Part I Observational overview

These notes will first present a very broad-brush review of the observations relating to galaxy formation that we would hope to explain with a complete theory, and then review theoretical approaches and methods aiming at explaining those observations.

1 Galaxy Structure and Morphology

One of the most obvious facts about galaxies is that they come in a variety of sizes and luminosities. Perhaps more surprisingly is that, for a given size and luminosity, there are usually a number of different types. Remarkably, the inherent reason for this variety is still not well understood. It is one of the key requirements for a successful theory: that we can reproduce the rich diversity of galaxies.

1.1 Galaxy Classification [Binney & Merrifield 4.3]

There are many ways of classifying galaxies but by far the most well-known is the Hubble system, which has originally suggested by Hubble in his 1936 book, *The Realm of the Nebulae*. Galaxy classification is a topic all of it’s own (for example see Sidney van den Bergh’s *Galaxy Morphology and Classification*), so here we just remind readers of the basic classes and the difficulties encountered in classification by any system.

In Hubble’s tuning-fork diagram, on the left are the sequence of elliptical galaxies spanning the range from circular to highly elliptical systems. These are denoted $E_n$, where $n = 10(1 - b/a)$ is computed by measuring the axial ratio $b/a$ of the ellipse formed on the sky. Note that these are measured from the projection on the sky, so that the intrinsic three-dimensional distribution may be considerably different (in fact, in general elliptical galaxies are likely to be triaxial, and the axes can change as a function of radius).

The elliptical systems are often further sub-divided into dwarf types (dE), although it is not clear if the dwarfs are simply small elliptical or a class of their own. Dwarf spheroidals (dSph) are a closely related type, the primary difference being that they are even less luminous and more diffuse (they are an interesting targets for dynamical studies because of their very high dark matter content). With the advent of new technology that permits detailed studies, the smallest galaxies have become a topical area of research. For a summary of known dwarf galaxies in the local group, see the article by Mateo et al. in the 1998 Annual Reviews of Astronomy and Astrophysics (Vol 36, p. 435).

After the ellipticals, Hubble’s diagram splits in two, with two parallel sequences differentiated by the presence (top sequence) or absence (bottom) of a stellar bar in the center. The first galaxies are the so-called *lenticular* or *spheroidal* galaxies, designed S0 (or SB0 if there is a bar). These systems appear to be strongly oblate like a spiral system but without any clear spiral structure. Like ellipticals, they usually have little dust or gas content, although some systems can contain significant amounts of dust: the strength of the dust is sometimes signified by a subscript ranging from S0$_1$ (no dust) to S0$_3$ (prominent dust lane).

After the spheroidals come the spirals, designated Sa, Sb or Sc depending on three characteristics: the tightness of the spiral arms, the size of the central bulge and the definitions of the arms. Spirals with tight, well-defined arms and a strong central bulge are Sa, while Sc systems have loose, poorly defined spiral arms.
and little or no central spheroidal component. Note that these three characteristics almost vary together, but exceptions exist which can make classification difficult. If the spiral has a bar, it is classified with SBA, SBb or SBc. Additions have been suggested to this system, including de Vaucouleurs addition of Sd and Sm spiral systems that include galaxies which are looser or have even more poorly defined arms than the Sc class.

Hubble thought that galaxies evolved along his sequence, going from elliptical ("early") galaxies to spiral ("late") galaxies. While this idea is no longer thought likely, the terminology has stuck. It is not uncommon to find a numerical T stage assigned to the Hubble type for the use in some forms of statistical analysis. This ranges from -5 (E0) to 0 (S0) to 6 (Sc) to 10 (Irr). Entirely outside of this sequence are the irregular galaxies Irr which lack symmetry or well-defined spiral arms, and the peculiar galaxies which may be undergoing a merger or interaction.

As we will see in more detail later, galaxies are not uniformly distributed in space. While isolated, field galaxies do exist, the majority of galaxies are found in groups or clusters. Very small groups, like the local group, consist of only a few large galactic systems. Large groups may have up to 100, while full-fledged clusters often have thousands of large galaxies. Galaxies of different types are not equally represented in different environments. In particular, ellipticals are much more likely to be found in dense cluster cores than in the field, while spirals show the opposite trend. This is most commonly called the morphology-density relation and is clearly telling us something about what sets galaxies morphology, but it is not clear whether spiral galaxies are transformed into ellipticals by the dense environments of clusters, or if such regions preferentially formed elliptical galaxies even before the cluster formed (clusters are thought, from both theoretical and observational evidence, to be younger than galaxies).

