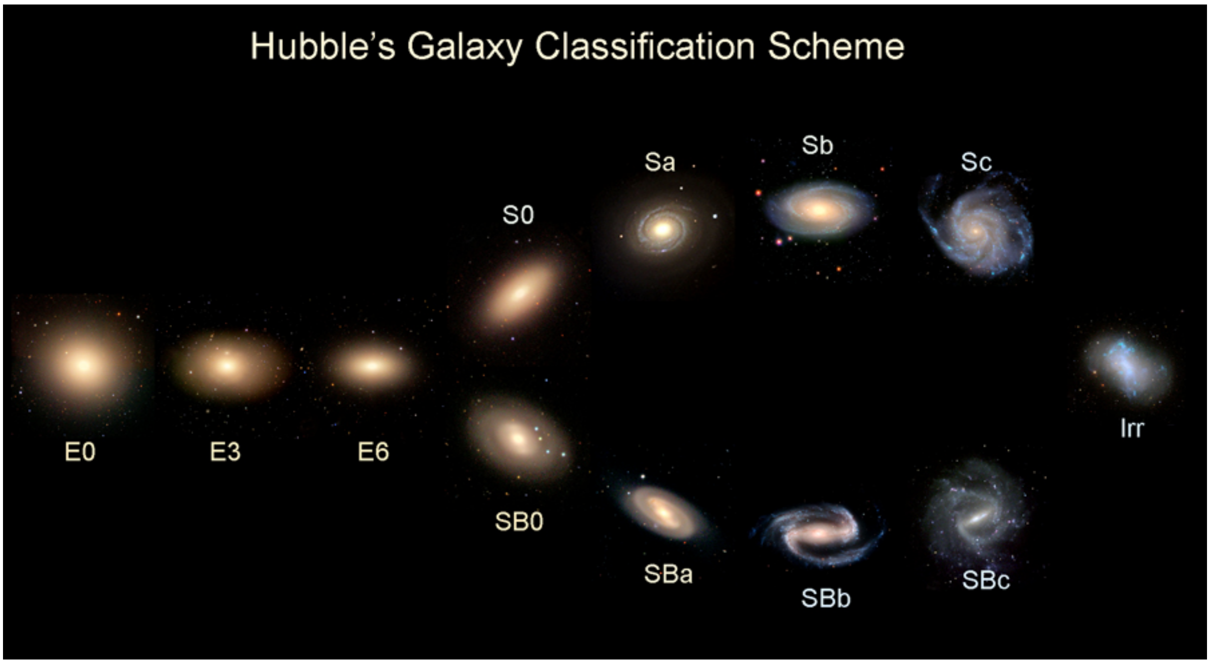

Figure 1: Hubble Tuning fork diagram. Elliptical "early type" galaxies are on the left and have a classification "E", but potentially extended to S0 lenticular galaxies that have a clear bulge and disk, but no spiral arms. Sa to Sd spiral and SBa to SBd barred spiral galaxies, all known as "late type," are on the right. Galaxies that do not fit this classification are irregular or "Irr". The "early" and "late" names are left over from initial incorrect ideas about galaxy evolution. Elliptical galaxies contain less gas and therefore have less star formation activity. They are sometimes referred to as red and dead galaxies. Credit: Alice Mortlock from `https://candels-collaboration.blogspot.com/2013/08/dont-judge-galaxy-by-its-cover.html`

is given a different name due to its rotation and the fact that it is partially rotation and dispersion supported. The two bulges are thought to have different formation mechanisms with classical bulges growing through mergers (i.e. the fusion of two galaxies after they collide) and pseudo-bulges forming as a galaxy evolves in isolation (secular evolution). Late types may also contain a bar at their centre which changes the galaxy's Hubble type from "S" to "SB." A bar is a boxy structure comprised of old stars moving in chaotic orbits. The bar often proves to be the most complicated part of a galaxy to model. Careful observations of edge-on disk galaxies demonstrate that some galaxies have a distinct "thin disk" and "thick disk," with star formation occurring in the younger thin disk. The age difference has led to speculation that thick disks are formed after an early thin disk is heated by mergers. Overall, the motion of stars and the galactic structures present in spirals depend on the mechanisms of the galaxy's formation. For another summary of the different galaxy components, see Table 1.

Clearly, galaxies are formed as a result of different processes so the Hubble types are an important categorisation. There are, however, a few issues with the classification:

1. it is subjective,
2. the classification depends on viewing angle and
3. it cannot be done quickly by a computer.

This is where a "morphological indicator" – an objectively measured property that can be used as a proxy for morphology – comes into play.

The most common observational morphological indicator is the bulge-to-total ratio, B/T which is the ratio of the integrated light from the fitted bulge component to that of sum of all the fitted components of the galaxy. A bulge (and an early-type galaxy) is described well by a Sérsic profile. For the spatially-resolved intensity of light, $I(R)$, the profile is

$$I(R) = I_e \exp\left\{-b_n\left[\left(\frac{R}{R_e}\right)^{1/n} - 1\right]\right\}, \tag{1}$$

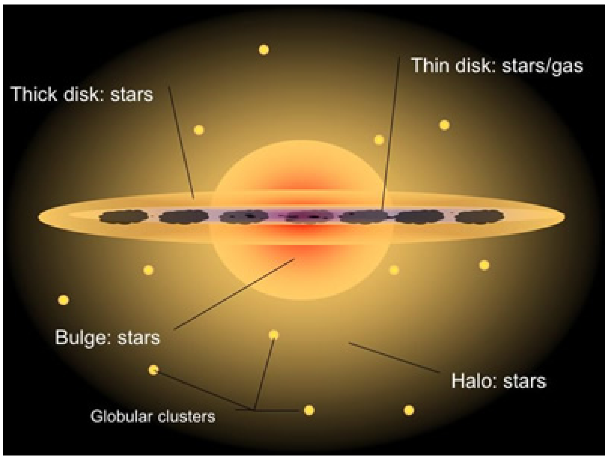

Figure 2: Illustration of the positions of stars in the major galaxy components. Note that the boundaries between each component overlap in space, which makes classifications in some regions more complicated. Some late type galaxies only contain a thin disk while early types don't contain disks at all. New stars are created by gas in the thin disk, represented by dark clouds in this picture.

where R is the projected radius from the centre, I_e is the central luminosity, b_n is a constant which is a function of n, R_e is the scale radius, and n is the Sérsic index (Sérsic, 1963, 1968). The bulge is fit to the profile and once the rest of the galaxy has been fit, a measurement of the bulge-to-total ratio, B/T, can be made. B/T is used as a proxy for morphology: galaxies with $B/T \gtrsim 0.5$ are bulge dominated and are usually ellipticals, while galaxies with $B/T \lesssim 0.5$ are disk dominated and have features of spiral galaxies.

