General Questions

1. Can you summarise your PhD in 5 minutes?
2. Did you enjoy your PhD?
3. What do you think the main result and impact of your thesis will be?
4. If you had to describe your thesis in a tweet, what would you say?
5. What are your plans for the future?
6. If you had the time, how would you follow up on this project?
7. What were the main research questions you were hoping to address?
8. What does Stephen Serjeant do? What work has he authored?
9. How does your work relate to the literature?
10. What is the Effectiveness of your Research?
11. What are the Limitations of this Thesis?
12. Are there existing similar works in the literature?
13. How has your research challenged or changed a topic?
14. Other approaches in your research?
15. What are the next steps for this area of research?

Abstract

1. Describe what is meant by hierarchical model.
   1. This is a model where smaller components combine and merge into larger ones. This happens over many times in a hierarchy of merging and accretion. In the context of the abstract, I mean LCDM which has within it a hierarchical model of galaxy evolution.
   2. I do not mean hierarchical observations. I mean that ‘hierarchical models’ and theories confirmed by observations.
2. What do you mean by the relation between physical processes and underlying parameters?
   1. An underlying process, in this context, is some process that leads from A to B. For example, star formation is an underlying process in a galaxy. These are then related to underlying parameters, for example the gas mass within the galaxy. Both can be parameterized, but one is describing movement of one thing to another while the other is simply a descriptive number.
3. Explain a Bayesian Convolutional Neural Network. What are its component parts?
   1. A Bayesian Convolutional Neural Network in this context is a deep neural network which is applied to images. A convolutional neural network takes an image, and breaks it down into smaller and more abstract parts. It passes each part, a representation, through many different layers which is altered by a weight which can be applied in a range of functions. These are passed through layer after layer until the final abstraction is then interpreted by a classification layer. However, in a Bayesian CNN, each weight is not simply a number which has been learned, it is a prior distribution of some kind. As we move through the BCNN, each distribution gives each abstraction a posterior distribution that can then be output in the final classification stage. Thus, we get our final answer but also a level of uncertainty within it.
4. You talk about active galactic nuclei here, what is it? Why does it matter with interaction?
   1. An active galactic nucleus is, essentially, the supermassive blackhole at the galactic centre which is growing. It has got an accretion disk around about it, which is incredibly hot, emitting high powered radiation and jets perpendicular to the plane of accretion. The high powered emission acts as a wind which goes through the galactic disk, driving outflows and causing ionization. This, in turn, leads to the cessation of star formation in the galaxy. Thus, the SMBH at the galactic centre is linked to the evolution of the galaxy as a whole. We call this co-evolution, and we see a distinct match between black hole mass and a host of galaxy properties. Most closely, the bulge and stellar mass of the galaxies.

One way in which AGN could be ignited is through galaxy interaction, and the rush of gas into the galactic core being the perfect way for material to accrete onto the black hole.

1. What is the dynamical timescale of an interaction? Does it have well defined beginnings and endings?
   1. The dynamical time, or the way I describe it, begins at the moment two galaxies are within each others dark matter haloes and dynamical friction is beginning to occur. It then ends when either the two galaxies have merged or they have escaped each others halos.
2. What is a confidence interval?
   1. A confidence interval is a measure of uncertainty about a defined parameter using Bayesian statistics. In a confidence interval, the true value of the parameter could be anywhere. It is often described by the percentiles about a value defined by assuming a statistical distribution. I’m worried that, in fact, I do mean a credible interval in this work and not a confidence interval.

A credible interval is measured using the mass of walker steps, as is done in my final Chapter. However, I have drawn the percentiles which dictate the different confidence interval, not the credible interval.

I have used credible intervals, not confidence intervals. Therefore, I need to re-write this part.

1. You mention Simulation Based Inference. What is that?
   1. Simulation based inference is a potential answer to the high computational cost of the MCMC approach described in this work. While I won’t really go into depth here, it involves front-loading the computational cost. We run our simulation across the parameter space, creating thousands of examples of images. These are then learned by an emulator or machine learning algorithm. This emulator then approximates the full probability space. It then can be applied to an observation, and recover the parameters of the observation.