1.2 Galaxy Luminosity Function

We define \( \Phi(L)dL \) as the number density of galaxies with luminosities from \( L \) to \( L + dL \). The luminosity (energy per unit time) \( L \) could be in a particular wave-band or it could be integrated over all wavelengths (the bolometric luminosity). In either case, the luminosity function is usually normalized so that,

\[
\int_0^\infty \Phi(L)dL = n_{gal}
\]

where \( n_{gal} \) is the local density of all galaxies (we have not specified it but the luminosity function can and does vary with position). The luminosity function is a difficult thing to measure but it is clear that it drops quickly at large luminosity and has a large extension to low luminosities. A useful parameterization is the known commonly as the Schechter function:

\[
\Phi(L) = \left(\frac{\Phi_\ast}{L_\ast}\right)(L/L_\ast)^\alpha \exp(-L/L_\ast),
\]

which is simply a power-law of slope \( \alpha \) with an exponential cutoff. The cutoff occurs at \( L_\ast \) and as long as \( \alpha > -2 \) (which is observationally always the case), the total light integrated over all luminosities is finite:

\[
L_{total} = \int_0^\infty L\Phi(L)dL = \Phi_\ast L_\ast \Gamma(2 + \alpha)
\]

For the typical value of \( \alpha = -1 \), the Gamma function is unity and since \( \Phi_\ast \) is approximately the number density of galaxies with luminosity of \( L_\ast \), the total light is clearly contributed mostly by galaxies around \( L_\ast \). Our own Milky-Way galaxy has approximately this luminosity, and it defines the typical brightness of a large galaxy.

On the other hand, the total number of galaxies is not always convergent and indeed large numbers of low-luminosity galaxies are observed (although there does appear to also be a lower cutoff which is not modelled by the Schechter function). The luminosity function is difficult to determine observationally for a number of reasons. First, the low luminosity end is difficult to determine because faint galaxies can only be seen nearby. More worryingly, the surface brightness of many galaxies falls close to the limit which can easily be observed. Even intrinsically luminosity galaxies can appear dim if they are large and spread across a large region of the sky. Indeed, recent work has uncovered a whole population of low-surface brightness (LSB) galaxies. These LSB galaxies are typically only a few percent of the surface brightness of the background.
sky and so are difficult to see. However, they are clearly of great importance both because they may be as numerous as their high-surface brightness cousins, but also because many appear to be largely dark-matter dominated and so provide good cases to study the distribution of non-luminous matter in galaxies.

2 Elliptical galaxies

We first address elliptical galaxies in more detail. Elliptical galaxies are in some ways simpler because they generally do not have gas and dust, so are largely stellar systems. On the other hand, they do not have the simple orbits that characterize disk systems. We start with radial structure of elliptical galaxies, and then move on to global properties of the systems and their scaling relations.

2.1 Radial brightness profile [Binney & Merrifield 4.3]

While elliptical galaxies are generally not spherical, it is often useful to examine how their brightness varies as a function of distance from the center. This is done either by fitting an elliptical model or by fitting the major and minor axes separately. In either case, it is found that the surface brightness (i.e. energy per unit second per unit area per unit solid angle) measured in magnitude per square arcsec ($\mu$) varies with projected distance $R$ as $R^{1/4}$. If we write surface brightness in linear units (recall that $I \propto 10^{-0.4\mu}$), then it is usually to express this relationship as:

$$I(R) = I_e \exp(-7.67[(R/R_e)^{1/4} - 1]).$$

The numerical factors are chosen such that if 1/2 of the total light (for a spherical system) is emitted within $R_e$:

$$\int_0^{R_e} \pi R I(R) dR = L_e/2.$$

For this reason, $R_e$ is a useful measure of a galaxies size and is called the effective radius. It is also useful to define the mean surface brightness within the effective radius: $L_e = \pi R_e^2 < I >_e$.

While most elliptical galaxies follow this distribution, some giant elliptical galaxies show an excess of stars at large radii. These cD ellipticals are generally found at the center of large groups or clusters of galaxies. The excess is commonly interpreted as an overall envelope of stars that more properly belong to the clusters as a whole than to the cD galaxy. This intra-cluster light can contribute significantly (20% in some cases) of the observed starlight from the entire cluster.

2.2 The fundamental plane of elliptical galaxies [Binney & Merrifield 4.3]

Elliptical galaxies are unlikely to be significantly supported by bulk rotation, as is the case for spiral galaxies. However, the stellar velocities still give an indication of the gravitational mass. Usually, it is impossible to measure individual stellar velocities in elliptical galaxies, but the distribution of line-of-sight velocities can still be measured as a function of position across the galaxy. This is done by measuring the spread of stellar emission line caused by the Doppler shifting of all of the stars at that (projected) distance. The velocity dispersion at the center of an elliptical galaxy is denoted by $\sigma_0$ and is usually straightforward to measure. It was noticed quickly that this quantity was strongly correlated with the galaxies luminosity:

$$L_e \sim \sigma_0^4$$

This is known as the Faber-Jackson relation. Other relations between these quantities and $< I >_e$ or $R_e$ were found, but all of them (including the Faber-Jackson relation) can be thought of as projections of a single correlation between three quantities. This relation defines a two-dimensional plane within the three-dimensional space define by (say), $R_e$, $< I >_e$ and $\sigma_0$. One common parameterization of this is given by:

$$\log R_e = 0.36\mu_e + 1.4 \log \sigma_0,$$

where $\mu_e$ is the surface brightness in units of blue magnitues per arcsec (recall $I >_e \propto 10^{-0.4\mu}$ , $R_e$ in units of kpc and $\sigma_0$ in km/s.
2.3 Globular Clusters

Although globular clusters are generally not viewed as galaxies in their own right, they do overlap in luminosity with dSph galaxies. However, they are many times more centrally concentrated (their effective radii are typically a few pc) and so appear to be a different type of beast. The distribution of GC luminosities appears to be Gaussian, with a central value of a few times $10^5$ solar luminosities. Both spiral and elliptical galaxies contain globular clusters (the Milky Way has approximately 160) and we can compute the specific frequency of globular clusters as the ratio of their number to the luminosity of the galaxy: $S_N \propto N_{GC}/L_{gal}$.