We want to investigate what processes are responsible for the different morphologies and galaxy structures and how they evolve over time. One limitation of observations is that galaxies evolve on extremely long time scales (compared to the 3 years of a PhD) so to examine a galaxy's evolution and the characteristic features present today, we must resort to simulations.

1.1.2 Simulations

There are several approaches to simulating galaxies. The two methods that are relevant for this project are idealised galaxy realisations and cosmological hydrodynamical simulations. Idealised realisations are "model" galaxies that are in equilibrium (i.e. are solutions to the collisionless Boltzmann equation), and have components with different predefined properties such as mass and size. They can generate a large variety of galaxies, with different structures and morphologies which is ideal for our purposes. Each "stellar particle" is assigned to a specific galaxy component e.g. disk or bulge. Galaxies with labelled stars are ideal for testing the accuracy of new component decomposition methods. Tests on ideal data have been crucial for calibrating decompositions with spectroscopic data and are even part of the observational decomposition process (Taranu et al., 2017); but these idealised tests have not been used to test methods of decomposing galaxies in simulations.

While idealised galaxy realisations are a good starting point, we want to investigate different types of galaxies and their realistic merger histories. This is where cosmological hydrodynamical simulations are invaluable. These types of simulations self consistently evolve a representative volume of the universe from early times to the present day. They are run with codes that employ a hydrodynamics scheme that tracks the movement, mass and energy exchange of dark matter, gas, star and black hole particles over time. All of the particles keep track of their own relevant properties, which are periodically printed out to a "snapshot" file. The simulation starts soon after the big bang with only dark matter and gas particles, before any star formation has occurred. When gas particles have high enough density and low enough temperatures, star formation occurs and the gas particle is converted to a stellar particle. Similarly, within sufficiently-massive dark matter halos (e.g. $> 10^{10} M_{\odot}$), a gas particle is converted to a super massive black hole. Stellar particles represent many stars (e.g. $\approx 10^6 M_{\odot}$) and clusters of them represent galaxies. These simulated galaxies are built up by the aggregated effect of star formation, mergers and cosmic accretion. They are our main target for preforming galaxy decomposition as we

Galaxy Structures	Kinematics	mean($j_z/j_c(E)$)	Stellar ages	Formation*
Disky				
thin disk	ordered, circular orbits	≈ 1	young	smooth accretion
thick disk	mostly ordered, almost circular	≈ 0.8	several Gyr	minor disturbance
spiral arms	same as disk	≈ 1	just formed	dynamical instabilities
Bulgy				
(classical) bulge	random orbits	≈ 0	oldest	mergers
pseudo bulge	mostly random but rotating	> 0	old	secular evolution
bar	complicated and rotating	$\gg 0$	old	bar instabilities
halo	large disordered orbits	≈ 0	old	accretion

Table 1: Comparison between different galaxy components. The formation of each component is speculative and is a question this project aims to answer quantitatively. Classical bulges, pseudo bulge and bars are all similarly centrally located, but only classical bulges and halos have similar kinematic properties. Observers classify bulges, pseudo bulges and sometimes bars as a "bulge complex", while simulators consider bulges and halos as a "spheroid."

want to investigate how the components of galaxies change during their evolution.

Since the first demonstration of these simulations' ability to create realistic galaxies in a cosmological context, the field has become extremely active. The current state-of-the-art simulations are `Illustris` (Vogelsberger et al., 2014), `Horizon-AGN` (Dubois et al. 2014), `EAGLE` (Schaye et al., 2015), `Massiveblack-2` (Khandai et al. 2015), `Magneticum` (Boccquet et al. 2016), `MUFASA` (Davé et al. 2016), `Romulus25` (Tremmel et al. 2017), `IllustrisTNG` (Pillepich et al., 2018) and `Simba` (Davé et al. 2019). Each simulation balances the trade-off between number of particles and simulation volume and allows us to compare different physical models. In general, simulations do a good job of producing galaxy populations that are similar to observed ones, even though they use different models (Vogelsberger et al., 2020, for a review).

`EAGLE` (Schaye et al., 2015) is the simulation that I will be using for my PhD. Astronomers at UWA, including my supervisors, Dr. Ludlow and Dr. Lagos, are experienced with the data and are members of the `EAGLE` collaboration. Furthermore, the Virgo Consortium is planning the next generation `EAGLE` simulation, named `EAGLE-XL`. It will have the same spatial and mass resolution, and similar subgrid physics models as the original run, but a volume 27 times larger $(300h^{-1}\text{Mpc})^3$, and will therefore contain a larger diversity of galaxies. This large volume is important for resolving rare objects, such as massive galaxy clusters and galaxies that form in unusual environments. `EAGLE-XL` has been awarded computer time (40 million CPU hours) and is currently (July 2020) coming to the end of its development. As part of ICRAR's involvement in the project, I will be running some small tests to see how well the code scales on GADI, Australia's fastest supercomputer. `EAGLE` will be sufficient for investigating the majority of the phenomena we are interested in but the better statistics and rare objects in `EAGLE-XL` are a bonus. Work carried out for `EAGLE` will be easily adapted to `EAGLE-XL`.

`EAGLE`'s galaxies have already been characterised using several "kinematic morphological indicators." This will allow for a useful comparison to any new morphological measurements. There is significant literature (e.g. Obreschkow and Glazebrook, 2014; Cortese et al., 2016) on the importance of angular momentum and mass as a proxy for morphology. The more (less) angular momentum, the more (less) rotation in the galaxy and the greater chance that it is a spiral (elliptical). One way this can be quantified is with the specific angular momentum of the stars,

$$j_* = \frac{1}{M_*} \left| \sum_i^* m_i \, \vec{r_i} \times \vec{v_i} \right| , \tag{2}$$

where $\vec{r}$ is the distance from the galaxy's centre, $\vec{v}$ is the velocity relative to the galaxy's rest frame and m_i is the particle's mass; $m_i\vec{r_i} \times \vec{v_i}$ is the particle's angular momentum which is summed and normalised by the galaxy's total stellar mass is M_*,. It has been shown using `EAGLE` that galaxies with low j_* are the result of mergers or a lack of star formation and such galaxies are usually early type (Lagos et al., 2017) and that galaxies that undergo gas rich (poor) mergers have higher (lower) j_* (Lagos et al., 2018).