Introduction

1. You talk about large scale structure, what do you mean? (page 1)
   1. The large scale structure is the overall filamentary structure we observe in the Universe. We get these nice redshift maps which show the filamentary structure of the universe on a grand scale. At a smaller scale, we then get over- and under-dense environments. Galaxy clusters, nodes, voids and the field. Each of these are made up, fundamentally, of galaxies.
2. Describe cosmic time to me. (page 1)
   1. Cosmic time describes the evolution of galaxies and the Universe’ age. It is from the Big Bang until now, and the Universe has expanded, changed and evolved with it. Cosmic time goes with the Hubble Flow, and assumes that the Universe is the same density everywhere at all times.
3. Can you fully elaborate on what you define as mutual interaction? (page 1)
   1. Mutual interaction is when two galaxies have a gravitational effect upon one another. This can be from dynamical friction, to the complete disturbance of their disks via a flyby.
4. Can you explain redshift? (page 1)
   1. Redshift is a result of light moving through an expanding Universe. As the universe expands, the distance between coordinates increases. This leads to a stretching in the wavelength of the light – a reddening effect of it. When applied to a galaxy, we see the internal spectrum being shifted by some amount in wavelength. As we know the true spectrum, and the speed of light, we can convert the redshift measurements into distances to the galaxies.
5. Define what you mean by a tidal feature. (page 1)
   1. A tidal feature is a structure in or about a galaxy which has been created by the effect of another galaxies gravity. For instance, as two galaxies flyby each other, they can shear off the outer regions of a galaxy and form a tidal tail and arm.
6. How else could we make this reliable identification of interacting systems? (page 2)
   1. Really, what we need is some measure of the distance towards them to show that they are truly at the same distance from us. We could potentially try to model them with simulations, and recreate their morphology. However, this may not be definitive. We could look for signs of increased star formation and tidal disruption. However, if we’re getting that we might as well measure the redshift of the system. Same with the velocity. What we really do need is the redshift more than anything.
7. What do you mean by exploring the relationship between parameters and characteristics of systems? (page 2)
   1. It’s sort of as I’ve said before. Interacting systems are characterized by increased star formation rates, potentially increased AGN fractions and tidal disruption in their disks. These are physical processes occurring within the galaxies that lead to these differences. I want to then map out the parameters which make this happen. Whether that be how long the two galaxies have been interacting for, the stellar masses of the systems, and the orientation angles of the interaction.
8. What is a super massive black hole? Why are they often at the centre of galaxies? (page 2)
   1. A supermassive blackhole is a black hole with a mass greater than 10^5 solar masses. Significantly larger than a stellar black holes (5 – 10 times solar mass). They are often, but not always, found at the centre of the galaxy. This is due to dynamical friction, where large mass objects sink in the gravitational potential due to friction with lots of smaller mass objects. Thus, we do no know if galaxies formed about SMBHs, or if galaxies formed with SMBHs within them, that then began to move into the galactic core due to dynamical friction.
9. Why may there not be supermassive black holes at the galactic centre? (page 2)
   1. This is a bit of a tricky one. The paper I’ve cited here is not saying there is no black hole at the centre of M33, but that using the known co-evolution arrangements between central velocity dispersion and black hole mass yields a mass between 0 and 1,500 solar masses. Therefore, there is something there, but there is clearly a space at the lower end of the co-evolution relation which we do not yet understand. They also do not really comment on why one might not be present. One could be there used to be many there, and they merged / interacted and ejected each other. However, these authors do not comment on this.
10. Could you describe the different components of the galactic disks. For instance, what is the thick and thin disk? (page 3)
    1. There are two components to the disk: the thin and thick disks. The thick disk is the oldest component of the disk, comprised of about 5% of the stars in the overall disk. It is much older than the thin disk and, surprisingly, much thicker: being about 5kpc in height. How this component forms is debated, but could be from interactions, or a later formation mechanism. The thin disk comprises 95% of the stars in the galactic disk, and is considerably younger than the thick disk. It is only about 600pc in height. This component likely formed through secular accretion over time.
11. What are the classifications for bulges? What is the difference between them? (page 3)
    1. There are many different kinds of bulges: classical, peanut, boxy, buckled, etc. However, the two main ones I talk about here are classical and pseudobulges (essentially the rest). A classical bulge is exactly what it sounds like, formed though interaction events and harassment thermalizing the galactic core. This is spherical, and composed of older bulges. Pseudobulges are formed via secular processes, internal to the galaxy. They are a result of the orbits of the stars within the centre of the galaxy in an ordered fashion, rather than the dispersion dominated classical bulges.
12. Could you just talk about galactic bars for a second? How do they form? What do they do?
    1. Bars are large, resonant structures in the centre of galaxies. They are composed primarily of stars and gas – which is being funneled within it into the galactic core. They form by a density wave radiating from the galactic core, and the wrapping of it around the fluid-like disk leaves to the build up of a bar over a long period of time.
13. Can you name the different galaxies? What is the primary differences between them? What about their evolutionary pathways, what can you tell me about them? (page 3)
    1. In this thesis, I primarily break down galaxies into three classifications: disk, elliptical and lenticular. Disk galaxies are shown as S(a-c) and Sb(a-c) on the Hubble tuning fork. These are large, young stellar systems which often have star formation occurring within them. They are observed to be blue. Elliptical galaxies are then E(0 - 9), where the number denotes their level of ellipticity. They are often older systems, that have been highly disturbed in a merger where the previous disk has been completely thermalized. They are red. Finally, there are lenticular galaxies. Initially, these were classified as at an evolutionary stage between elliptical and disk galaxies. However, it turns out the truth is more complicated. They are incredibly compact objects, and have highly irregular morphologies. They are often gas poor, and red, but exhibit a disk. Lots of work is still ongoing into their origins.
14. You say that the instabilities in the early Universe were caused by “physical processes”, what do you mean specifically? Why were they caused by the small scale of the Universe? (page 4)
    1. Physical processes here refers to Gaussian density fluctuations throughout the early universe. If you assume an initial condition of a box of uniformly distributed gas, which is very hot, and that is expanding in every direction, you eventually get areas of slight over- and under-densities. Due to the small size of the Universe at this time, these over-densities lead to gravitational accretion of objects.
15. The Universe was expanding. Why? (page 4)
    1. Well, that is certainly the question. You would expect that with lots of baryonic and dark mass within the Universe, that it would be slowing down due to gravity. However, what we see is an expanding Universe and the need for dark energy which is causing the acceleration. We assume that dark energy, or the energy density of the Universe, is constant despite expansion. Hence, it has began to dominate over gravity. However, we know it comes out of our normal cosmological equations as Einstein’s constant. This represented the energy density of the Universe.
16. Why did the rate of mergers and galaxy interaction go down? Why is it lowest now compared to the past? (page 5)
    1. Simply, I’d assume, the Universe was smaller. With a smaller total volume of the Universe, the probability that two galactic halos would merge was significantly higher. As galaxies had just formed, they were small mass accreting systems. As some systems merged with others to form larger, more massive systems they gravitationally pulled in more systems and had a run away effect of higher merger rates.

However, as the universe expanded and many systems had coalesced into larger, massive systems, the probability of interaction and merging decreased.

1. As above, but for cosmic star formation. Why is it lowest now compared to the early Universe? (page 5)
   1. Well, that is another open question. I want to say that it’s because the merger rate is declining, however there are many works in the literature that dispute how much of a driver merging actually was for cosmic star formation (see Schawinski, see Kaviraj).
2. You say that mergers play a significant role in star formation of the past, but don’t really back this up. Why do you say this? (page 5)
   1. This is primarily from works such as Kaviraj et al. (2016) and even some works from Sarah Ellison which are showing that actually the enhancement in interaction and merging we see is not enough to account fully for the large difference in cosmic star formation at early times. There are a few other works that say this. This comes from measuring the enhancement in major and minor merging galaxies out to about redshift 3. Thus, they postulate that other, secular processes could be driving the cosmic star formation rate at these times.

The secular processes that dominate this are the balance between gas accretion and feedback processes, such as by supernovae and AGN. (A lot of this came from Madau & Dickinson, 2014).

1. You say that elliptical galaxies are purely an effect of mergers. Are there not secular processes which lead to the formation of ellipticals? (page 7)
   1. Initially, they were thought to form in an initial burst of star formation at redshift greater than 5. Thus, giving rise to a tight relation between their stellar mass and internal velocity dispersion. However, this is unlikely to be true. This monolithic view of their formation was also helped by the finding that they likely formed in a very short timescale.

However, with further observations of signs of recent star formation and numerical simulations being able to replicate elliptical galaxies through merging of like-massed disk galaxies, a secondary avenue has been opened into their likely origin. With gas rich mergers, the fundamental-plane relation is able to be reproduced.

1. Can you further expand on what a stellar population is? (page 7)
   1. A stellar population, in this context, is a set of stars that are at the same age and described by one initial mass function. I can go into what the IMF is later, but when we take this group that is formed and know the number counts of stars that will be formed of different stellar types. This group, or population, will then have a flux distribution we can predict using stellar population synthesis codes and mode.
2. How does a stellar population relate to the star formation rate? (page 7)
   1. They relate in a bit of a strange way. So, to find the flux distribution of a stellar population, we need to make assumptions about the metallicity of the population and the star formation rate. For instance, in the one I use here (a Bruzual and Charlot 2003 SSP), we assume that all the stars are formed in a delayed exponential burst. We account for the mass formed in this sudden burst and distribute it through our IMF. This then informs the flux distribution of the resultant population. As the population ages, different classes of stars begin to die. The most obvious of these are the OB-type stars, which die after a few million years. This significantly alters the flux distribution as the lower mass FGK stars begin to dominate the light profile.

So, the star formation rate relates to our SSP model as it dictates the total mass in each stellar ‘bin’ in the underlying population. In our model, we have then assumed that the star formation rate exponentially declines with time, allowing the FGK stars to dominate.

1. Can you describe the process of gas cloud fragmentation into forming stars? (page 7)
   1. First, we have a large gas cloud close to or at a critical mass and density. This critical mass and density is the Rayleigh Jeans criterion. This defines the length and mass at which a gas cloud must be for self-gravity to overcome the stability of the cloud, and it begins to collapse. This collapse begins to increase the density within the cloud, which increases the self-gravity and increases the collapse. As enough gas collapses to the core, nuclear fusion and ignition of the core begins: forming a star.

However, there often must be something to cause the gas to overcome the Rayleigh-Jeans criterion. Often, this is a shockwave from a nearby supernova. This shockwave compresses the gas and, temporarily, causes different parts of the cloud to overcome the RJ criterion and collapse. This process, of a larger cloud having multiple areas of it overcome the RJ criterion, is called fragmentation. Thus, we often form many stars from a single massive gas cloud.