More precisely the definition is given in terms of the galaxies V magnitude $M_V$:

$$S_N = N_110^{-0.4(M_V+15)}.$$  

(8)

There appears to be a trend with $S_N$ decreasing from left-to-right in the Hubble diagram. It is not obvious why bulges and ellipticals should have relatively more globular clusters for their given luminosity. In fact, the origin of globular clusters in general is not understood. The Milky Way is observed to have a bi-model distribution of GC’s, with roughly half been old and metal-poor, with the rest being younger and more metal rich.

2.4 Introduction to Collisionless dynamics [Binney & Tremaine 4.1, 4.2]

The stars$^1$ in galaxies are in an interesting dynamical state: gravity attempts to bring them to a central point but their own motions resist this compression. The situation is reminiscent of the hot gas in a star, but crucially different in one important respect. The cross-section for physical collisions between stars is very small: a star would have to orbit for a period much longer than the age of the universe before it would encounter a single collision with another star. This is very different from the situation inside a star, where collisions are vastly more frequent than almost any other timescale. The result is that in a star, the velocity distribution is Maxwellian and we can define a unique temperature and velocity at any position. We don’t need to integrate the trajectories of lots of individual atoms because we are only interested in the bulk properties such as the pressure. It is the pressure of lower layers which supports the upper ones against gravitational infall.

In galaxies, we can no longer assume that the velocity distribution at a given point is Maxwellian and so there is no unique temperature or pressure at a given location. Instead, the most general thing is to integrate the trajectories of all of the stars. The relevant equations are Newton’s equations of motion plus gravity. For a galaxy with a million or so stars in it, this approach might be possible if computationally taxing. For galaxies with $10^{11}$ stars it is impossible to follow every single star.

Another useful approach is to develop an equation for the distribution of stars as a function of position and velocity. We define the distribution function:

$$f(\mathbf{x}, \mathbf{v}, t)d^3\mathbf{x}d^3\mathbf{v},$$

(9)

as the number of stars which are within the volume element $d^3\mathbf{x}$ centered at position $\mathbf{x}$ and with velocities within $d^3\mathbf{v}$ of $\mathbf{v}$. The density of stars at a given position $\mathbf{x}$ can be found simply by integrating over all velocities:

$$\rho(\mathbf{x}, t) = \int f(\mathbf{x}, \mathbf{v}, t)d^3\mathbf{v}$$

(10)

If we make the assumption that stars change their position and velocities smoothly (i.e. that there are no sudden collisions – more on when this assumption is valid below), then the rate of change of the number of stars in some small region of phase space should just be equal to the net flux of stars into or out of that region. This is just saying that the distribution function should obey a conservation equation, which can be written as:

$$\frac{\partial f}{\partial t} + \sum_{i=1}^{6} \frac{\partial (f \dot{w}_i)}{\partial w_i} = 0.$$  

(11)

$^1$and cold dark matter if present
To stress the six-dimension nature of the distribution function, we have adopted a six-vector \( \mathbf{w} = (x, v) \) and used the usual dot notation to indicate a total derivative with respect to time: \( \dot{\mathbf{w}} = (\dot{x}, \dot{v}) = (v, -\nabla \phi) \), where \( \phi \) is the gravitational potential (\( \nabla^2 \phi = 4\pi G\rho \)).

This equation can be simplified which can more easily be seen if we write out the sum as two parts:

\[
\sum_{i=1}^{6} \frac{\partial (f \dot{w}_i)}{\partial w_i} = \sum_{i=1}^{3} \frac{\partial (f v_i)}{\partial x_i} + \sum_{i=1}^{3} \frac{\partial (f \dot{v}_i)}{\partial v_i} = \sum_{i=1}^{3} v_i \frac{\partial f}{\partial x_i} - \sum_{i=1}^{3} \frac{\partial \phi}{\partial x_i} \frac{\partial f}{\partial v_i} \tag{12}
\]

We can take \( v_i \) out of the partial differential with respect to time because \( v_i \) and \( t \) are independent variables. Similarly, the potential does not depend on velocity (in the last term we have rewritten \( \dot{v}_i \) in terms of the potential gradient. This gives us:

\[
\frac{\partial f}{\partial t} + \sum_{i=1}^{3} v_i \frac{\partial f}{\partial x_i} - \sum_{i=1}^{3} \frac{\partial \phi}{\partial x_i} \frac{\partial f}{\partial v_i} = 0. \tag{13}
\]

or we can write it in vector notation as:

\[
\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla f - \nabla \phi \cdot \frac{\partial f}{\partial \mathbf{v}} = 0. \tag{14}
\]

This is the collisionless Boltzmann equation or the Vlasov equation. Sometimes it is written as \( \frac{Df}{Dt} = 0 \), where \( D/Dt \) is the so-called convective or Lagrangian gradient and refers to the rate of change of \( f \) for an observer moving with the flow. This says that the value of the distribution function doesn’t change for a given star.

