4 Elliptical galaxies

4.1 Inferring the density from the surface brightness

The surface brightness profile, which essentially traces the distribution of stars, can be used to infer the underlying 3D matter density distribution.

The projected stellar surface density distribution is the integral over the 3D stellar density distribution. This in turn can be viewed as the total matter 3D density distribution scaled by a suitable mass–to–light ratio (in practice the use of a single M/L may be simplistic, though useful, assumption).

So what can we learn about the underlying 3D matter density distribution from the distribution of starlight?

Consider a power law distribution of 3D density, e.g. $\rho \propto r^{-\gamma}$. The projected surface brightness distribution will be a scaled version of the projected density, i.e.

$$I(R) \propto \int_0^\infty (z^2 + R^2)^{-\gamma/2} \mathrm{d}z \tag{1}$$

where the z-axis is defined by the line-of-sight to the observer. Taking $g = z/R$, this integral becomes

$$\int_0^\infty \frac{R\ \mathrm{d}g}{R^\gamma (g^2+1)^{\gamma/2}} = R^{-\gamma+1} G(\gamma) \tag{2}$$

where $G(\gamma) = \int_0^\infty (g^2+1)^{-\gamma/2} \mathrm{d}g$ and depends only on γ.

Therefore a power-law 3D density of slope $-\gamma$ projects onto a surface density profile of slope $-\gamma+1$.

One of the drawbacks of the de Vaucouleurs surface brightness distribution is that it does not have an analytic counterpart in 3D density. Various density profiles have been suggested that provide a good match to observed surface brightness profiles when projected, e.g. the Hernquist (1992) profile

$$\rho(r) = \frac{M}{2\pi} \frac{r_c}{r(r+r_c)^3}. \tag{3}$$

The Hernquist model is particularly appealing as it arises from the numerical simulation of the merger of two equal mass disk galaxies, each embedded within a dark matter halo.

4.1.1 Cuspy versus cored

Elliptical galaxies display a range of surface brightness properties from luminous, cored galaxies which show near constant surface brightness within some core radius and cuspy galaxies where the central surface brightness continues to rise to a sharp peak at inner radii.

Consider the slide comparing the galaxies NGC 1399 (a cD galaxy, $M_V = -21.9$) and NGC 596 ($M_V = -20.9$, about half as luminous).

Though the inner surface brightness of NGC 1399 is $I(R) \propto R^0$ i.e. constant, we can see that this corresponds to a mass density $\rho(r) \propto r^{-1}$. NGC 596 displays $I(R) \propto R^{-0.5}$ and thus $\rho \propto r^{-1.5}$ and possesses and even steeper central mass density profile.

Recall that $1 L_\odot \mathrm{pc}^{-2} = 26.4\ \mathrm{M_V}\ \mathrm{arcsec}^{-2}$. Therefore, the central surface brightness of NGC 1399 of $I_V(0) = 16$ corresponds to a stellar surface density of $\mathrm{dex}[(26.4 - 16)/2.5] = 14,450\ \mathrm{L_\odot pc^{-2}}$. The central stellar density of NGC 596 is approximately $5 \times 10^5 L_\odot \mathrm{pc}^{-2}$. Recall that the central stellar surface density of a spiral galaxy such as NGC 7331 reaches only 350 $L_\odot \mathrm{pc}^{-2}$.

These surface brightness trends were first quantified by Kormendy (1977). Luminous galaxies are increasingly core dominated according to the relation

$$\mu_e = 20.2 + 3 \log R_e \quad \text{or} \quad M_B = -19.3 - 2 \log R_e. \tag{4}$$

Taking $\mu_e = -2.5 \log I_e$ and $M_B = -2.5 \log L_B$, one obtains $I_e \propto R_e^{-1.2} \propto L^{-1.5}$ and $L_B \propto R_e^{0.8}$ (and notice that $L \propto I_e R_e^2$ as expected).