In a galaxy interaction, the shockwave is caused by the collision of gas clouds. As gravitational torques cause the gas to lose angular momentum and actually move around, they begin crashing into each other. This leads to shockwaves, density waves and over densities. These then collapse rapidly to form stars.

1. Does the ultra-violet fall in our blue filters? (page 7)
   1. Well, it depends on the system. Often no, the ultraviolet is it’s own category of filter called u or U. However, depending on the redshift, UV quickly enters the blue filter. I see what you mean though, my text may not be strictly true here. The U band is very high frequency wavelengths.
2. Can discuss further the different states of gas. Why is it only molecular gas that can fragment and form stars? (page 9)
   1. It’s all to do with the thermal energy within the gas itself, I’d assume. As I’ve talked about the Jeans criterion, and requiring a size of cloud with a certain mass to collapse, this can be offset by higher thermal energy. I.e. the force of gravity trying to get the gas to collapse has more force to overcome. Also, if the particles are moving too fast (i.e., ionized) they simply won’t be affected by self-gravity and collapse in over-dense regions.
3. How was the Kennicutt-Schmidt relation found? Why was the number found to be n = 1.3? (page 9)
   1. Simply, they had to measure the star formation rates in different galaxies using tracers. We’ve already sort of discussed how this is achieved. The two main ones are with integrating over the ultraviolet measurement to capture the emission of high powered, young OB-type stars, measuring the distribution of ionized Hydrogen and then measuring in the Far Infrared emission.

The first is to find lots of these young stars, where the SFR has been found to scale linearly with luminosity.

The second is to measure the ionized hydrogen in the galaxy, and quantify that. This is because these hydrogen clouds will be the area where stars are born – nebulae – but with the stars inside them causing ionization of the gas. However, this is very prone to contamination by other ionization sources.

Finally, we look in the far infrared, where we can detect the emission from ionization of the ISM which is being absorbed by the dust in the system. Usually, we would combine the UV and FIR in one equation to calculate the SFR in the system, with the Hydrogen then being a confirmation.

We also can measure the ratio in the 4000 angstrom break, which is directly related to the SFR of the galaxy.

1. If, like you said, ionized hydrogen can’t be use in SF, then how do we use it to measure the SFR? (page 10)
   1. It acts as a tracer of SFR through the galaxy. As new stars are formed, and OBA-type stars exist they have the power to ionize the surrounding gas. This is a temporary affair, and therefore the H\alpha eventually becomes molecular again. So, if H alpha is present, we know SF must be occurring.
2. Could you speculate about the evolutionary pathway of red spirals and blue ellipticals? (page 11)
   1. Red spirals are an interesting phenomenon, where they are likely drained and quenched of their gas via many different processes. Interestingly, in the Galaxy Zoo work done on these they are often found in cluster environments, close to the cores. Thus, it is likely by ram pressure stripping they may have lost their gas.

Blue ellipticals are a different story. Often, when observed with more powerful, higher resolution telescopes, we do actually observe a disk component to them. Some had some kind of AGN component associated with them, which could be driving star formation as well.

1. You say we measure the SFRs of these galaxies. Could you explain how we do this? (page 11)
   1. Sure, so, there are a few different ways to do this. We know that a few line emissions have a direct relationship to the SFR that we can use as an estimate. For instance, as I’ve spoken about above: the Hydrogen Alpha line. This line is produced by high mass stars which have been formed and then ionizing the surrounding Hydrogen gas. We can measure this emission line and then integrate it through out filter and convert it to a luminosity. With said luminosity, we can use well known SFR to luminosity relations which are derived from synthesis models and the assumed IMF.
2. Can you confirm your definition of a filament galaxy? (page 11)
   1. In this context, I’ve referred to a filament galaxy as one where the environmental density is above that of a field, but the galaxies are not in an over-dense cluster environment. Filmaments form the backbone of the large scale structure of the Universe, and are the webs that connect to the nodes with high over densities. Along filaments, gas and galaxies flow. Thus, this is a different environment from those galaxies in the field in total isolation and cluster galaxies.
3. How would you define these different environmental classifications? (page 11)
   1. A field galaxy – A galaxy in total isolation. Not necessarily in a void, but with almost no near neighbours.
   2. A filament galaxy – A galaxy with a few neighbours, but not in an overdense region like a cluster. It is co-moving with it’s neighbours, but not bound to them.
   3. A cluster galaxy – A galaxy in a larger galactic structure with many members. Often have very many neighbours.
4. Can you describe the parameters you’re referring to that change interaction? (page 12)
   1. Ofc, the main two we know that change how an interaction turns out is the masses of the systems and the orientation of the galaxies as they interact. The masses lead to different results for both the galaxies – from complete destruction to just tidal disturbance. The orientations then dictate the tidal features that will form for various reasons.

There are a few other parameters: the relative sizes, the velocity of the encounter and the total dynamical time. Quick encounters, between small low mass galaxies will have very little impact on the interactors. Slow encounters between two massive galaxies will be devastating for them.

1. Could you describe how, observationally, we define a wet, dry and mixed merger? (page 13)
   1. We would have to use the colour of the galaxies as a proxy for this. For instance, we need to know the gas content of the galaxies involved for a true measure. However, this is very costly to calculate. Therefore, we would use the measured broad-band photometry to do this. If both the galaxies involved were blue, we would say that this is a wet merger, both red would be dry and one or more of each would be a mixed merger.
2. Could you explain why wet mergers have an increased AGN fraction? (page 13)
   1. This is because of the increased amount of gas within the galaxy. Due to the torques of the interaction, the gas within the galaxy loses angular momentum and begins to move towards the center of the galaxy. This over density of gas could lead to accretion beginning in the galactic core, and therefore, black hole growth and AGN.
3. Why would micro mergers be a driver of cosmic star formation compared to major ones? (page 14)
   1. This would simply be a question of frequency, to be honest. Minor and micro interaction fractions and occurrences are more frequent than major ones. These interactions also leave one of the two galaxies relatively unchanged by the interaction. So, the secondary is destroyed and any gas that it contains will be accreted into the larger galaxy. Thus, this could act to replenish the gas of a large primary galaxy and also lead to the conditions necessary for cloud fragmentation and star formation.
4. Can you describe the work of Toomre and Toomre further? Why was their work so ground-breaking? (page 15)
   1. Since writing this, I’ve actually found a paper from 1948 which talks about the potential interactions of galaxies with N-body simulations. Anyway, at around this time, it was thought that the strange tidal features in ‘peculiar galaxies’ could be the result of interaction, but the full extent and likelihood of this seemed rather remote. Toomre and Toomre were the first major work to use a basic numerical simulation to accurately recreate the tidal features of different passages of galaxies of various masses. Here, they proved that a passage of another galaxy could form fine, filamentary tidal features simply from gravitational effects.
5. Can you tell me what violent relaxation is? (page 15)
   1. Violent relaxation is transfer of the rapidly changing gravitational potential into the orbits of the stars.
6. Also, can you expand on the definition of dynamical friction? (page 15)
   1. This is essentially a drag force that is exerted upon the two galaxies in the system. Imagine they are moving past each other. Once they are passed, their gravitational forces cause tidal features to form in bhind them. This causes a dragging force on the two systems which cause them to lose angular momentum and velocity. If this is enough to cause them to be trapped in the orbit, they will continue to spin towards each other.
7. What do you mean by changing gravitational fields? (page 16)
   1. Here, I mean the disturbance in the gravitational potentials as it becomes more complex than a single isolated system. This results in the torques on the interacting systems, and the shearing of the outer regions.
8. Do we see increases in the star formation rate in tidal features of interacting galaxies? (page 19)
   1. Interestingly, yes. The torques and forces I refer to here act much more strongly with the gas than with the stars. So, while some stellar material will be sheared from the disk in a tidal tail, the primary component is gas. So, I would think, with the features being primarily gas being significantly compressed, we would see large star formation rates in these tidal features. In fact, I have been accepted for a WEAVE proposal where we will be looking at this specific problem, and using a statistical argument to find if the SFR in the tidal features is more or less than the disk and then in isolated galaxies.
9. What is the long, complicated process of accretion by the super massive black hole? (page 19)
   1. Here, really, I want to talk about emission, build of the accretion disk and then the heating of the surrounding gas.