This equation is not easy to solve mostly because \( f(x, v, t) \) is a six-dimensional function. There are a range of applications for various simplified cases, such as spherical symmetry as we shall see later.

One key assumption we have made is that we can describe the system with the smoothly varying function \( f(x, v, t) \) (which comes from averaging over many stars in small regions of phase space). In particular we have assumed that the potential comes from a smoothly varying density function, or, in other words, encounters between individual stars have been ignored. This is fine as long as a star is only affected by the potential of many stars acting together (described by the distribution function) and doesn’t have it’s velocity significantly changed by encounters with individual other stars (which is not described by the distribution function). Generally, as \( N \) (the number of stars in the system) decreases, this becomes approximation becomes less and less accurate.

We can work out an estimate for when we can use the collisionless Boltzmann equation by computing how long it takes before the effects of individual encounters become noticeable. This timescale, usually called the relaxation time, is approximately given by

\[
t_{\text{relax}} = n_{\text{relax}} t_{\text{cross}} = \frac{N}{8\ln N} \frac{R}{v}, \tag{15}
\]

where \( t_{\text{cross}} = R/v \) is the crossing time for a spherical system of characteristic radius \( R \), characteristic velocity \( v \) and mass \( M \) (which we have assumed in the derivation to be related via the virial theorem such that \( v^2 = GM/R \)). The number of times a star must cross the system, \( n_{\text{cross}} \), before it has accumulated enough close encounters with individual stars depends mostly on \( N \).

For systems which are younger than the relaxation time, the effect of individual encounters is negligible and they can be well described by the collisionless Boltzmann, or Vlasov equation. Essentially all galaxies fall in this category. On the other hand, Globular clusters and clusters of galaxies are systems for which two-body scattering effects may be important. One way to model this effect is to add an extra term to the right-hand side of the Boltzmann equation to model the effects of collisions. In standard kinetic theory of gasses, where collisions are important, this is exactly what is done.

### 3 Disk Galaxies

#### 3.1 Radial Brightness profiles [Binney & Merrifield 4.4]

While ellipticals follow the de Vaucouleurs \( R^{1/4} \) profile, spiral galaxies have an exponential fall-off to their surface brightness profile. In order to measure this, it is necessary to measure the inclination angle of the
disk (the angle between the observer and the normal of the disk) and correct for the effect of this inclination. Once this is done, we find \( I(R) = I_0 \exp(-R/R_d) \), where \( R_d \) is the disk scale-height and typically takes values of a few kpc. In fact, this is a significant oversimplification because the vast majority of observed disks show a central excess which can be fitted by a spheroidal profile. This central spheroidal or bulge can be seen as a spherical addition in the centers of most spiral galaxies (including in the Milky Way). The relative importance of the disk and the bulge can be measured by decomposing the profile into the two components and computing the ratio of the total light in each component. This is known as the disk-to-bulge ratio \((B/D)\) and is correlated with the spiral type (Sa’s having a smaller value than Sb’s).

3.2 Stars in disk galaxies [Binney & Merrifield 10.3, 10.4]

As we noted above based on the radial light distribution, disk galaxies actually consist of two components: the disk and the bulge. This has long been known to be the case for our own Milky Way, for which an examination of the distribution of globular clusters and local high-velocity stars showed that there were actually two populations of stars. The stars in the disk (population I) tend to be younger, bluer and more metal rich, while the stars in the halo (population II) are older, redder and more metal poor. The fact that the halo appears to have formed first, followed by the disk, gives us an important hint about how the Milky Way formed.

The integrated vertical structure of the disk is often given as \( I(z) \propto \text{sech}^2(z/2z_0) \), where \( z_0 \), the vertical scale-height is typically around 0.5 kpc. This gives the disk its characteristic thin plate-like morphology. For the Milky-Way, however, considerably more is known about the disk and this simple prescription turns out to be an oversimplification. It appears that our disk itself consists of two populations, each described by an exponential power-low. One component (the thin disk) has a scale-height of about 300 pc and the other (the thick disk) of 1400 pc. While there is substantial overlap, it does appear that the populations of stars in these components differ in their physical properties as well (although the difference is not as clear cut as that between disk and halo stars). In particular, the thick disk stars have a larger velocity dispersion and are generally older and more metal-poor. It should be noted that the gas and dust out of which stars form today has a very short scale-height. There a natural interpretation is that the thick disks stars were originally formed close to the disk but were then heated, either by a sudden event such as the merger of a smaller galaxy, or by a gradual process such as scattering off of molecular clouds or spiral arms in the disk.