At this point we emphasize again that this relation holds for *bright* elliptical galaxies. As we shall see, faint ellipticals *appear* to follow a different linear relation.

The variation in the central surface brightness properties certainly points to variations in their evolutionary histories. Indeed, the debate as to whether bright, cored ellipticals and faint, cuspy ellipticals are separate populations or not still continues actively. We will consider this further below.

However, at this point we can note that within the hierarchical view of galaxy formation that sees massive galaxies constructed from the accretion of smaller sub units, the presence of lower surface density cores in bright galaxies may well arise from the increased energy present in stellar orbits resulting from past mergers.

4.2 Stellar population evolution

Stellar populations in elliptical galaxies are old and simple.

Combined with the absence of gas and dust in elliptical galaxies we can make some simple statements regarding their evolutionary history.

The absence of stars of spectral type A or earlier indicates that typical elliptical galaxies have not experienced major bursts of star formation within the last 5×10^8 years – the approximate main sequence lifetime of an A-star.

(In fact the presence of strong Balmer absorption in the spectra of elliptical galaxies – the signature of a significant A-star population – defines the "E+A" or "k+a" galaxy type: a post-starburst elliptical/early-type galaxy.)

This is consistent with the absence of gas and dust (often correlated with giant molecular clouds and active star forming regions).

We will see that later fundamental plane and colour magnitude diagram arguments applied to populations of ellipticals point to early, coeval star formation.

In the absence of young, bright OB stars, the light from ellipticals is dominated by the red giant population, i.e. $L \propto N_{rg}$.

The number of red giant stars at some time t will be equal to the number of stars with main sequence lifetimes $t - \Delta t_{rg} < t_{ms} < t$. Where Δt_{rg} is the red giant lifetime.

The luminosity-mass relationship for main sequence stars is $L \propto M^{\alpha}$, with $\alpha \approx 3$ for low mass stars.

This corresponds to a main sequence lifetime $t_{ms} \propto M/L = M^{1-\alpha}$ if we assume that a fixed fraction of the mass of each star is converted to energy during the main sequence lifetime.

With this relationship, the range of stellar ages contributing to the red giant population becomes a range of stellar masses, i.e. $M(t_{ms})$ to $M(t_{ms} - \Delta t_{rg}) = M(t_{ms}) + \Delta M$.

Defining the stellar initial mass function (IMF) as $\mathrm{d}N/\mathrm{d}M \propto M^{-1-x}$ as the number of stars in the mass range M to $M + \mathrm{d}M$, then the number of red giant stars can be written as

$$N_{rg} = \frac{\mathrm{d}N}{\mathrm{d}M} \times \Delta M = \frac{\mathrm{d}N}{\mathrm{d}M} \times \frac{\mathrm{d}M}{\mathrm{d}t_{ms}} \, \Delta t_{rg}. \tag{5}$$

Re-writing $t_{ms} \propto M^{-1/\theta}$ with $\theta = 1/(\alpha - 1)$, the number of red giant stars becomes $N_{rg} \propto t^{-1+\theta x}$. Note that we assume that the red giant lifetime is constant and thus acts only as a scaling constant.

Taking $x = -1.35$ (the Salpeter IMF) and $\theta = 1/3$, the luminosity of a galaxy dominated by red giant stars becomes

$$L \propto N_{rg} \propto t^{-0.6}. \tag{6}$$

These arguments were first formulated by Gunn, Tinsley and Larson in the mid 1970s. The 1980s and beyond saw the advent of the first stellar population synthesis codes for creating evolutionary tracks for integrated stellar populations, e.g. Bruzual and Charlot (1983). However, the above analysis approximates fairly well to the luminosity evolution of bright ellipticals.

4.3 Dynamics

4.3.1 Orbital structure

The 2D shapes of elliptical galaxies results from the 3D distribution of stars.

This in turn may be thought of as a reflection of the 3D orbital structure of the galaxy.