So, the internal structure of an AGN is such that we have a SMBH surrounded by a very hot accretion disk which is embedded in a very dusty torus. There is various densities of dust perpendicular to the torus. As the AGN accretes material, the accretion disk heats up to the point it begins emitting radiation that then is Compton scattered off of the surrounding clouds as X-rays. The other emissions (and even the X-ray emission) then depends on our orientation angle of the AGN.

1. Could you break down all the different classifications for AGN for us, and explain the difference between them? (page 20)
   1. Sure, we have many different classifications of AGN based on how we observe them and their orientation. So, first, we have Seyfert Galaxies and Quasars which we classify based on luminosity. Seyfert galaxies are then subdivided into Type 1 and Type 2 Seyfert. A Type 1 has both broad and narrow lines present in its spectra, while Type 2 has only narrow line emissions. We also have two other classifications due to emission lines: LINERs and Blazars. A LINER (Low-Ionization Nuclear Emission-Line Regions) has very weak narrow line emissions. Finally, there are then radio loud and radio quiet AGNs, and simply depend on if they have radio emission.
2. What is AGN flickering? (page 20)
   1. AGN flickering is an idea that comes about from a disagreement in the time of accretion required for black hole masses observed with the calculated lifetime of an AGN. For a blackhole to grow to the masses we measure, they would need to accrete for close to 10^{9} years. However, measured AGN accretion rates is significantly less than this, about 10^{5} – 10^{6} years. Thus, it has been proposed that AGN flicker on and off, holding both ideas.
3. Why would a delay in the AGN activation make sense? What would cause the delay? (page 20)
   1. In interaction, I think it makes sense. You’ve thrown tonnes of gas into the galactic core, but perhaps it takes time for the gas to be funneled into the galactic black hole. It may take time foe the accretion setup to form. I think that would be reasonable.
4. What other mechanisms could be in the AGN itself causing suppression? (page 20)
   1. I’ll admit, the works I’ve read on this don’t really tend to speculate. I would argue that perhaps it uses all of the material about it very quickly (in galactic terms) and then more material must infall for the system to begin again. This is often called chaotic accretion, and is based on very precise requirements for the angular momentum of the material for it to actually accrete.
5. What are AGN winds and how do they blow out gas in the galaxy? (page 21)
   1. When I say AGN winds, I mean there are a few. We have radiative processes that drive out high energy photons. These photons then blow out material and exert radiative pressure upon them. These radiative processes are driven by the accretion in the galactic centre.
6. Can you describe the structure of a neural network? How does it actually work? (page 23)
   1. This depends somewhat on the neural network we’re discussing, but here I’ll try to remain broad. A neural network is a set of connected nodes which are various layers deep. Each node that is interconnected takes an input – a number, an array, whatever – and then usually activates or de-activates. Whether it does either is dependent on the value of the input that has entered it. This number is weighted by some value between nodes, with input coming from every node in the previous layer. These weights can be tuned in what is called training. In supervised learning, we define a training set with which we know what the neural network should answer. So, we input our training data into the neural network, it splits the input into many different layers with the weights acting on each and activating and de-activating depending on the function. It then outputs some final answer based on a classification step. We then compare that to our known answer, and check if the neural network got it right. If not, we tweak the weights in the neural network and go again.

How we tweak the weights is dependent on the learning rate and optimizer of the neural network. The optimizer is method you are using to measure how you are doing, and then tweak the weights. So, for instance, a gradient descent algorithm, which checks that the loss is constantly reducing. The loss function is some way that you measure how the much the neural network is getting wrong.

1. How are such classifications of a set actually made? (page 23)
   1. So, in the context of Zoobot, all the output values (1,280 of them) go into a SoftMax output. This output takes each value, and converts it into a single number. It does this by processing the data (I think summing it) and then taking the exponential. This exponential value is then divided by the sum of all individual weighted values of the exponential going into it.
2. Can you describe some different optimizers and loss functions? (page 23)
   1. Sure, I actually have a list of them in my thesis.
3. What’s different about a Convolutional Neural Network and a regular neural network? (page 24)
   1. A neural network mainly acts on images. It has a set of layers which would necessarily be applied to a regular NN. These are the convolutional and pooling layers. The pooling layers reduce the size of different inputs which are sections of the image. These either bring out features or gradients in the image. There are then convolutional layers which make a convolution with a kernel to, again, smooth the smaller layers and encode the features of an image. Once this has been done, the pooled and convolved images are flattened and sent into a neural network as described previously.
4. Where does the inconsistency with training a neural network on simulations and applying to observations come from? (page 24)
   1. Simulated data, simply, cannot be fully represented of observed data.
5. Was it a good idea to name everything Chapter? (page 25)
   1. That was at the behest of my supervisor. I can’t possible comment.