The halo of our galaxy is still poorly understood but new large-area surveys are providing information about the distribution of stars. For example, the density of stars appears to fall off roughly as a power-law with \( \rho \sim r^{-3.5} \) out to at least 50 kpc. Although the halo is largely smooth, there are some indications that it was not formed entirely at the same time. The recent discovery of the Sagittarius galaxy and its associated tidal streams indicate that stars are still being added to the halo, although it seems unlikely that the halo was entirely built by this process.

3.3 The Interstellar Medium

Besides stars, spirals galaxies like the Milky Way contain gas, which can be observed in a number of ways. One of the primary ways is via observations of the 21 cm emission from the hyperfine transition of neutral hydrogen (HI). Such observations of the external galaxies show that the HI disks often extend considerably further than the stellar disk. The lack of star formation at large radii is probably due to the fact that the disk at that point has a density below the critical density for gravitational instabilities to form.

The HI disk is also often observed to deviate from cylindrical symmetry. For example, spiral disks are often observed to be lopsided and have warps (the stellar disk can also warp). Some galaxies exhibit holes in the HI disk, due in part to ionization from recently formed massive stars or to supernovae explosions. In both cases, it usually requires a cluster of recent star formation to make a noticable hole.

One of the most important applications of HI observations is to measure the rotation curve of a galaxy and so determine the density distribution of gravitating mass. The gas (or stellar) redshift can be converted to a circular rotation speed which, assumes spherical symmetry, gives the enclosed mass via \( v_c^2 = GM(r)/r \). Such observations gave some of the first and strongest indications of the existence of dark matter. Unfortunately, for typical disks such as the Milky Way, the local disk is dominated by stars and gas rather than dark
matter (which is more important at larger radii). Therefore, a correction for the stellar component is necessary which requires assuming a mass-to-light ratio for the stars. This can be circumvented by make observations of so-called Low Surface Brightness (LSB) disks which have a particularly low gas density and so are dominated by dark matter over most or all of the disk. These observations are usually fit to one of two dark matter profiles:

$$\rho_{SO}(r) = \frac{\rho_0}{1 + (r/a)^2} \quad \rho_{NFW}(r) = \frac{\rho_0 a^3}{r(a + r)^2}$$ (16)

The first is similar to the profile for an isothermal sphere and has a constant density within the core radius $a$, while falling as $r^{-2}$ at large radii. The second profile comes from N-body simulations of Cold Dark Matter halos and exhibits a cusp at small radius, while falling steeply as $r^{-3}$ for $r >> a$. It appears that a substantial fraction of observed LSB galaxies are not good fits to the NFW profile, but it is still not clear if this is a true contradiction between CDM N-body simulation and observations or if one or other will change as the observations and models (which currently do not include stars and gas in a self-consistent way) improve.

HI observations often show a large hole in the center of the disk. Radio and sub-mm observations of molecular emission lines, in particular CO, often show emission from these regions, indicating that much of the gas is dense and cold enough to be molecular.

Infrared emission is associated with dust, which is mostly a good black-body although there are interesting spectral signatures at various wavelengths. Dust is a strong absorber of UV radiation and in starbursting galaxies most of the radiation is reprocessed by dust into the infrared.

3.4 Spiral Structure [Binney & Tremaine 6.2]

One of the most intriguing questions about spiral galaxies is how and what are the spiral arms. It seems very unlikely that they are fixed physical structures because the disk is differentially rotating: the outer parts have a lower angular velocity than the inner parts, causing any physical spiral structure to wind up on a timescale of few times the orbital times of the galaxy, which is typically a few hundred million years, much lower than the age of the entire system. Instead, the prevailing idea behind spiral structure rests on spiral density wave theory. This describes the (spiral) wave patterns that result from perturbations in self-gravitating stellar disks.

Stars in orbits which are nearly, but not quite, circular can be described as perturbations on top of a circular orbit. Here, we develop a linear evolution equation for the deviation of the orbit from circular. The acceleration of a star is described by Newton’s equation: \( \ddot{r} = -\nabla \phi \). In cylindrical co-ordinates with radius $R$, azimuthal angle $\phi$ and height $z$, this can be written as:

$$\ddot{R} - R \dot{\phi}^2 = -\frac{\partial \phi}{\partial R}$$

$$\frac{d}{dt} (R^2 \dot{\phi}) = 0$$

$$\ddot{z} = -\frac{\partial \phi}{\partial z}$$ (17)