The debate as to the orbital structure of bright ellipticals was eventually resolved via resolved absorption line spectroscopy during the 1970s and early 1980s.

Ellipticals are not isotropic systems whose 2D morphology arises from rotational flattening.

They are slow rotators compared to their random velocity dispersion. The 2D morphology of bright ellipticals results from an anisotropic mix of stellar orbits.

As the stellar orbits are incoherent (both positive and negative along the line of sight) let us consider the mean square velocity $\langle v^2 \rangle$ along each axis. We refer to the mean square velocity along each axis as σ^2 (i.e. $\langle v_x^2 \rangle = \sigma_x^2$ where σ is the velocity dispersion).

Consider a galaxy rotating in the x-y plane (i.e. about the z-axis) with a rotational velocity V. The velocity dispersions are isotropic, i.e. $\sigma_x^2 = \sigma_y^2 = \sigma_z^2$.

The radial extent of stars along each axis should be proportional to their kinetic energies.

Assuming we view the system from the y-axis (side on), then a non zero axis ratio should result from rotational flattening according to

$$\frac{b}{a} \approx \frac{K_z}{K_x} \approx \frac{\sigma_z^2}{V^2 + \sigma_x^2}. \quad (7)$$

The 1D velocity dispersion $\sigma_r = \sigma_y$ which is equal to the other components following our isotropy condition.

We therefore have

$$\frac{b}{a} \approx \frac{\sigma_r^2}{V^2 + \sigma_r^2}, \tag{8}$$

which can be re-arranged to yield

$$\left(\frac{V}{\sigma}\right)^2 \approx \left(\frac{a}{b} - 1\right) = \frac{\epsilon}{1 - \epsilon}, \tag{9}$$

where $\epsilon = 1 - b/a$ is the observed ellipticity.

As can be see from the slides, to obtain $b/a = 0.5$ (an E5 elliptical) requires $V/\sigma \approx 1$. As a typical bright elliptical may display $\sigma = 250$ kms^{-1} this means it would have to rotate as fast as a massive spiral. In practice many bright ellipticals rotate much slower than this limit.

One concludes that rotational flattening does not contribute to the 2D shapes of ellipticals and that they are instead caused by anisotropic orbits.

Clearly some ellipticals do rotate to the extent that rotational flattening contributes to their morphology. These are generally lower luminosity ellipticals and their morphologies as often referred to as "disky" (as opposed to the "boxy" bright ellipticals) – thought to indicate a rotating stellar disk.

4.3.2 Faber-Jackson and the virial theorem

Faber and Jackson (1976) determined that the luminosity of bright ellipticals is related to their velocity dispersion via $L \propto \sigma_r^n$ with $3 < n < 5$.

More luminous ellipticals are more massive.

One can take a simple approach to the FJ relation by noting that $v^2 \propto GM/R$ and $L \propto I_e R_e^2$. This indicates that

$$L \propto \frac{v^4}{I_e (M/L)^2}, \tag{10}$$

which nominally reproduces the FJ relation if both the M/L ratios and surface brightness properties are relatively constant for bright ellipticals.

4.3.3 Faber-Jackson - a more detailed treatment

As we have noted, the FJ relation is a reflection of the virial theorem applied to elliptical galaxies, assuming that their surface brightnesses and mass-to-light ratios are relatively constant.

Beginning with the virial theorem we can write $2K + U = 0$ for $\ddot{I} = 0$. For the kinetic energy of the system we have

$$K = \frac{1}{2}\sum_i m_i v_i^2 = \frac{1}{2} M \langle v^2 \rangle. \tag{11}$$

We further note that $\langle v^2 \rangle = \sigma^2 = 3\,\sigma_r^2$, where σ_r is the 1D velocity dispersion.

The potential energy takes the form

$$U = \sum_{i>j} -\frac{G m_i m_j}{|r_i - r_j|} = -\frac{GM^2}{r_g}, \tag{12}$$

where r_g is a weighted average separation of the stars in the galaxy. Generally we may write

$$U = -\alpha \frac{GM^2}{R_e} \tag{13}$$

where α is a constant of order unity whose value depends upon the form of the density profile.