Chapter 1: Zoobot

1. What are the debates about the fueling of AGN? (page 28)
2. What more could be done to remove contamination by close pairs? (page 30)
3. Could you describe how ESA Datalabs works? (page 31)
4. What other ways will ESA Datalabs impact the field? (page 31)
5. What future work would you do with ESA Datalabs? (page 31)
6. Why did you choose to use ACS, *F814W*, etc for your dataset? (page 31)
7. What time is ESA Datalabs actually saving you? Why can’t you just use TAP services to download the cutouts? (page 31)
8. Could you further discuss the applicability of the Shapely Python package? (page 32)
9. What do you mean by affects of interpolation? (page 32)
10. Describe more the functionality of Zoobot. How does it specifically work? (page 33)
11. Explain representation learning, as I’m unsure what you specifically mean here? (page 34)
12. In the original W+22 work, they used flattened 3-colour images. How does using only a single-band image affect your classifications? (page 34)
13. If you are not using the prediction score as a probability score here, then what does it mean when you have a cutoff of 0.95? (page 34)
14. Why do we need a smaller training set size when we are conducting finetuning? (page 36)
15. Why are the image contrasts changing with source size? What affect might this have had on training? (page 38)
16. In general, what is active learning? (page 40)
17. Is conducting this image augmentation and adding to your training set valid? What impacts does this have on your accuracy? (page 41)
18. Why is your validation set not completely balanced as well? (page 41)
19. What effect does this bi-modality tell you about your results? (page 41)
20. What effects on your results would be reducing you 0.95 cutoff? 0.95 does seem very stringent? (page 43)
21. If your value does not correspond to probability, what does it actually mean when you have a cutoff of 0.95? (page 43)
22. What do you mean by ‘despite removing 50% of the catalogue’? Do you think you would be finding 63 million interacting galaxies? (page 43)
23. As you say, using a balanced dataset leads you to be biased towards classifying a galaxy as interacting. Why didn’t you use an un-balanced dataset? (page 46)
24. Can you extend what you say about hierarchical clustering? How does it actually work? (page 47)
25. Describe Euclidean Linkage. Why didn’t you use a different linkage system? (page 47)
26. You talk about representation learning. What is this? (page 48)
27. Explain Principal Component Analysis. (page 48)
28. How do the representations relate to the morphology of the galaxy? (page 48)
29. Explain an AutoEncoder. What is it doing and how does it take the 40-dimensional representation and reduce it to a 2-dimensional projection? (page 49)
30. What would happen to your results if you changed your X and Y mappings? Why did you select these specific values? (page 49)
31. Why would each of these sources be in the HSC, but not in any other archives? (page 53)
32. Does a source not existing in Simbad, ViZier or NED really mean that it’s unknown? What other databases exist where these could be within them? (page 53)
33. What is bootstrapping? How did you actually conduct this here? (page 55)
34. Why was there final contamination at the end? Why not remove it using further visual classification? (page 55)
35. You’ve mentioned them, so I’m going to ask. What are: submillimeter galaxies, quasars, jellyfish galaxies, galactic jets, gravitational lenses, Lyman-alpha emitters, transitional stellar objects and supernova remnants? (page 57)
36. 5” for a match seems rather large… Why did you use such a wide criteria to match to Simbad or ViZier? (page 57)
37. Have you done the further work in order to confirm the classifications of these extra gem systems requiring multi-wavelength data? (page 57)
38. You are looking for hard and soft X-Ray emission for AGN. What’s the difference? What do they tell us about the system being observed? (page 57)
39. What do you mean by heterogeneous selection and analysis procedures? (page 58)
40. Why would you expect a second locus in this parameter space? How can you be so sure you’re not just picking up high starforming objects due to the *F814W* filters dependence on the UV emission? (page 62)
41. Explain Figure 2.13. What are you talking about? What does this actually show that we would expect about this sample? (page 64)
42. What value would you use to split the red and blue sample here? For each panel? (page 64)
43. In your conclusions, you don’t really discuss that these are morphologically identified interacting galaxies. Why? (page 65)

Chapter 2: Mergers in COSMOS

1. In your Introduction, you define the dynamical timescale. Is this actually correct? (page 67)
2. Can you define the project separation in this context? (page 67)
3. Why do you not use the 3D separation between your systems, and map this out? (page 67)
4. Which COSMOS2020 catalogue did you use? Why this one? (page 69)
5. What is broadband photometry? (page 70)
6. How is it used to measure these ancillary parameters? (page 70)
7. If at redshift 1.2, you are saying that tidal features are difficult to identify, how could Zoobot in the previous chapter? (page 70)
8. You choose limits of stellar mass down to 6.5 and star formation rate down to -5. Is this not incredibly low? Can a galaxy actually have a stellar mass down to 10^6.5? (page 70)
9. Why do degeneracies appear in the plots when you use the same software? What does this tell you about your results? (page 71)
10. What are the limitations of the visual approach taken in the COSMOS project? (page 73)
11. What could you do differently to find more secondaries? (page 73)
12. The large % of your catalogue not having an identified secondary is incredible surprising… Expand on why you think this is acceptable. (page 73)
13. You use a very large separation to identify your galaxies. Why did you do this? (page 73)
14. Are the fractions of secondaries with stage expected? Break down each one. (page 74)
15. Could you explain what you mean by de-duplication? (page 75)
16. What did you do to check that your control galaxy wasn’t interacting? (page 75)
17. When you are finding additional interacting systems, why were they not picked up by your catalogue? (page 75)
18. What do you mean the mass-limited sample gives you uniform sensitivity across the volume? (page 77)
19. How does the environment affect the SFR within a galaxy? Why could it match an interacting galaxy? (page 79)
20. How would you go about breaking this degeneracy in the dynamical timescale? (page 80)
21. Are you biased in the inclination of your systems? (page 81)
22. Your COSMOS cutouts are 30” by 30”. What distance would the secondary have moved to not be within the cutout anymore? (page 82)
23. Why would post-mergers be classified in your merger stage if a key criteria is that a secondary must be present? (page 82)
24. Why does a galaxy in a cluster environment have, on average, a higher SFR than a field galaxy? (page 86)
25. Could you describe the different methods of measuring the galactic environment? Why did you stick with Darvish et al’s method? (page 86)
26. What are you talking about with bootstrapping and approximating the selection effects in your paired sample? (page 90 - 91)
27. Could you explain the KS and AD tests? What’s the difference? What does each test actually measure? (page 94)
28. Could you describe clumpy galaxies to us? As you have written about it. (page 97)
29. Could you expand on the relations you expect to uncover from the different mass ratios? How would this affect your results if you were dominated by major interactions rather than micro or minor ones? (page 99)
30. In your description of your samples, you have jumped between the mass and volume- limited samples. You need to summarise this better in Chapter 3. (page 99)
31. Does the weighting scheme you have used in stellar mass and SFR keep your results valid? (page 101)
32. How does balancing based on the stellar mass lead you to have equal number counts in each bin? (page 101)
33. Could you summarise the error analysis you used from Cameron (2011)? What are you talking about with the beta function? (page 105)
34. Fully break down and explain equation 3.1. You say it’s from Aird et al (2019), but you do not explain the fundamentals of this. (page 106)
35. Are these results what are expected for the different stages of interaction? If so, why? (page 110)
36. Why have you not used the confirmed merging galaxies in your pair sample? Could these not be assumed to be all at 0kpc separation? (page 110)
37. You change your bin widths through the projected separation space. How does this affect your results? Is this something you can legitimately do? (page 110)
38. How does taking the average measure the excess of the SFR from interaction? (page 110)
39. What about only the pericenter and apocenter measurements on Figure 3.21? (page 111)
40. How did you apply the Cameron (2011) methodology with respect to the projected separation? (page 112)
41. Is the projected separation distribution with SFE for the full sample as expected? Why? (page 112)
42. Should you not say something here about the timelines of the AGN and interaction? (page 113)
43. Your weighting scheme in the AGN fractions is not clear… Did I actually do this two step fraction? (page 113)
44. Do we expect that the AGN fraction in the approaching pair phase is actually supposed to be this high? (page 113)
45. How do the low counts in this sample affect your results? Would you expect these results to remain consistent with a higher number counts? (page 113)
46. What other tests could be conducted which lead to a more optimal examination of these low count AGN results? (page 113)
47. Explain the physical mechanisms behind a delayed AGN ignition. What’s going on there? (page 116)
48. Could you define what you mean by a delay in AGN ignition? Surely this is just because you have no mergers in this sample? (page 118)
49. Could you expand on the change in the SFR other works have observed? Do these often use observations or simulations? (page 119)
50. You argue from your results that you have found evidence of no starburst, but actually of a slow increase in star formation rate with time. How are you able to do this? Why do your results reflect this conclusion? (page 119)
51. I’m going to need you to really expand on the small increase in the star formation rates over a longer time on page 120. How do you get this from your results? (page 120)
52. Is this idea of long term decline of SFE in escaped galaxies and quick decline of SFE in merging galaxies supported in the literature? (page 120)
53. How do you know that post-merger galaxies are quenched? You haven’t got many in your sample? (page 120)
54. You said you might have contamination from post-merger galaxies in your final sample. Doesn’t this mean that you’re finding high star formation rates in these as well? (page 120)
55. Please explain the difference between delayed AGN ignition and AGN flickering? (page 121)
56. What is a delayed AGN? (page 121)
57. What is AGN flickering? (page 121)
58. Using the projected separation of the interaction, you make a temporal argument for AGN flickering or delay. Could you explain how you made this leap? (page 122)
59. Could you explore, for a moment, what would occur if the two galaxies were overlapping and merging? Would your measure of star formation and mass change? (page 123)
60. How significant is finding that the projected separation is not a perfect proxy for stage? Should we use something else besides interaction stage? (page 124)