These are the $R$, $\phi$ and $z$ components respectively, where we have assumed the disk potential is axisymmetric (i.e. $\phi(R, z)$ is a function of only $R$ and $z$). The middle equation is an expression of the conservation of angular momentum in the z-direction: $L_z = Rv_z = R^2 \dot{\phi}$ is a constant. We can use this definition to re-write the first equation as:

$$\ddot{R} = -\frac{\partial \phi}{\partial R} + \frac{\partial}{\partial R} \left( \frac{L_z^2}{2R^2} \right) = -\frac{\partial \phi_{eff}}{\partial R}$$ (18)

which we have re-written in terms of an effective potential $\phi_{eff} = \phi + L_z^2/2R^2$. The exact form of this effective potential depends on the potential well of the disk, but has a minimum at the radius $R_c$ of a circular orbit with angular momentum $L_z$. Since we are looking for small perturbations away from this position, we can perform a Taylor series expansion around the point $R_c$, defining the distance away from $R_c$ as $x = R - R_c$:

$$\phi_{eff} = \frac{1}{2} \frac{\partial^2 \phi_{eff}}{\partial R^2} x^2 + \frac{1}{2} \frac{\partial^2 \phi_{eff}}{\partial z^2} z^2 + \text{higher order terms}$$ (19)
Note that we have dropped the $\frac{\partial \phi}{\partial R}$ and $\frac{\partial \phi}{\partial z}$ terms because the unperturbed orbit is at a minimum in the potential (and the constant term can be set to zero because it has no physical meaning anyway). Putting this into the equation for $\ddot{R}$ (similar results apply for $\ddot{z}$), and noting that $\ddot{R} = \ddot{x}$, we find:

$$\ddot{x} = -\kappa^2 x$$

which describes simple harmonic motion with a frequency known as the epicyclic frequency:

$$\kappa^2 = \frac{\partial^2 \phi}{\partial R^2} + \frac{3L_z^2}{R_c^4}$$

The epicyclic approximation is useful because it describes the timescale for typical non-circular perturbations to oscillate (or grow) in the radial direction of a disk.

If we describe the rotation frequency of a star on a nearly circular orbit as $\Omega = v_c / R$ and its radial frequency as the epicyclic frequency $\kappa$, then the orbit is close back on itself if the ratio of $\Omega / \kappa$ is a rational number. More generally, the orbit doesn’t quite close but the point where the radial motion returns to its original position is some constant angle $\delta \theta$ ahead or behind of the starting angle. If we look at the system in a frame which is rotating at a speed such that we move this angle in one rotation of the circular orbit (i.e. at a frequency given by $\Omega_p = \Omega - n\kappa / m$ where $n$ and $m$ are integers), then the orbits do appear to close. Therefore if we can arrange the orbits to create a pattern (e.g. a spiral pattern), at one epoch then this pattern will rotate with a frequency given by $\Omega_p$ above. Of course, this only works if $\Omega_p$ is independent of $R$ so that the pattern at all radii moves with the same angular frequency. Remarkably, for typical spiral galaxy rotation curves this is true over a wide range of radii for selected values of $n$ and $m$. In particular, for $n = 1$ and $m = 2$ this holds; these orbits generate elliptical orbits and a slow rotation of the position angle of the ellipse with radius creates a two-armed spiral (see Figure 6-11 of Binney & Tremaine for an example of this).

This discussion is purely kinematical, it does not take into account the fact that the axisymmetric perturbations will change the potential and so modify the result. However, a more detailed investigation shows that a process known as swing amplification can actually help grow a small spiral pattern into a more pronounced one. While this effect is clearly important in many real galaxies, it is also quite possible that spiral structure is related to other physical processes, including the increased star formation created by the density enhancements.

### 3.5 Scaling relations: The Tully-Fisher relation and the Virial Theorem

Much like the Fundamental plane for elliptical galaxies, the Tully-Fisher relation is a correlation between the galaxy’s luminosity and the maximum rotational velocity of the gas, often measured via the HI 21 cm line width:

$$L \propto V_{\text{max}}^\alpha$$

where $\alpha \sim 4$ (although the value appears to depend on the wavelength at which the luminosity is measured).

This is strongly reminiscent of the Faber-Jackson relation, implying some common origin. Because we are dealing with self-gravitating, equilibrium systems, with turn to the virial theorem which provides a relation between the kinetic energy (i.e. velocity) and potential energy (which is related to mass or luminosity) for such systems. The virial theorem, which can be derived from the Boltzmann equation (by first multiply by the velocity and integrating over velocities to get the Jeans equation and then multiplying by position and integrating over all positions), can be stated as: $W = -2K$, where we define the potential and kinetic energies as:

$$W = -\frac{GM}{R_g} \quad K = \frac{1}{2} M \langle v^2 \rangle$$

To relate the mass $M$ to the luminosity $L$ of the galaxy, we use the usual definition $Upsilon = M/L$ and in addition define $C_v = \langle v^2 \rangle / v_{\text{max}}^2$ so that

$$L = \frac{R_g C_v}{G T} v_{\text{max}}^2$$

3
If we assume that the radius $R_g$ is linearly related to the scale length of the disk and that all spiral galaxies have the same surface brightness $I_0$ so that $L = \pi R_g^2 I_0$, then we get:

$$L = \frac{C_v^2}{\pi I_0 G^2 \Upsilon^2} V_{\text{max}}^4$$

which reproduces the scaling of the Tully-Fisher relation. While this appears successful, it hides a number of serious problems. The first is that the surface brightness of all galaxies is not constant – in fact, the scatter in the surface brightness of spirals is several orders of magnitude. However, the Tully-Fisher relation itself is very tight and so we have created something that is more than the sum of its parts: the final relation has a better correlation than its individual components.