A similar indicator is κ_{co}, a measure of kinetic energy invested in co-rotation. A pure disk of circularly orbiting stars would have $\kappa_{co} = 1$. The parameter was recently presented in Correa et al. (2017), defined

$$\kappa_{co} = \frac{1}{K} \sum_{i}^{r<30\mathrm{kpc}} \frac{1}{2} m_i [j_{z,i}^{>0}/R_i]^2, \tag{3}$$

where $j_{z,i}^{>0}$ is the positive value of a particle's specific angular momentum (i.e. in the direction of rotation), R is the projected distance from the centre in the plane of the disk, making $[j_{z,i}^{>0}/R_i]$ a co-rotational velocity and $\frac{1}{2} m_i [j_{z,i}^{>0}/R_i]^2$ the co-rotational kinetic energy, normalised by $K \equiv \sum_i^{r<30\mathrm{kpc}} \frac{1}{2} m_i v_i^2$, is the galaxy's total kinetic energy. Galaxies with larger (smaller) amounts of energy invested in co-rotation $\kappa_{co} > 0.4 (< 0.4)$ are mostly spiral (elliptical) galaxies and were found to be a good way to separate galaxies by colour (Correa et al., 2019). κ_{co} correlates with S/T (Trayford et al., 2019).

A morphology indicator that is similar to the observational B/T is the spheroid-to-total ratio S/T, which is twice the counter-rotating mass ($j_z < 0$), formally,

$$S/T = 2 \frac{\sum_i^{j_z<0} m_i}{M_*}. \tag{4}$$

Non-rotating dispersion-dominated spheroids will have equal mass co- and counter-rotating, so the spheroid mass is double the counter-rotating mass with $S/T = 1$, where as disks with no counter-rotating matter will have $S/T = 0$. S/T and B/T are different since the spheroid is comprised of halo stars but the bulge is not, while the bulge includes central components that may be rotating but would not be included in the spheroid, see Table 1. Tracking this simple estimator through time reveals different galaxy formation tracts based on stellar mass (Clauwens et al., 2018) and that over the history of the universe, $z \approx 0.5$ was the time where the most stars were in galaxy disks (Trayford et al., 2019).

Alternatively, the simulations can be modelled using similar techniques to observations. Such "like-to-like" mock observations can be generated with a code such as `SKIRT` (Baes et al. 2003, 2011). It was run on the `EAGLE` galaxies (Trayford et al., 2017) which allowed extensive comparisons to observations. The galaxies' mock observations were morphologically classified and it was found that there was a significant correlation between the photometric morphologies and the kinematic morphological estimators (Bignone et al., 2020).

The morphological estimators have been used to separate galaxies into early type, late type and irregular and their evolution over time in `EAGLE` has been well studied. Nevertheless, we think there are significant improvements to be made. In both hydrodynamical simulations and idealised models the full six dimensional position and velocity information is available. With full information we hypothesise that a robust technique can be developed that will be able to disentangle each galaxy component. We believe that an approach that uses clustering to find galaxy components in phase space is the way to this. This will allow us to study the mechanisms responsible for the build up of different components.

1.1.3 Finding Galaxy Components with Clustering

We have established that the stars in a galaxy's components move differently. This is formalised by the concept of a "Phase space," which is a space where each stars' orbit is represented by a point. For example two stars in the same location in phase space will be on the same orbit. Since the stars in different galaxy components move differently, each galaxy component populate a different region of phase space. Fortunately, there are several algorithms that are designed to find clusters of points in a phase space. By applying these clustering algorithms to stars in phase space we hope that each identified cluster of points is equivalent to a structural component. However, I describe below the many uncertainties associated to linking the numerical clusters to physical components.

When using clustering algorithms, the most impactful choice is that of the correct phase space. One quantity that encodes important orbital information is the ellipticity, $j_z/j_c(E)$, first introduced in Abadi et al. (2003). Ellipticity is the particle's angular momentum in the direction of the galaxy's total angular momentum (the z direction), j_z, normalised by the angular momentum of a particle with the same energy, E in a circular orbit, $j_c(E)$. Ellipticity contains useful information on orbits; e.g.

$$j_z/j_c(E) \begin{cases} < 0, & \text{counter-rotation} \\ = 0, & \text{no rotation} \\ > 0, & \text{co-rotation} \\ = 1, & \text{circular orbit} \\ > 1, & \text{unbound orbit.} \end{cases} \tag{5}$$

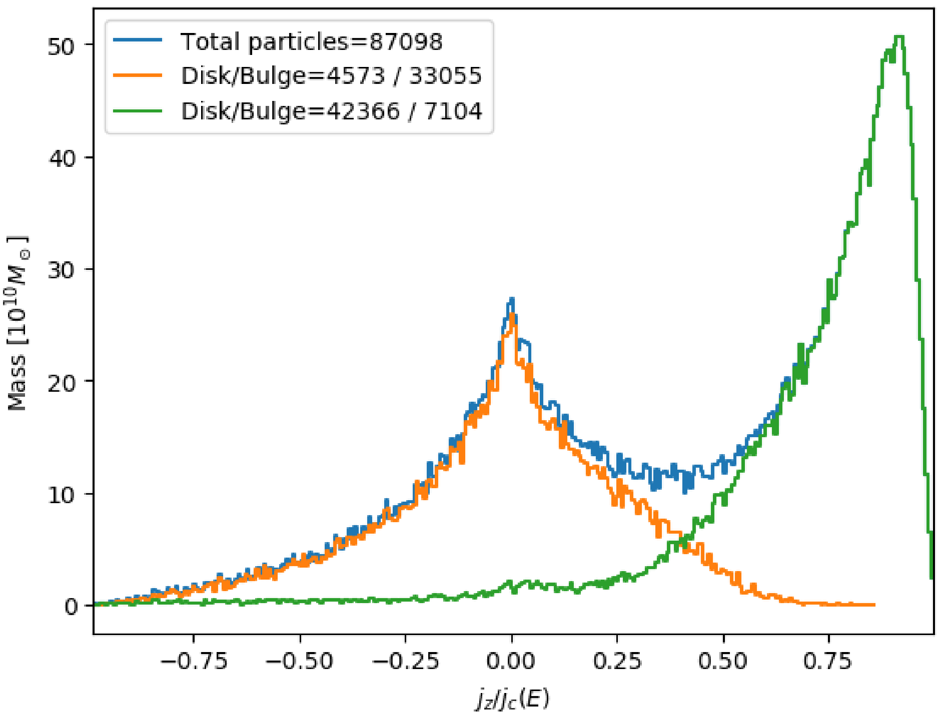

Figure 3: Two component Johnson's S_U distribution fit to $j_z/j_c(E)$ for the same 2 component `GalIC` galaxy as in Figure 4. Each star is probabilistically selected into components. The number of stars sorted into their correct component is almost as good as the three dimensional Gaussian mixture model.