Chapter 3: Inferring Galactic Parameters

1. Can you describe in a bit more detail cosmological simulations? How do they work? (page 127)
   1. Well, there are many kinds of cosmological simulations. A set of initial conditions are set for the simulation. This can be the gas distribution, the distribution of the baryonic particles within it or even just the size of the box. They then run this with the particles having mutual gravity between each other. With the correct initial conditions and physics involved, we then find the cosmological structure that is comparable to the observed universe at z = 0. There are then many ways to describe the gas within these simulations. Often, there are a set of gas cells that overlay the simulation. Thus, things like star formation can be calculated. There a few approaches to dealing with gravity in these cosmological simulations. Primarily Tree methods and Mesh methods. The Tree method groups particles together for the calculation of forces based on their density. A mesh method calculates the forces on an increasingly refined mesh. These both are for computational purposes.
2. Can you expand on and describe the Galaxy Zoo: Mergers process? (page 127)
   1. I have done many times.
3. You randomly bring up the Vera C. Rubin observatory here. Why? Why does the age of the Vera C. Rubin observatory matter? (page 127)
   1. This is true, this is actually a result of the writing of this thesis and this project. Initially, we started by wanting to develop this code for the LSST collaboration. This would be able to quickly find posteriors to the interacting galaxy underlying parameters. VCR is a large survey telescope that is nearing completion in Chile. As I stated previously, it is estimated to bring in more galaxies than have been observed to date.
4. You say samples of thousands of interacting galaxies are going to be found a week with LSST. Is this really true? (page 127)
   1. As I stated previously, it is estimated to bring in more galaxies than have been observed to date. By that fact alone, I think it will be the most interacting galaxies seen to date.
5. What is a probability distribution? (page 128)
   1. A probability distribution is measure of the chances a parameter has a certain value. So, for instance, in our mass measurements we find the probability distribution of the parameter and which value it is most likely to have.
6. You discard 8 of these systems because they are not in SDSS. Can you not just use the best fit underlying parameters that GZM found and incorporate these into your sample? (page 129)
   1. I thought this might come up… Don’t worry, I have already created the other 8 simulation images and am running them as we speak.
7. Also, could you not simply assume the redshift of these systems and gain another 3 systems in your sample? (page 129)
   1. I don’t think that would gain us anything, really. In early iterations of the code I did try this and just assumed a redshift of 0.01. However, it just messes the scaling of the image up and leaves to really poor results. As I state the user must have the redshift, I didn’t want to add in the really bad results with the ones which are already variable based on a few other issues.
8. What effects on the morphology of the system would combining the white images be? (page 129)
   1. I think it’s more a question of what we could miss if we didn’t have this combined in image. As we know, galaxy morphology can change based on the filter being used to observe. Simply, some filters are worse at picking up things than others.
9. Why did you conduct the block reduce to 100 x 100 pixels? (page 129)
   1. This seemed like the right trade off between resolution and pixel number. If we reduced it further, say to 50x50, then often the specific shape of the tidal features could be lost and become blocky. At 150 x 150, we often didn’t have the particle resolution to successfully recreate the distribution fully.
10. Could you provide a summary of the JSPAM code for us? Why is it very efficient and fast? (page 131)
    1. The Java Stellar Particle Animation Module is an N-body code which makes use of a unit galaxy in order to achieve high efficiency. A unit galaxy is a one which is broken down into the components that I describe in the thesis, and then has a mass of 1 simulation unit – about 1x10^11 solar masses. This is then scaled by the user input. This is concentrated in a point at the centre of what will be the particle distribution. So, rather than attempting to calculate the orbit of a host of particles, we predict the path of this unit galaxy from the final position to the initial position. We then add the particles and integrate them forward using a 4th order Runge-Kutte. This two step process gives us a huge boost in efficiency.
11. Define a reasonable time resolution. (page 132)
    1. A single time unit in the simulation is 57.7Myrs. So, we use a timestep of 0.01 units of 0.577Mys. This is incredibly high resolution in time, I think.
12. Could you explain a fourth order Runge-Kutta methodology? (page 132)
    1. The Runge-Kutta is an iterative method by which we calculate the change in particle position and velocity of a series of discrete timesteps. It is complicated by both the position and velocity being functions of each other. So, we calculate the next position based on the velocity in the current timestep. We then increase the time by one timestep value, and recalculate the new velocity on the particle in the new position. Thus, it is called a fourth order approximation as we update the position based on the outputs of four separate functions approximating the slopes at four different points. The equations are in my thesis.
13. What is a softened point mass approximation? (page 132)
    1. Both of these questions are ways in which we calculate the new gravitational potentials which are then presenting the forces upon the particles in the simulation. The SPM approach has a single simple equation that describes the gravitational potential anywhere in the simulation. This is then translated into the 3D acceleration of the particles rather easily. The N-body approximation directly accounts for the gravitational potential with a bulge, disk and dark matter component. This increases the complexity of the gravitational potential, and therefore the acceleration the particles undergo.
14. You say you elect a softened-point mass over the N-body approximation for computational efficiency. How different are the different approximations? Are then any other physical reasons that you might choose one over the other? (page 132)
    1. The to approaches do have somewhat different morphologies, but hardly different. When applying this to the simulations that were created with the same potential, this was fine. However, with observations, it is a slightly different story…
15. You mention genetic algorithms. What are these? (page 132)
    1. Essentially, this is a hierarchical model with a lot of biological terminology in its description. We begin by estimating one or two parameters and generating their images. The best fits of these are then passed to the next ‘generation’ which attempt to fit further parameters. This continues until all parameters have been fit and allows us to compare images one by one. It is, essentially, a slower hierarchical MCMC which I did look into to little avail.
16. How would a genetic algorithm be used in the context of finding the parameters of interacting galaxies? (page 132)
    1. Described as above.
17. What are semi-analytic models? (page 133)
    1. SAMs are similar to N-bdoy and hydro codes in that they are designed to study galaxy formation and evolution. However, where hydro codes directly and numerically calculate the result of lots of physical processes, SAMs use analytic expressions to approximate these. For instance, in a hydro simulation, gas would be directly measured. This gas would then undergo star formation when the gas particles themselves are at a certain density etc etc. In a SAM, a gas reservoir would exist where would assume that some equation dictated the SFR in any given time frame.
18. You state you use a Bruzual & Charlot simple stellar population. What is this? What does it look like? (page 133)
    1. The BC03 SSP is a simple stellar population which was created by BC in their paper in 2003. This was a new SSP which advanced previous models based on observations of existing stellar spectra. This one also accounted for a few more things, at the time, such as thermally-pulsating stars. They also assumed a Chabrier IMF. Each library is a large dataframe of expected emissions at different wavelengths at the expected age of the SSP.
19. You then use a Chabrier IMF. How do you use it specifically? What is the initial mass function in general? (page 133)
    1. Well, I don’t really use it. It is built into the model. However, the IMF is an assumption about the number of stars that will be born in an SSP at the same time in a given mass bin. Think of having a set of 10,000 stars. This population would then be weighted between the different mass intervals. It is a two power law, with an exponential dependence on mass for stars less than 1 solar mass and then a power function above that. This knee in the IMF differs from earlier works, such as the Salpeter IMF which is just a power law, if I recall correctly.
20. What is the e-folding timescale? (page 133)
    1. In general physics, we describe the e-folding timescale as the time in which it takes a quantity to increase by a factor of e. Therefore, in this context of an exponent starburst, it is the time in which the starburst occurs over.
21. What is a spectral energy distribution? (page 133)
    1. It is essentially a spectra.
22. You say this matches a bunch of works. Why does it have this value of about 1.7 Gyrs? (page 133)
    1. Admittedly, this may be a little bit low. Works such as Horizon AGN and Phillips find the e-folding timescale to be between 2 – 4 Gyrs. However, these were with specific kinds of galaxies. Peng et al. (2010) was working with field galaxies and found the value to be between 1.5 and 2.5 Gyrs, if I recall correctly. We chose 1.7 Gyrs as when we used this in our star formation histories, it recreated a very small, intense starburst.
23. You define a gas fraction in these galaxies as 0.15. Is this reasonable? Where did this number come from? (page 134)
    1. Previous works have actually found this value to be about 0.12, and is dependent on the measured absolute magnitude of the galaxies involved. We elect to have a higher gas mass in our models to fully explore the results of a sudden increase in the SFR.
24. Is the initial galaxy age of 10Gyrs reasonable? (page 134)
    1. Yes, I would say so. Perhaps with recent works from JWST, showing disks are much more prominent in the early Universe this number should be updated to be a little bit older. However, in early iterations of this, we found that changing the age significantly did not alter our results.
25. You say you model new stellar populations as the interaction continues. What do you mean? How do you do that? (page 134)
26. How is your semi-analytic SFR enhancement similar to that done in CIGALE? (page 134)
27. What is CIGALE? (page 134)
28. Does your enhancement parameter have any basis in the literature? It seems rather arbitrary. (page 135)
29. You define an SFR\_{galaxy} in equation 4.3, but do not define this. (page 136)
    1. This is the total reservoir of SF in the galaxy. This was a mistake on my part.
30. Why have you not defined a gas distribution model? (page 136)
    1. The gas distribution here actually follows the particle distribution due to the SFR being a density dependent thing upon binning.
31. Can you show that your assumptions do mimic the simulations direct calculations? (page 136)
    1. This was primarily based off of the simulations of Jorge Moreno, 2021. Where the model their star bursts as large, quick exponentially declining bursts.
32. What are the limitations of removing particles with no neighbours to zero? (page 138)
33. Your mock observations appear to still have holes in their tidal features. So, did you flux distribution algorithm not work? (page 142)
    1. It did work for the tighter tidal features. The really large ones, such as this one in the image, there is simply no way that we could have the flux distributer make the entire thing with no holes in.
34. Why did you opt to use MCMC here? And not something like, say, simulation based inference? (page 143)
    1. Honestly, it was more about timing than anything else. Simulation Based Inference has been around for a while, but the first major works of its application in astronomy for what I would use it for were with Jeffery et al. (2019). These then have come to fruition in the last few years. If I could start this project again, I would go into SBI and focus on the development from there. This project was supposed to a 1 – 2 year project which would show the proof of concept and development of such an automated technique. The actual development time of the project has been much longer than that.
35. Please explain how a typical MCMC would work. (page 145)
    1. Ok, the basis of an MCMC is that of the Monte Carlo simulation. This is a simulation which takes in a function which predicts, from a set of input parameters, an outcome that we can map. For instance, predicting how prices change in the stock market. The Monte Carlo simulation takes in a set of randomized parameters under some prior, and then maps out the outcomes and we can check how likely each outcome is from the number of times that outcome is found.