We can therefore write

$$\begin{aligned}
3M\sigma_r^2 &= \alpha \frac{GM^2}{R_e} \\
M &= \frac{3\sigma_r^2 R_e}{\alpha G} \\
L &= \frac{3\sigma_r^2 R_e}{\alpha G (M/L)} \\
L &\propto \frac{\sigma_r^2 R_e}{(M/L)} \quad \text{but we also have } L \propto I_e R_e^2 \quad \text{therefore} \\
L &\propto \frac{\sigma_r^2 (L/I_e)^{1/2}}{(M/L)} \\
L &\propto \frac{\sigma_r^4}{I_e^2 (M/L)^2}.
\end{aligned} \tag{14}$$

Which corresponds to the FJ relation assuming that I_e and M/L are relatively constant for bright ellipticals (see below).

4.3.4 Further maths how do we compute α?

Consider a uniform density sphere of radius r_s. The potential energy may be thought of as an integral over successive mass shells $\mathrm{d}M = 4\pi\rho\, r^2 \mathrm{d}r$, i.e.

$$\begin{aligned} U &= -G\int_0^\infty \frac{M(<r)\mathrm{d}M}{r} \\ &= -G\int_0^{r_s} \frac{16}{3}\pi^2\rho^2 r^4 \mathrm{d}r \\ &= -\frac{16}{15}\pi^2 G\rho^2 r^5, \end{aligned} \tag{15}$$

where we have assumed $M(<r) = \frac{4}{3}\pi\rho r^3$.

To obtain α we need to express r_s in terms of the projected half mass radius R_e.

We can consider the mass as a density weighted volume integral ($\int \rho\, \mathrm{d}V$) or as a surface density weighted area integral ($\int \sigma\, \mathrm{d}A$). Consider the change of variable $r^2 = x^2 + y^2 + z^2$ with the z-axis orientated toward the observer. We can further define the projected distance $b^2 = x^2 + y^2$.

We then integrate over the z axis, i.e.

$$\sigma(b) = 2\int_{z_{min}}^{z_{max}} \rho\, \mathrm{d}z \tag{16}$$

where the factor 2 represents the integral over two hemispheres. The z limits cover $|z| < (r_s^2 - b^2)^{1/2}$. We may express this as a radial integral following

$$\sigma(b) = 2\int_{r_{min}}^{r_{max}} \rho \frac{\mathrm{d}z}{\mathrm{d}r}\, \mathrm{d}r. \tag{17}$$

Taking $z_{min} = 0 \Rightarrow r_{min} = b$ and $z_{max} = \sqrt{r_s^2 - b^2} \Rightarrow r_{max} = r_s$.

Furthermore, with $z = (r^2 - b^2)^{1/2}$ we have $\mathrm{d}z/\mathrm{d}r = r/(r^2 - b^2)^{1/2}$ and the integral becomes

$$\sigma(b) = 2\int_b^{r_s} \rho \frac{r\, \mathrm{d}r}{(r^2 - b^2)^{1/2}} = 2\rho[(r^2 - b^2)^{1/2}]_b^{r_s} = 2\rho(r_s^2 - b^2)^{1/2}. \tag{18}$$

The projected mass computed using $\sigma(b)$ takes the form

$$M(<b) = \int_0^b 2\pi\sigma b\, \mathrm{d}b$$

$$
\begin{aligned}
&= 4\pi\rho \int_0^b b\,(r_s^2 - b^2)^{1/2}\mathrm{d}b \\
&= -\frac{4}{3}\pi\rho[(r_s^2 - b^2)^{3/2}]_0^b \\
&= \frac{4}{3}\pi\rho[r_s^3 - (r_s^2 - b^2)^{3/2}] \\
&= \frac{4}{3}\pi\rho r_s^3\left[1 - \frac{(r_s^2 - b^2)^{3/2}}{r_s^3}\right]. \qquad (19)
\end{aligned}
$$