See Figure 3 for an example of ellipticity for the stars in a two component galaxy.

The first attempt at galaxy decomposition using clustering by Doménech-Moral et al. (2012) used j_z/j_c and two other parameters, the perpendicular angular momentum $j_p/j_c(E)$, where $j_p = ||\vec{j} - j_z\hat{z}||$ is the angular momentum perpendicular to the disk, normalised by the circular angular momentum, and the binding energy, $e/|e|_{\max}$, normalised by the binding energy of the most bound particle, $|e|_{\max}$. A component with a large spread in $j_p/j_c(E)$ means that particles are not ordered into a disk. The binding energy describes the gravitational strength of the galaxy on the particle, which is a proxy for distance from the centre. For an example of this phase space and a decomposition see Figure 4. There has been no published literature for simulated galaxies exploring clustering phase spaces other than these three parameters.

The other most impactful choice in clustering is which clustering algorithm to use. The first attempt at galaxy decomposition with clustering in Doménech-Moral et al. (2012) used the simplest clustering algorithm, k-means. k-means tries to find the best location for each cluster's centre based on the distance to all the points. It is a fast algorithm, but has several limitations that the authors observed, for example it requires the number of clusters, n, to be known beforehand. The best method for deciding is for a human to look at the output and estimate the most realistic n. Doménech-Moral et al. (2012) decided that three clusters looked best, corresponding to thick disk, thin disk and spheroid and then the spheroid was to be further decomposed into bulge and halo after the algorithm was run.

The next improvement was in Obreja et al. (2016), who used the Gaussian Mixture Models (GMM) clustering algorithm to analyse the galaxies in the `NIHAO` simulations. GMM is more sophisticated than k-means, in that it tries to fit Gaussian probability distribution functions (PDF) to the data probabilistically. There is a fast and easy to use implementation in python in the package `scikit-learn` (Pedregosa et al., 2011). The probabilistic nature of this algorithm allows clusters to overlap, which is needed to describe realistic galaxies (Scannapieco et al., 2010), but does not overcome the same problem k-means had of requiring a human to interpret the most physical number of clusters (n). The PDF that it uses is a Gaussian so clustering with this method has the additional complication that it assumes components are Gaussian shaped in phase space, which is not necessarily true.

To address concerns about the number of components, Obreja et al. (2018) presented decompositions

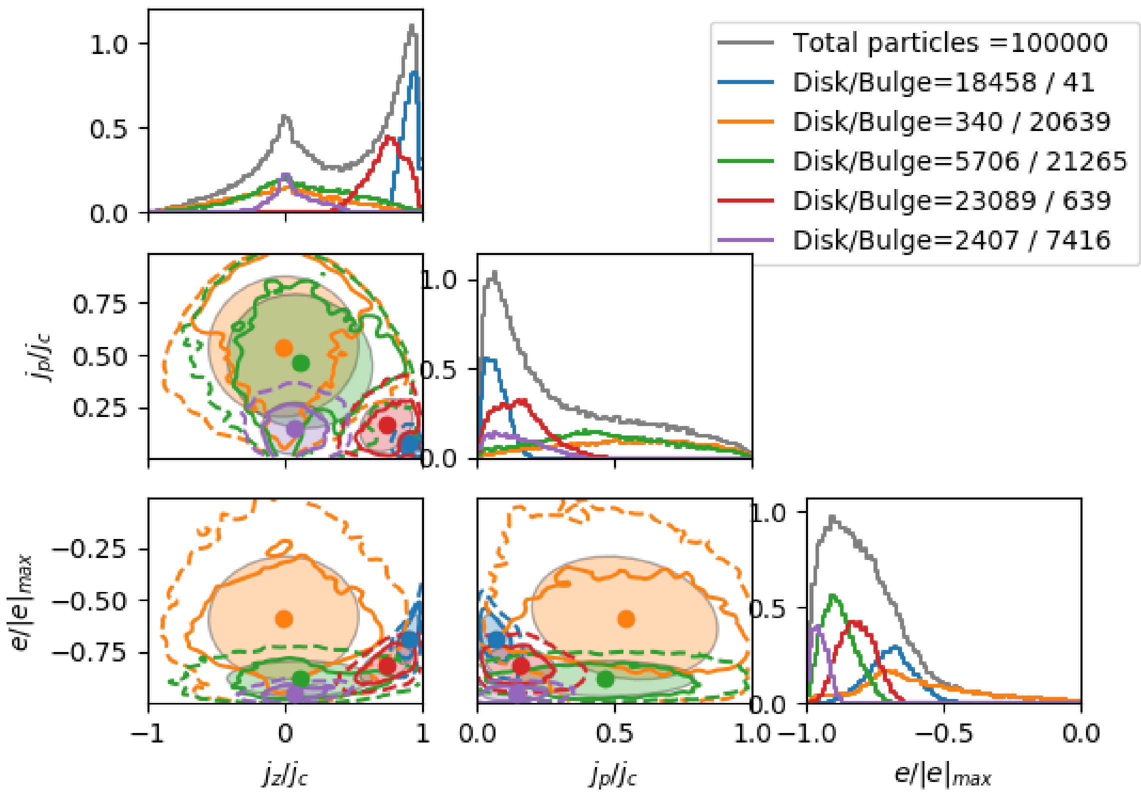

Figure 4: Corner plot of the phase space for the best fitting Gaussian Mixture Model for this `GalIC` galaxy. Each component's 68% and 95% contours are drawn as well as the oval which represents the projected shape of the three dimensional multivariate Gaussian that was fit to the data. The number of particles according to `GalIC` that belong in the disk and bulge that were classified into each component are displayed in the legend. The green bulgy component contains a large number of disk particles and the purple central rotating component has a strange mix. The orange halo component and the red and blue disk components both contain mostly they one population of stars.

for $2 \leq n \leq 5$ and suggest that depending on the application, a particular n could be chosen. The code was publicly released as `galactic structure finder` (`gsf`). In Du et al. (2019), the authors wanted to create decompositions for all the galaxies in the `IllustrisTNG` simulation so they introduced the Bayesian Information Criterion (BIC) as a method for automatically deciding the best number of components found with `gsf`. BIC is a probabilistic measure of goodness of fit that penalises fits with too many free parameters i.e. penalises decompositions with too many clusters. A smaller BIC means the model better describes the data without overfitting. The decomposition with the smallest number of clusters with a BIC below a threshold, C_{BIC}, is decided as the best decomposition (see Figure 5 for an example). Their method was tested by visually inspecting several galaxies from their sample. We think a more rigorous test is needed.