An MCMC adds in a Markov Chain to this simulation. So, instead of randomly finding the outcome of a situation, we map the probability that an outcome will happen and then want to find where that probability is highest. MCMC includes walkers, which are sets of the parameters you are investigating with different values. They ‘walk’ around the parameter space, predicting outcomes for each set of parameters and then calculating the probability that such an outcome is correct. The walkers move as a function of a step size and the distribution of the other walkers (in general, we use some more advanced ones here). So, by the end of the MCMC, the walkers should have all clustered around the area of highest probability (as that is what they are looking for) and we can measure the posterior distribution of each parameter.

Lot of drawing going to take place in this one.

1. What is a posterior distribution? (page 145)
   1. A posterior distribution is a measure of a likelihood function through a parameters space. I.e. it is a curve where each point is a measure of the probability of each value.
2. What is a confidence interval? (page 145)
   1. A confidence interval is the measure over the full probability space, this curve. As I said at the start, I did exactly what I didn’t want to do and used the wrong terminology for how I measure my results. I use credible intervals, which are measures of the mass of the walkers which are in certain regions of the space.
3. Please explain how emcee works in the context of your work. (page 145)
   1. As above, really.
4. Please explain what a Differential and Snooker-Differential Evolution move is? How do these differ from a typical walker move? (page 145)
   1. Typically, and most basically, a walker move is calculated as a function of the distribution of the nearest few walkers. However, for a Differential Evolution move, the movement and vector of all chains in the ensemble is taken into account when choosing the next step. A random two chains in the ensemble are selected, and their previous step vectors are subtracted and then multiplied by factor dependent on the dimensions of the parameter space. A further random vector step is then applied to move within the small space.

So, dependent if the walkers have found a high probability, they will move in small areas at the peak. It multiple peaks have been found, they can jump between these and the walkers will cluster over the most probable peak.

A Snooker-Differential Move then uses the above, but accounts for another, completely different chain, and the density along the proposed movement of the line.