We can now solve for $R_e = b$ as the projected radius which satisfies $M(< R_e) = 1/2M_{total} = 1/2 \times \frac{4}{3}\pi\rho r^3$, i.e.

$$
\begin{aligned}
1 - \frac{(r_s^2 - b^2)^{3/2}}{r_s^3} &= \frac{1}{2} \\
\frac{r_s^3}{2} &= (r_s^2 - b^2)^{3/2} \\
\frac{r_s^2}{2^{2/3}} &= r_s^2 - b^2 \\
R_e = b &= r_s\left(1 - \frac{1}{2^{2/3}}\right)^{1/2} = 0.608\, r_s. \qquad (20)
\end{aligned}
$$

4.3.5 The Fundamental Plane

Djorgovski and Davis (1987) were among a number of researchers to note that two parameter scaling relations such FJ and Kormendy contained real scatter in which the residuals in one plot correlated with those on the other. This suggested the existence of a three parameter relation encompassing the above relationships, i.e. a tilted plane of points in 3D of which the FJ and Kormendy relations are 2D projections.

The Fundamental Plane (FP) relation for bright ellipticals takes the form

$$\log R_e = 1.4 \log \sigma_e + 0.36\Sigma_e + \mathrm{constant}, \qquad (21)$$

where σ_e is the 1D velocity dispersion measured within R_e and $\Sigma_e = -2.5 \log I_e$.

The FP relation can be reconstructed using the following arguments

$$R_e^2 = \frac{L}{2\pi I_e} \quad (22)$$

$$R_e = \frac{M}{c\sigma_e^2} \quad (23)$$

with this second equation being a statement of virial equilibrium and c denoting a combination of physical constants.

Dividing these two equations one obtains

$$R_e = \left(\frac{c}{2\pi}\right)\left(\frac{M}{L}\right)^{-1} \sigma_e^2 I_e^{-1}$$

$$\log R_e = 2\log\sigma_e - \log I_e + \log\left[\frac{2}{2\pi}\left(\frac{M}{L}\right)^{-1}\right]$$

$$\log R_e = 2\log\sigma_e - 0.4\Sigma_e + \log\left[\frac{2}{2\pi}\left(\frac{M}{L}\right)^{-1}\right] \quad (24)$$

which is close to but not exactly equal to the observed FP relation.

We conclude that

1. Bright ellipticals are in virial equilibrium.
2. To 1st order M/L ratios and structural parameters are very similar.
3. Therefore, their stellar populations, ages and DM properties are very similar.
4. To obtain an exact match to observed FP data requires $M/L \propto M^{0.2}$, i.e. massive ellipticals are slightly older than less massive counterparts.

4.4 Reconciling bright and faint ellipticals

Morphological classification and magnitude cuts lead to the definition of giant and dwarf ellipticals as bulge-dominated systems with $M_V < -18$ and $M_V > -18$ respectively.

In addition, dwarf spheroidal galaxies form an ill-defined class of very faint spheroidal (i.e. not a disk and not irregular) galaxies with $M_V > -11$ or so.

Plotting such objects on 2D scaling relation diagrams (e.g. Kormendy) reinforces the idea that Es, dEs and dSphs are physically distinct classes of objects.

As such we need to explain their origins.

However, an alternative view is to consider them as essentially a single class of galaxy with a continuum of slowly changing physical properties.

Put another way, if one is to classify them as separate galaxy class, where should you draw the line? Is an elliptical galaxy with $M_V = -18.2$ physically distinct from an elliptical with $M_V = -17.8$? One manifestation of this continuum of physical properties is the variation of central (i.e. $R \approx 0.02R_e$) surface brightness properties as a function of magnitude.