After the clustering analysis is completed, the next step is to interpret the physical meaning of the clusters that are found. It is straightforward for an astronomer to assign to each component a physical meaning from Table 1, but none of the clustering algorithms are taught anything specific about the motions of stars in galaxies. To assign labels to clusters automatically, Du et al. (2020) use the location in the $j_z/j_c(E)$ vs $e/|e|_{max}$ plane of all of the galaxy clusters to decide each cluster's label. If there were too many clusters, they would be combined until the galaxy had at most 4 components: thin disk, thick disk, bulge and halo. This interpretation led Du et al. (2020) to conclude that disk structures make up $\sim 55\%$ of the stars in galaxies at $z = 0$ in `IllustrisTNG` among other results. The next interesting step for their work is to analyse the time evolution of galaxy components.

Overall, the current state-of-the-art techniques need to be more thoroughly verified. The current

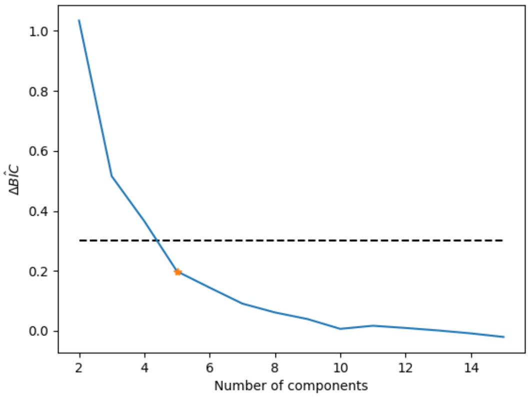

Figure 5: Plot of the normalised asymptotic Bayesian Information Criterion that was used to decide that 5 components were correct for Figure 4 as described in Du et al. (2019). For this selection $c_{BIC} = 0.3$. We remind the reader that by construction this galaxy has only 2 components, and hence the solution of 5 components is incorrect. Therefore, we cannot trust the Bayesian Information Criterion method to find the correct number of components.

numerical decomposition methods and they way they are linked to physical galaxy components have a number of possible improvements that need to be explored. A thorough method of testing the decomposition needs to be created and the current methods need to be tested along with any alternatives such as a decomposition on a different phase space or with a different clustering algorithm. Performing this structural decomposition on galaxies as they evolve over time has not been done and analysing such data will reveal new information about galaxy evolution.

1.2 Aims

The overall aim of my PhD is to devise a method to characterise galactic components and morphologies to examine how they evolve and to characterise the physical processes that drive their evolution. Below are a few more focused research questions that I aim to answer over the course of my PhD:

- What is the best method for identifying simulated galaxy components?
  - What is a good definition of a distinct galaxy structural component?
  - Can component identification be done using only kinematic data?
- How do galaxy components relate to morphology?
- How do galaxy components evolve over time?
  - How do centrals, satellites and isolated galaxies compare?
- How do simulated galaxy components compare to observations?
  - Can component identification techniques used for simulated galaxies also be used for observed ones?

To help answer these research questions, I will produce a catalogue of `EAGLE(-XL)` galaxy decompositions similar to the one Du et al. (2020) provided for `IllustrisTNG`. I aim to improve upon existing work by testing existing simulated galaxy decomposition techniques and then seeing if any new methods perform better.

1.3 Preliminary Results

Decomposition using GMM where the number of Gaussians is chosen by the BIC has not been demonstrated to work for a system with known structural properties. We want to demonstrate that the current techniques either work or need to be improved. To test this we have identified two criteria to evaluate a successful decomposition:

1 the number of clusters found is the same as the number of galaxy structural components and

2 the the particles found in a cluster belong to a single structural component

I used `GalIC` (Yurin and Springel, 2014) to generate 1 and 2 component galaxies and I then analyse them with my implementation of `galactic structure finder` (Obreja et al., 2018). My initial tests have revealed interesting results:

- Depending on the range of allowed C_{BIC}, the algorithm found 4 to 7 clusters were needed to describe a galaxy model generated with 2 structural components. This method failed the first criterion.
- When set to the known number of clusters, $n = 2$, the number of particles sorted into the correct component is $\approx 75\%$.
- The GMM requires that a galaxy be sampled with $\gtrsim 2000$ particles for the BIC to give consistent values. Drastically different results are found with $\lesssim 1000$ particles.
- Using the Du et al. (2020) prescription to combine the components into just a spheroid and a disk generally correctly identifies $\approx 80\%$ of particles. There was one case (out of 50 tests) where a component was associated with the incorrect component which led to a larger number of incorrectly tagged particles.

We conclude that this constitutes a failure of the BIC method for choosing the number of components. When the mixture model algorithm suggests that a seven-cluster fit is the best statistically, then we think that the method of combining components to give them physically meaningful labels is not well justified. Nevertheless, we do acknowledge that such a prescription is useful for comparing to previous studies. My investigation so far suggests that this failure is due to the components not being Gaussians nor sums of Gaussians in this phase space. This pushed us to try "non-Gaussian mixture models".

I am testing whether galaxy components can be better fit with a different PDF. To do this, I first rewrote the mixture model to work for arbitrary probability distributions. My algorithm currently only works for a 1-dimensional space and generalising it higher dimensions is proving to be complicated. While I was testing the algorithm, I found a trick that speeds up the algorithm so that it is now faster than GMM for larger numbers of components. The Johnson's S_U distribution was found to be a good fit to $j_z/j_c(E)$ for all the `GalIC` structural components. The distribution is parameterised

$$f(x; a, b) = \frac{b}{\sqrt{x^2+1}}\phi(a + b\log(x + \sqrt{x^2+1}), \tag{6}$$

where $\phi(x)$ is the pdf of a standard normal and a, b, are the two variables that modify the shape of the distribution. The whole fitting routine was applied to `GalIC` galaxies with $n = 2$ and found that $\approx 75\%$ of particles were sorted into the correct components. See Figure 3 for an example fit and Figure 6 for a projection of the stars that were sorted by the method. I am working on a method to automatically select the number of components with my non-Gaussian mixture model, but it looks like in this case the BIC is not appropriately penalising the number of clusters. The disk cluster found in the example Figure 6 shows a significant number of stars above the disk were misclassified which is leading us to investigate how to extend this decomposition to include disk height.