Another problem is that rotation velocity curves are set in large part by the dark matter component (because the $v_c$ curves are flat, the dark matter must dominate beyond a few disk scale lengths). This means the amount of dark matter inside a few disk scale lengths must correlate with the mass of baryons. This is, at face value, a surprising result because the baryon distribution is set by its specific angular momentum content (i.e. when rotation balances gravity). while the dark matter structure is presumably set by something else (e.g. the velocity dispersion of the dark matter).

Despite these difficulties, it seems likely that we are on to important element explaining this relation. We can apply the same virial reasoning that we just used on the Tully-Fisher relation to the elliptical galaxies’ Fundamental plane. For example, we can again take the scalar virial theorem and define dimensionless quantities relating physical quantities to observables: $C_v = \langle v^2 \rangle / \sigma_0^2$ and $C_R = R_g / R_e$ where $\sigma_0$ is the central velocity dispersion of the elliptical galaxies. This gives us

$$R_e = \frac{C_v C_r \sigma_0^2}{G \pi \Upsilon \langle I \rangle_e}$$

where we have related the gravitating mass to the luminosity via the mass-to-light ratio $\Upsilon$ and employed the mean surface brightness inside the effective radius defined in the section on ellipticals. This bears some similarity to the Fundamental plane as we can see by taking the log of the above expression and writing the surface brightness in terms of magnitudes ($\mu$):

$$\log R_e = \log \frac{C_v C_r}{G \pi} + 2 \log \sigma_0 + 0.4 \mu - \log \Upsilon.$$  

The 0.4 exponent in front of $\mu$ is close to the observed 0.36, but $\sigma_0^2$ differs quite a bit from $\sigma_0^{1.4}$. Possible explanations are non-homology (i.e. the structure of the halo does not scale so $C_v$ is not really a constant but depends on $\sigma_0$), or “tilt”, an expression indicating the $\Upsilon$ varies with elliptical mass (or $\sigma_0$).

### 3.6 The Black-hole galaxy relation

Recently, it has become clear that all local galaxies with significant spheroidal components host supermassive blackholes in their centers. Even more remarkably, there is a fairly tight relation between the mass of the black-hole and the mass of the bulge. This (“Magorrian”) relation indicates that roughly 0.5% of the mass of the bulge is in the supermassive blackhole. It is sometimes written differently, in terms of the black-hole mass $M_{\text{BH}}$ and the velocity dispersion of the bulge or spheroid:

$$M_{\text{BH}} \propto \sigma^{4.5}.$$  

It is clear that these blackholes where quasars at some point in the past. In fact, assuming that these black holes grow from the inspiral of gas in an accretion disk accompanied by the emission of roughly 10% of the gravitational energy of the infalling mass (as General Relativity predicts), then the resulting radiative energy agrees well with the total energy that quasars are observed to produce. This comparison is known as the Soltan argument.

The fact that the mass of black holes is tightly correlated to the bulge mass tells us that black-holes are closely connected to their large halos. While it is not clear which way causality runs in this case (i.e. do galaxies control how much gas is fed onto black holes or do quasar outflows regulate star formation in spheroids), it is not hard to imagine that the vast amounts of energy involved in quasars must have a substantial impact on the galaxies that host them.
4 Distribution of galaxies in space

The large-scale distribution of galaxies in space is a very useful probe of cosmology because it is a reflection of the source of inhomogeneity in the early universe. Although we will largely stay away from large-scale structure in these notes, distribution statistics are important to us for two reasons. First, in order to convert the distribution of the galaxy to that of the underlying matter, we need to understand what is the relation between mass and galaxies (sometimes called the bias). Secondly, galaxies are themselves built up from inhomogeneities from the early universe so we need to know at least a little about this topic in order to understand how galaxies come together.

The first topic in this section should properly be a discussion of distance measures, in particular a discussion of the so-called distance ladder and how we actually work out the distance to extragalactic objects. However, this would take us too far outside the main development of the course and so we just refer readers to other source (for example, chapter 7 of Binney & Merrifield’s *Galactic Astronomy*).

4.1 The Correlation Function

The correlation function quantifies the clustering properties of galaxies above and beyond a uniform distribution. More precisely, the two-point correlation function $\xi(r)$ is defined via the probability of finding two galaxies separated by a distance $r$ in volume elements $dV_1$ and $dV_2$:

$$dP(r) = n_0^2[1 + \xi(r)]dV_1dV_2,$$

where $n_0$ is the mean density of galaxies and we have assumed the distribution is statistical isotropic. We can relate this to the overdensity distribution $\delta(x)$ through the definition $n(x) = n_0[1 + \delta(x)]$. Given this density distribution of galaxies, we can find the probability of finding a galaxy at $x$ AND at $x + r$:

$$dP(x, r) = n_0^2[1 + \delta(x)][1 + \delta(x + r)]dV_1dV_2.$$

Averaging this over all positions $x$ and assuming statistical isotropy, we get:

$$dP(r) = \langle dP(x, r) \rangle = n_0^2[1 + \langle \delta(x)\delta(x + r) \rangle]dV_1dV_2,$$

since averages over the overdensity ($\langle \delta(x) \rangle$) are zero by definition. This leaves us with a definition for the two-point correlation function in terms of averages over all positions:

$$\xi(r) = \langle \delta(x)\delta(x + r) \rangle$$

The observed galaxy two-point correlation function is well-fit by a power-law:

$$\xi_{\text{gal}}(r) = \left(\frac{r}{r_0}\right)^{-1.8},$$

where $r_0$ is the so-called correlation length and for typical bright galaxies is approximately $5h^{-1}\text{Mpc}$. Note that the correlation function depends on the objects being correlated (of course), with elliptical galaxies and clusters of galaxies being more highly correlated.

Note that even for an isotropic distribution, the two-point distribution does not contain all of the possible statistical information. Higher-order correlation function (three-point, etc) can be similarly defined and measured. For a Gaussian random field (more on this later), all of the higher terms can be expressed in terms of the two-point function, so it contains all of the statistical information of such a field.

4.2 The power spectrum

It is often useful to discuss correlation in terms of the Fourier transform of the real overdensity field $\delta(x)$:

$$\delta(x) = \frac{V}{(2\pi)^3} \int \delta(k)e^{-ik\cdot x}d^3k.$$
This is the usual Fourier transform definition, where \( V = L^3 \) is the volume of the box and the inverse transform is given by:

\[
d(\mathbf{k}) = \frac{1}{V} \int \delta(\mathbf{x}) e^{i\mathbf{k} \cdot \mathbf{x}} d^3\mathbf{k}.
\] (35)

We have used integrals in these expressions and so assumed that the box is arbitrarily large. Similar results apply for a finite sized periodic box with integrals replaced by sums. Keep in mind that the Fourier modes are plane waves in real space so each \( \delta(\mathbf{k}) \) is telling us the amplitude of an infinite plane wave. The superposition of many of these waves results in the observed real-space distribution of the overdensity field.

To relate this to the correlation function is straightforward. We evaluate equation (32) with the definition given above. To start, we write:

\[
\delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) = \frac{V^2}{(2\pi)^6} \int \delta(\mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{x}} d^3\mathbf{k} \int \delta(\mathbf{k}') e^{-i\mathbf{k}' \cdot \mathbf{x}} e^{-i\mathbf{k}' \cdot \mathbf{r}} d^3\mathbf{k}'
\]

\[
= \frac{V^2}{(2\pi)^6} \int \int d^3\mathbf{k} d^3\mathbf{k}' \delta(\mathbf{k})\delta(\mathbf{k}')^* e^{-i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{x}} e^{i\mathbf{k}' \cdot \mathbf{r}}
\] (36)

In the second line, we have re-arrange the integral. We have also replaced the integral of \( \mathbf{k}' \) by its negative and exploited the fact that the \( \delta(\mathbf{x}) \) function is real so that we can use the symmetry relation \( \delta(-\mathbf{k}) = \delta(\mathbf{k})^* \), where the asterisk indicates complex conjugation. The next step is to perform the average operation. Recall that the average is given by \( \langle \ldots \rangle = V^{-1} \int \ldots d^3x \) and we can interchange the integrals over \( \mathbf{x} \) and \( \mathbf{k} \) to get:

\[
\xi(\mathbf{r}) = \langle \delta(\mathbf{x})\delta(\mathbf{x} + \mathbf{r}) \rangle = \frac{V}{(2\pi)^6} \int \int \delta(\mathbf{k})\delta(\mathbf{k}')^* e^{i\mathbf{k}' \cdot \mathbf{r}} \int e^{-i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{x}} d^3x d^3\mathbf{k} d^3\mathbf{k}'
\] (37)

The integral over space is zero unless \( \mathbf{k} = \mathbf{k}' \) and otherwise it is equal to \( (2\pi)^3 \) (i.e. it is a delta function), so the expression simplifies considerably to:

\[
\xi(\mathbf{r}) = \frac{V}{(2\pi)^3} \int \delta^2(\mathbf{k}) e^{i\mathbf{k} \cdot \mathbf{r}} d^3\mathbf{k}
\] (38)

The quantity \( \delta^2(\mathbf{k}) \) is the power spectrum and we now see that the correlation function and the power spectrum are simply Fourier Transforms of each other. As usual, we can assume statistical isotropy and so only the magnitudes of the vectors \( \mathbf{k} \) and \( \mathbf{r} \) are important. Since \( P(k) = \delta^2(k) \) only depends on the magnitude \( k \), we can do some of the integral in the previous expression to get:

\[
\xi(\mathbf{r}) = \frac{V}{2\pi^2} \int P(k) \frac{\sin kr}{kr} k^2 dk
\] (39)

where we have used the fact that we are only interested in the real part of the transform.