Taking a sample of bulge dominated galaxies from either the Virgo of Fornax clusters one observes that bright ($M_V < -18$) galaxies display a central luminosity deficit with respect to a single de Vaucouleurs/Sersic model, i.e. an approximately constant surface brightness core. This core is modeled as a power law of slope α with a smooth transition to a larger scale Sersic profile – the so-called core-Sersic model.

At fainter magnitudes ($M_V > -18$), galaxies display a central luminosity excess with respect to a single model fit, i.e. a bright nuclear region fit with an additional Sersic component (a double Sersic model).

Viewing the sample of galaxies as a function of magnitude one observes that the trend from central deficit to central excess galaxies is relatively (there is some scatter) smooth and continuous.

Finally, one can plot all of the parameters normally considered in 2D scaling relations, e.g. μ_0, Sersic n, R_e, μ_e, $\langle\mu\rangle_e$, as a function of magnitude. This reveals a smooth, continuous variation of galaxy properties from giants, through dwarfs, to dwarf spheroidals.

Certainly if one selected any pair of these properties, e.g. μ_e and R_e (the Kormendy relation), one would compute different linear relations for dwarfs and giants and perhaps conclude that they represented different classes of galaxy.

Clearly though the changing gradient of quantities such as μ_e versus M_B do reflect different physical histories as a function of brightness (mass) and we consider these below.

4.5 The formation of elliptical galaxies

How do ellipticals form? In a hierarchical universe massive galaxies are predicted to be the result of the successive mergers of less massive galaxies (White and Rees 1978).

The weak dependence of the M/L ratios of bright ellipticals on their mass indicates that star formation – potentially associated with the merging/mass assembly process – occurred earlier in more massive galaxies. This is reasonable insofar that galaxies (or pre-galactic clumps) in dense regions of the universe would be expected to merger faster and earlier, leading to the production of more massive galaxies.

Merging produces an incoherent mix of stellar orbits.

As we have seen, long two-body relaxation times preserve the memory of individual encounters in the form of stellar streams, tidal tails and stellar shells.

However, violent relaxation – stellar encounters in a rapidly changing potential – results in more complete orbital mixing.

The products of N-body simulations recreating the merger of massive disk dominated galaxies produce slowly rotating, pressure-supported bulge dominated galaxies whose Hernquist-like 3D density distributions indicate that they would result in de Vaucouleurs-type surface brightness distributions (Hernquist 1992).

The central surface brightness profiles of bright ellipticals are "core-like", i.e. flattening toward the centre. This suggests that they formed via "dry" or gas-poor merging. This is a collisionless process – little angular momentum is lost via two-body encounters – and the additional orbital angular momentum of the merging galaxies results in a "puffing up" of the central stellar distribution.

However, as one moves to fainter ellipticals there is a relatively smooth trend to observe high-surface brightness stellar nuclei. This suggests an increasing importance of "wet" or gas-rich mergers as one considers lower mass ellipticals.

Gas-rich mergers are collisional and the orbital angular momentum of the merging galaxies can be dissipated away. This allows the gas to accumulate in the centres of such systems, potentially triggering dense, nuclear starbursts.

Bright ellipticals and faint ellipticals form a single family; they are not fundamentally different objects. This can be seen from plots showing the continuity of scaling relations of red-sequence selected galaxies (ellipticals).

The relations linking bright and faint ellipticals are not linear. Put another way, if you tried to define linear scaling relations (e.g. the Kormendy relation) for bright and faint ellipticals you would compute different gradients and assume that the objects are different.

However, it is clear that different physical effects have played a role in the evolution of bright and faint ellipticals and that there appears to have been a relatively smooth transition between competing effects as a function of luminosity (mass).

Elliptical galaxies dominate the galaxy populations in the centres of rich clusters. However, the Butcher-Oemler effect indicates that the ellipticals we observe in rich clusters today have undergone relatively recent transformation.

This suggests that merging/interactions, accentuated by ram pressure stripping, have played a role in their formation.