Further work will involve finding a theoretical distributions for different galactic components or developing a new PDF specifically for galaxy components. This approach is less like identifying arbitrary components – which was one of the advantages to this approach – and more like parametric fitting. This model fitting approach, however, could result in more physical components that are easier to label. Such an approach is similar to photometric decomposition.

Another current major avenue of inquiry is investigating different phase spaces and different clustering algorithms. The mixture model has advantages but it will always require a predetermined probability distribution function. Now that we have devised a simulated galaxy structural decomposition test, we have the tools to demonstrate the accuracy of any technique.

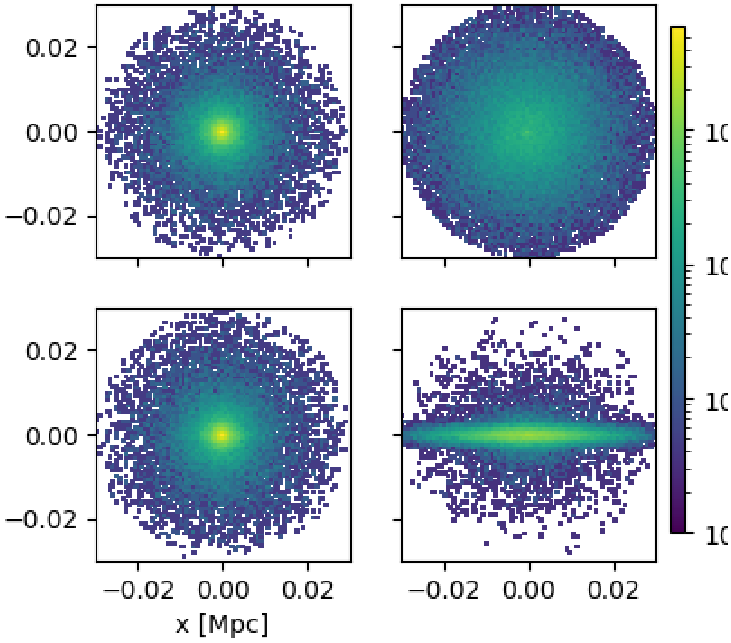

Figure 6: The projection in x and y (top) and x and z (bottom) for the two components found by the Johnson's S_U mixture model fitting. The disky component (right) has stars that are above and below the disk that are probably incorrectly identified bulge stars. Fitting that includes $|z|$ would probably improve the fit in this case.

2 Research Project Details

2.1 Confidential and Sensitive Information

I will be part of the team collaborating on the `EAGLE-XL` simulation. The full particle results of the runs will be privileged data not publicly released due to the large file sizes. I need to be careful that any copies of these data are password protected.

2.2 Intellectual Property Information

The `EAGLE-XL` code is being supplied to me by Dr Aaron Ludlow. When the paper describing the code comes out, I will be crediting it in all publications, for now the previous `EAGLE` simulation is described in Schaye et al. (2015). So far I have used the publicly available code `GALIC` (Yurin and Springel, 2014) and have adapted the public codes galactic structure finder (Obreja et al., 2018) and scikit-learn (Pedregosa et al., 2011) for my own research.

Astronomers often provide open source code. Open source licenses generally require that the authors of the code are credited in publications, so I need to be mindful about when I use code written by others and ensure the authors are properly credited.

2.3 Fieldwork Information

Most of my research will be conducted at the UWA campus. When the possibility of attending interstate or international conferences arises, I will abide by the GRS's guidelines.

2.4 Facilities

Most of my research will be conducted at the UWA campus. I will mostly be using the supplied laptop for writing code and doing basic analysis. I will used the Hyades computer cluster at ICRAR for simulation analysis and data storage. When more computer power is required I will obtain access to the

Research Skill	Current level	Evidence	Desired Level
Data Analysis and Writing Code	Competent	The work in my honours thesis and my PhD to date demonstrate ability to adapt and understand new concepts. I ideally want to produce similar high quality of work that takes both less time to write and less time to run.	Proficient
Academic Paper Writing	Basic	My honours thesis is basic academic writing. I want more paper writing experience that it becomes a more natural part of my job.	Proficient
Academic Presentations	Basic	I have given talks in group meetings and presented papers at journal club. I would like to be more confident and be able to communicate my points more effectively.	Proficient
Collaboration	Basic	Being a part of ASTRO-3D Genesis simulation group for the past 3 months has shown me that my work can have more impact if I participate more.	Competent
(Super-)Computer Skills	Basic	I have run code on computer clusters both in undergrad and postgrad. There are certainly techniques and tricks that I don't know that I know.	Competent
Writing Proposals for Computer or Observation time	None	Part of being a top astronomer is winning competitive time. I want to be involved in asking for more resources if needed.	Basic

Table 2: Skill audit table. I have some experience in most of the skills I think are important for a career as an astronomer but I want to improve in all areas.

supercomputers at Pawsey (Western Australia) or use my access to the supercomputers at the National Computer Infrastructure (NCI Australia) depending on need, availability and permissions.

2.5 Statistical Component

I have some previous experience applying scientific statistics and can learn new methods from textbooks or papers when I need to. As part of the rules for my degree I will be taking the master's level course "Astro-statistics".

2.6 Skill Audit

See Table 2. In general I have some basic skills that I would like to improve over the course of my degree.

2.7 Research Project Communication

I will present my research proposal to a panel of 3 ICRAR researchers on the 15$^{\text{th}}$ of July 2020. I will have yearly reviews which are adjudicated by the same panel in late January 2021-2023.

To communicate my work to the wider astronomical community, I plan to publish approximately a thesis chapter worth of work once it is of an appropriate standard. Ideally this will be done at a pace of approximately 1/year. From this point my proposed chapters will be:

1 **Automatic Galaxy Component Identifier.** This paper will investigate the performance of the contemporary automatic methods for identifying galaxy components on known idealised solutions. If successful, I will publish my adjusted phase space clustering component identification method and provide justification for its wider adoption. The paper should also include a detailed example galaxy from `EAGLE`.

2 **The Evolution of Galaxy Components in Galaxies in `EAGLE(-XL)`.** This paper will investigate the properties of the components (identified with the best method found in my first paper) in galaxies over time. We will to quantify how the different components are built up in different environments and through mergers. We will like to separate galaxies based on morphology and see if any physical differences can be attributed to its initial galaxy components for example.