1. What motivated the choice of moves for the walkers? (page 145)
   1. As I stated above, both the DE and the DES moves are very good with highly degenerate parameter spaces due to their ability to move between different peaks. However, the differences between the peaks turned out to be so small, it was very difficult to achieve this in the computational expense provided.
2. Why are these moves good for high dimensional parameter space? (page 145)
   1. As above.
3. Define a prior. (page 145)
   1. A prior is just a way of incorporating known information into the MCMC algorithm and Bayesian statistics.
4. What motivated the decision to have uniform priors? (page 145)
   1. We wanted to uncover the degeneracies and such.
5. By adding to the gamma parameter as you describe, it sounds like you do not have uniform priors? (page 146)
   1. Potentially, this is true. I’d more describe this as a weighting to our prior.
6. Explain further this constant C, in equation 4.6. (page 147)
   1. This is a use case of the filter parameter as described in Holincheck et al. 2016 in the GZM paper.
7. You state that you find the underlying probability distribution is not Gaussian. Therefore, how does this affect your results? Could making another decision here lead to a better outcome? (page 147)
   1. Potentially, however, I won’t try to argue what should be done here. The underlying parameter space seems so complex dependent on the parameter that I don’t think I could even comment on what could be done here.
8. You bring up GALFIT here, what is that? What’s it for? (page 147)
   1. GALFIT is used to measure different morphological parameters of galaxies. It models flux distributions of different light profiles, sersic indexes, inclinations, etc. It then compares the mock image to the observed image and finds the minimum of a chi squared.
9. Why does taking the log help with computation time? (page 148)
   1. In theory, it removes the flatness of the probability distribution. Any value, when not logged, would be between 0 and 1. Therefore, by taking the log, low probability spaces are many thousands and cause the walkers to move faster to the best fit values.
10. How should I publish all the probability distributions? Currently, on a GoogleDrive. (page 148)
    1. For the examiners to decide.
11. Why does the best fit and observation images look so different in Figure 4.2? (page 149)
    1. Pretty much, this shows the whole issue with the approach. The found best fit simulation is just that – a best fit simulation. There are still significant differences between them. Hence, I describe these best fit images as different from the observations.
12. What is the Corner Python package? How does it work? (page 150)
13. Why do we find degeneracy in the orientations? Why is it a 4-fold degeneracy? (page 152)
    1. Ok, up at the board.
14. Why did you use the most massive system in your sample as your example? (page 152)
    1. Highest flux distribution, and should be the best and most easy system to constrain.
15. Why is the mass the easiest to constrain? (page 152)
    1. Because it’s directly related to the luminosity and flux distribution, which is what is directly in our Chi-Squared measurement.
16. Are there other approaches you could have taken besides a chi-squared approach? Why weren’t these considered? (page 152)
17. Would this movement of the secondary in the image actually change the errors you have? (page 153)
18. Why does the secondary disk contribute more to the tidal debris? (page 153)
19. Why does the simulation being in the frame of the primary mean that the 3D velocity of the secondary is more? (page 153)
20. What are you talking about with the backward integration? It’s not very clear. (page 155)
21. Why does the simulation give non-physical results when we start at or very close to pericenter? (page 155)
22. Is your time parameter actually within this confidence interval? This does not seem constrained at all. (page 155)
23. Why does the skewing of the time parameter to small values seem excellent compared to the skewing of the position and velocity parameters? (page 155)
24. Is your gamma parameter really necessary? If it is preventing you finding the best fit parameters of these galaxies, then surely it is a hindrance? (page 156)
25. What do you mean by changing each parameter by a small amount based on the posterior? (page 156)
26. You say that your results appear to have converged. Is this what you really mean? (page 157)
27. You’ve described the Geweke diagnostic. How does it actually work in practice in ChainConsumer? (page 157)
    1. Essentially, what we’re doing is comparing the variance and the mean of the movement of the walkers in two different parts of the chains. At the start, we have the walkers ‘burning-in’. This means they are essentially moving randomly around the parameter space. So, the logic goes, if we conduct this means test on the start and at the end and they match, it means that the walkers are still moving essentially randomly at the end of the chain that you have run. This indicates that the walkers have not actually converged on a decent estimate of the posterior distribution.
28. Why does the Geweke diagnostic work? What does comparing the first 10% and the last50% of each chain tell us? (page 157)
29. Why does the degeneracies in the orientation space lead to a disruption of the Geweke diagnostic? (page 157)
    1. Pretty much, as above. The means T test almost entirely spans the parameter space. So, while it has converged in four different areas, when we calculate it.
30. What does it mean to flatten and thin the walker chain? What does this actually do to it? (page 159)
    1. Flattening the walker chains simply mean taking each chain for each parameter and combining them into one long measure of each walkers position through the entire run. Doing this allows us to measure the mass of where walkers were at different stages of the run, and find the peak of this. This reveals our posterior.

Thinning is simply an effect of the highly auto-correlated nature of the walker positions and mass. There is no reason to store this, as we can remove them and retain statistical power as long as we are above the auto-correlation time. In most cases we are, and therefore we remove a multiple of the auto-correlation time of the walkers.

1. What effect would throwing away more steps have on your results? What about less? (page 159)
   1. You can keep all of the walkers and positions if you want to and it didn’t change the results very much. I simply opted to do the thinning as it reduced the computation time in ChainConsumer with little effect on the found marginalized distributions.
2. What does it mean to ‘burn-in’ the walker chain? (page 159)
   1. Burn-in is the first random walks of the entire ensemble across the parameter space. These are almost entirely random as the walkers are initially distributed throughout parameter space. This is, therefore, essentially noise in the mass of our particle positions. Removing these give us a cleaner posterior distribution.
3. Is assuming the circular velocity at different radii not enough to ascertain the rotational velocity? What do you think the rotational velocity is? (page 161)
   1. Sorry, it is. I was just unsure of how to combine this and create a map of this with the time left in the project. I think that what I have right now with the LOS velocities does demonstrate the power of adding velocity without requiring to go in and find this.
4. Do you expect to do anything with WEAVE in the future on this project? (page 162)
5. What are IFUs? How do they work? (page 162)
6. Why is your ability to constrain dependent on the stage of the interaction? (page 162)
7. Is it a bit of a cop out to discuss Arp 240 again in this section? (page 162)
8. Why do all of you mass probability distributions have a banana shape? (page 162)
9. What do you mean by stage 2, 3 and 4 here? (page 162)
10. You’re saying that two overlapping disks is the same as an ongoing merger or merger remnant? That doesn’t sound right… (page 162)
11. You seem to consistently find under-estimates for the time and mass parameters. Why do you think this might be? (page 165)
12. What further investigation have you made into exploring a more representative parameter space in the velocity? (page 167)
13. What do you mean by folding the parameter space? (page 167)
14. This is a very insightful discussion on the outputs of your simulation. So, how could we go about fixing this? Is it reasonable in a small timescale? How could you account for the over-estimates due to the ring-distribution of your particles in the galaxy? (page 167)
15. Why would the systems require a high beta parameter to increase the time of the closest approach? (page 168)
16. Where does this discrepancy between the scales of the H16 images and your observations come from? (page 169)
17. How are the positions of the secondary galaxies different between H16 and the observations? (page 169)
18. Why does having noise surrounding the galaxy lead to higher uncertainty in the radial probability distribution? (page 171)
19. How do you make a conversion of total to stellar mass in the Arp 240 best fit simulation? (page 171)
20. What are the errors within the He et al. (2020) paper for their measurements on the Arp 240 stellar mass? (page 171)
21. What about the other parameters of Arp 240? Can you make any further comparisons? (page 171)
22. Why do you say that some systems may simply not be able to be recreated with restricted N-body simulations? (page 174)
23. In your observational corner plots, there almost seems like a lot of binomial distributions. What’s going on there? (page 172 - 173)
24. What would happen if your MCMC tried to constrain an image at the wrong resolution? (page 174)
25. You say your MCMC needs a reasonable number of degrees of freedom to reach convergence. What do you mean? (page 175)
26. Explain different methods of improving computational efficiency. You talk about GPUs, but why would they actually achieve increased computational efficiency? (page 176)
    1. GPUs are composed of thousands of individual threads which lead to excellent levels of parallelization.
27. In your conclusions, you say you had to scale up the observed images. What do you mean by this? (page 178)
28. You say in your conclusions that this took thirty days, but in the main text forty days. Which is it? (page 178)
29. You talk about gaussian processes, machine learning and simulation based inference in the conclusions of this Chapter. What are these? Could you expand on how these could help you? (page 178)
    1. Sure, the main one I’ll focus on is Gaussian Processes, as I feel I’ve spoken bout the others enough earlier. GPs are a non-parametric, supervised way to model distributions via machine learning. My idea for this would be that we could break down our images into 1D flux distributions, and then attempt to write a Gaussian Processes that could link the parameter distributions to the 1D flux distribution. Then, rather than running the simulation many millions of times, we could enhance the exploration of parameter space with training a GP.

Conclusions

1. Do conflicting results emerge from using projected separation and morphological stage with projected separation? You found the same overall behavior with project separation anyway? (page 180)
2. Is it confirmation of AGN flickering? Please explain this further. (page 180)
3. What is amortization? (page 184)
4. Classic question, you explain the next steps of the development of the MCMC algorithm. Are you actually going to carry them out? (page 184)