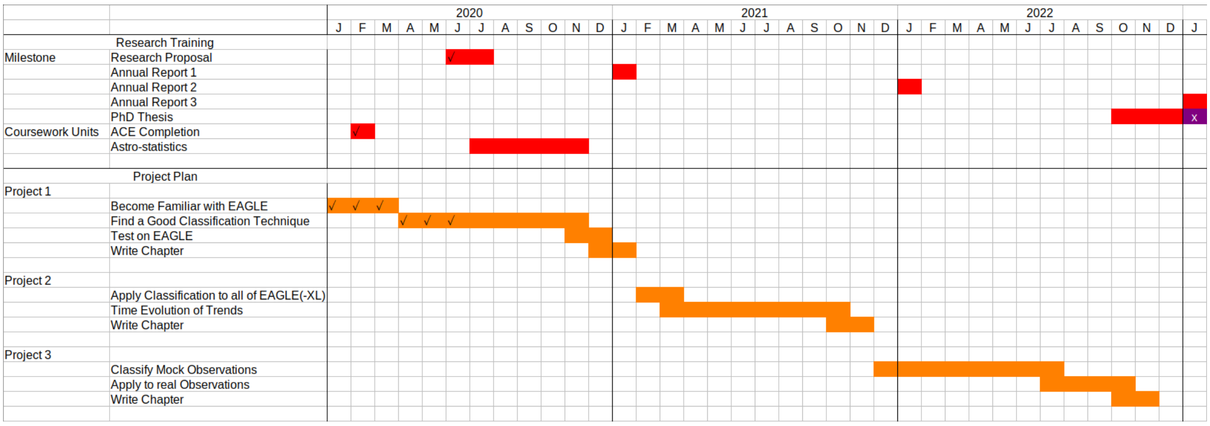

Figure 7: Gantt chart detailing the projects in my PhD. I'm giving my first project the most amount of time partly because of a loss of productivity because of covid-19 but mostly because my findings will directly influence the details of the later projects.

3 **Observations of Galaxy Component Evolution.** Assuming that the galaxy component identifier proves to be a useful tool for describing galaxies, we will investigate if the component identifier can be adapted for integral field spectroscopy. We will investigate the similarities and differences between real and simulated galaxies when compared in a like to like way.

Once I have some results, I will also present my work, either with a talk or a poster at one of the several regular astronomy conferences. Conferences that I have attended this year and intend to present at in the future include: ESOz, the joint Australia-European Southern Observatory conference which focused on galaxies across the whole electromagnetic spectrum, partially in preparation for the new Australian Square Kilometre Array radio telescope. The Astronomical Society of Australia holds a yearly conference which covers all research done in Australia and includes many student talks.

2.8 Approvals

My project encounters no foreseeable ethical issues, hazardous material or anything else that would require GRS approval. If that changes the GRS will be notified appropriately.

2.9 Data Management

Code that I write will be stored both on my laptop supplied by the university and in a personal git repository. While in the process of analysing data the code will be private, but when results are published I will make the code public to allow my results to be repeatable. My git links are: `https://gitlab.com/Matt_W` and `https://github.com/Matthew-Wilkinson`.

Data sets will be stored on my laptop. Data that I am currently analysing will be also kept on the Hyades computing cluster maintained at ICRAR.

2.10 Research Project Plan

See Figure 7 for the proposed timeline of my PhD.

3 Research Training

3.1 Research Training Plan

The Gantt chart provides a timeline of my planned research training.

3.2 Confirmation of Candidature

For my confirmation of candidature, I will be required to complete the tasks listed here with expected due dates.

Item Description	Estimated Cost	Funds Available	Funding Source
Laptop Computer	$1,250	$1,250	ECM
Additional IT Supplies	$500	$500	ICRAR
Domestic Travel	$1,500	$1,000 $500 (subject to approval)	ASA ICRAR ASTRO-3D
International Travel	$3,000	$1,850 $1,150 (subject to approval)	GRS ICRAR ASTRO-3D
Total	$6,250	$6,250	-

Table 3: Budget break down by item and funding sources. ASTRO-3D gives students funding for travel if they apply.

- AACE1000 Academic Conduct Essentials: March 2020 (completed)
- Project proposal report: Due 26th July 2020
- Project proposal presentation: Scheduled 15th July 2020
- Coursework to a total of 6 points: Astro-statistics to be taken from July 2020 to November 2020
- Annual progress report: Due 26th January 2021, 2022 and 2023
- Annual progress seminar: To be presented in January in 2021, 2022 and 2023

3.3 Working Hours

My usual working hours are 9am - 5pm, Monday - Friday with variations as required to total at least 37.5 hours a week.

4 Budget

See Table 3 for budget break down. I have already obtained the laptop.

5 Supervision

Principal Supervisor: Dr. Aaron Ludlow(50%) Dr. Ludlow and I will meet on a weekly basis (usually at the same time as Dr. Lagos). Dr Ludlow is an expert on cosmological and galaxy simulations. He has extensive knowledge of what is possible to know from the simulations as well as technical coding expertise.
Coordinating Supervisor: Dr. Claudia Lagos(50%) Dr. Lagos and I will meet on a weekly basis (usually at the same time as Dr. Ludlow). Dr. Lagos is an expert on galaxy evolution and galaxy simulations. She provides invaluable insights into the types of tests and measurements that lead to a deeper understanding.

References

Abadi, M. G., Navarro, J. F., Steinmetz, M., and Eke, V. R. (2003). Simulations of Galaxy Formation in a Λ Cold Dark Matter Universe. II. The Fine Structure of Simulated Galactic Disks. *ApJ*, 597(1):21–34.

Bignone, L. A., Pedrosa, S. E., Trayford, J. W., Tissera, P. B., and Pellizza, L. J. (2020). Non-parametric morphologies of galaxies in the EAGLE simulation. *MNRAS*, 491(3):3624–3642.

Cappellari, M. (2017). Improving the full spectrum fitting method: accurate convolution with Gauss-Hermite functions. *MNRAS*, 466(1):798–811.

Clauwens, B., Schaye, J., Franx, M., and Bower, R. G. (2018). The three phases of galaxy formation. *MNRAS*, 478(3):3994–4009.

Correa, C. A., Schaye, J., Clauwens, B., Bower, R. G., Crain, R. A., Schaller, M., Theuns, T., and Thob, A. C. R. (2017). The relation between galaxy morphology and colour in the EAGLE simulation. *MNRAS*, 472(1):L45–L49.

Correa, C. A., Schaye, J., and Trayford, J. W. (2019). The origin of the red-sequence galaxy population in the EAGLE simulation. *MNRAS*, 484(4):4401–4412.

Cortese, L., Fogarty, L. M. R., Bekki, K., van de Sand e, J., Couch, W., Catinella, B., Colless, M., Obreschkow, D., Taranu, D., Tescari, E., Barat, D., Bland-Hawthorn, J., Bloom, J., Bryant, J. J., Cluver, M., Croom, S. M., Drinkwater, M. J., d'Eugenio, F., Konstantopoulos, I. S., Lopez-Sanchez, A., Mahajan, S., Scott, N., Tonini, C., Wong, O. I., Allen, J. T., Brough, S., Goodwin, M., Green, A. W., Ho, I. T., Kelvin, L. S., Lawrence, J. S., Lorente, N. P. F., Medling, A. M., Owers, M. S., Richards, S., Sharp, R., and Sweet, S. M. (2016). The SAMI Galaxy Survey: the link between angular momentum and optical morphology. *MNRAS*, 463(1):170–184.

Doménech-Moral, M., Martínez-Serrano, F. J., Domínguez-Tenreiro, R., and Serna, A. (2012). Formation of galaxies in Λcold dark matter cosmologies - I. The fine structure of disc galaxies. *MNRAS*, 421(3):2510–2530.

Du, M., Ho, L. C., Debattista, V. P., Pillepich, A., Nelson, D., Zhao, D., and Hernquist, L. (2020). Kinematic Decomposition of IllustrisTNG Disk Galaxies: Morphology and Relation with Morphological Structures. *ApJ*, 895(2):139.

Du, M., Ho, L. C., Zhao, D., Shi, J., Debattista, V. P., Hernquist, L., and Nelson, D. (2019). Identifying Kinematic Structures in Simulated Galaxies Using Unsupervised Machine Learning. *ApJ*, 884(2):129.

Hubble, E. P. (1926). Extragalactic nebulae. *ApJ*, 64:321–369.

Lagos, C. d. P., Stevens, A. R. H., Bower, R. G., Davis, T. A., Contreras, S., Padilla, N. D., Obreschkow, D., Croton, D., Trayford, J. W., Welker, C., and Theuns, T. (2018). Quantifying the impact of mergers on the angular momentum of simulated galaxies. *MNRAS*, 473(4):4956–4974.

Lagos, C. d. P., Theuns, T., Stevens, A. R. H., Cortese, L., Padilla, N. D., Davis, T. A., Contreras, S., and Croton, D. (2017). Angular momentum evolution of galaxies in EAGLE. *MNRAS*, 464(4):3850–3870.

Obreja, A., Macciò, A. V., Moster, B., Dutton, A. A., Buck, T., Stinson, G. S., and Wang, L. (2018). Introducing galactic structure finder: the multiple stellar kinematic structures of a simulated Milky Way mass galaxy. *MNRAS*, 477(4):4915–4930.

Obreja, A., Stinson, G. S., Dutton, A. A., Macciò, A. V., Wang, L., and Kang, X. (2016). NIHAO VI. The hidden discs of simulated galaxies. *MNRAS*, 459(1):467–486.

Obreschkow, D. and Glazebrook, K. (2014). Fundamental Mass-Spin-Morphology Relation Of Spiral Galaxies. *ApJ*, 784(1):26.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12(null):2825–2830.

Pillepich, A., Springel, V., Nelson, D., Genel, S., Naiman, J., Pakmor, R., Hernquist, L., Torrey, P., Vogelsberger, M., Weinberger, R., and Marinacci, F. (2018). Simulating galaxy formation with the IllustrisTNG model. *MNRAS*, 473(3):4077–4106.

Scannapieco, C., Gadotti, D. A., Jonsson, P., and White, S. D. M. (2010). An observer's view of simulated galaxies: disc-to-total ratios, bars and (pseudo-)bulges. *MNRAS*, 407(1):L41–L45.

Schaye, J., Crain, R. A., Bower, R. G., Furlong, M., Schaller, M., Theuns, T., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J. C., Jenkins, A., Rosas-Guevara, Y. M., White, S. D. M., Baes, M., Booth, C. M., Camps, P., Navarro, J. F., Qu, Y., Rahmati, A., Sawala, T., Thomas,

P. A., and Trayford, J. (2015). The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *MNRAS*, 446(1):521–554.

Sérsic, J. L. (1963). Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Boletin de la Asociacion Argentina de Astronomia La Plata Argentina*, 6:41–43.

Sérsic, J. L. (1968). *Atlas de Galaxias Australes.*

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., and McConnachie, A. W. (2011). A Catalog of Bulge+disk Decompositions and Updated Photometry for 1.12 Million Galaxies in the Sloan Digital Sky Survey. *ApJS*, 196(1):11.

Taranu, D. S., Obreschkow, D., Dubinski, J. J., Fogarty, L. M. R., van de Sande, J., Catinella, B., Cortese, L., Moffett, A., Robotham, A. S. G., Allen, J. T., Bland-Hawthorn, J., Bryant, J. J., Colless, M., Croom, S. M., D'Eugenio, F., Davies, R. L., Drinkwater, M. J., Driver, S. P., Goodwin, M., Konstantopoulos, I. S., Lawrence, J. S., López-Sánchez, Á. R., Lorente, N. P. F., Medling, A. M., Mould, J. R., Owers, M. S., Power, C., Richards, S. N., and Tonini, C. (2017). Self-consistent Bulge/Disk/Halo Galaxy Dynamical Modeling Using Integral Field Kinematics. *ApJ*, 850(1):70.

Trayford, J. W., Camps, P., Theuns, T., Baes, M., Bower, R. G., Crain, R. A., Gunawardhana, M. L. P., Schaller, M., Schaye, J., and Frenk, C. S. (2017). Optical colours and spectral indices of z = 0.1 eagle galaxies with the 3D dust radiative transfer code skirt. *MNRAS*, 470(1):771–799.

Trayford, J. W., Frenk, C. S., Theuns, T., Schaye, J., and Correa, C. (2019). The star formation rate and stellar content contributions of morphological components in the EAGLE simulations. *MNRAS*, 483(1):744–766.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., and Hernquist, L. (2014). Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *MNRAS*, 444(2):1518–1547.

Vogelsberger, M., Marinacci, F., Torrey, P., and Puchwein, E. (2020). Cosmological simulations of galaxy formation. *Nature Reviews Physics*, 2(1):42–66.

Yurin, D. and Springel, V. (2014). An iterative method for the construction of N-body galaxy models in collisionless equilibrium. *MNRAS*, 444(1):62–79.