Dark Matter in Spiral Galaxies [and Discussion]

T. S. Van Albada; R. Sancisi; Maria Petrou; R. J. Tayler


Stable URL:
http://links.jstor.org/sici?sici=0080-4614%2819861217%29320%3A1556%3C447%3ADMISG%5B%3E2.0.CO%3B2-O

Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences is currently published by The Royal Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/rssl.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The JSTOR Archive is a trusted digital repository providing for long-term preservation and access to leading academic journals and scholarly literature from around the world. The Archive is supported by libraries, scholarly societies, publishers, and foundations. It is an initiative of JSTOR, a not-for-profit organization with a mission to help the scholarly community take advantage of advances in technology. For more information regarding JSTOR, please contact support@jstor.org.
Dark matter in spiral galaxies

BY T. S. van Albada and R. Sancisi
Kapteyn Astronomical Institute, Groningen University, Postbus 800, 9700 AV Groningen, The Netherlands

Mass models of spiral galaxies based on the observed light distribution, assuming constant $M/L$ for bulge and disc, are able to reproduce the observed rotation curves in the inner regions, but fail to do so increasingly towards and beyond the edge of the visible material. The discrepancy in the outer region can be accounted for by invoking dark matter; some galaxies require at least four times as much dark matter as luminous matter. There is no evidence for a dependence on galaxy luminosity or morphological type. Various arguments support the idea that a distribution of visible matter with constant $M/L$ is responsible for the circular velocity in the inner region, i.e. inside approximately 2.5 disc scalelengths. Luminous matter and dark matter seem to 'conspire' to produce the flat observed rotation curves in the outer region. It seems unlikely that this coupling between disc and halo results from the large-scale gravitational interaction between the two components. Attempts to determine the shape of dark halos have not yet produced convincing results.

1. Introduction

The aim of this paper is to review the observational evidence for dark matter in spiral galaxies, and to discuss some of its consequences. That dark matter is needed can be inferred from a comparison of the distribution of light and the rotation curve. Considerable progress has been made in this area in recent years. The amount of dark matter detected so far in spiral galaxies is rather modest from the cosmological point of view. Total mass:light ratios of individual galaxies in blue light may approach $20 M_\odot/L_\odot \times H_0/75 \text{ km s}^{-1} \text{ Mpc}^{-1}$, but the corresponding mass density is still only a few hundredths of the density needed to close the Universe.

Since the concept of dark halos surrounding galaxies now seems to be widely accepted it may be useful to recall that the underlying facts were still disputed only a few years ago. At the Besançon IAU symposium, Kalnajs (1983) pointed out that the approximately flat rotation curves (i.e. circular velocity in the disc of the galaxy as a function of radius) of four galaxies could be reconciled with the observed radial distributions of light by adopting an appropriate constant value for the mass:light ratio, $M/L$. Another case is UGC 2885, an Sc galaxy of high luminosity. Because of its flat rotation curve extending to about 80 kpc (Rubin et al. 1980; Rubin et al. 1986; Roelfsema & Allen 1985; $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$), UGC 2885 has been regarded as a pre-eminent example of a galaxy with a dark halo, even though the light distribution had not been measured. CCD photometry obtained recently by Kent (1985) shows, however, that the light profile produces an approximately flat rotation curve out to 60 kpc; only beyond 70 kpc is there an indication for the need of dark matter (see figure 1).

† Distances in this paper are based on $H_0 = 75 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (1 pc $\approx 30857 \times 10^{12} \text{ m}$), but results from cited papers using a different Hubble constant have been left unchanged. In those cases the value of $H_0$ used is indicated.
From these examples it is clear that a flat rotation curve in itself does not imply the presence of dark matter. To establish the existence of dark halos it is necessary to measure the circular velocity well beyond the turnover radius of the disc (i.e. the radius where the predicted rotation curve of luminous matter in the disc starts to decline; for an exponential disc this occurs around 2.2 disc scalelengths). For this purpose 21 cm radio observations of neutral hydrogen gas in the outermost regions are indispensable. In the inner regions, optical rotation curves are often superior because beam smearing effects or a deficiency in \( \text{HI} \) can be a severe limitation to the radio measurements.

The plan of this paper is as follows. We first discuss in §2 the evidence for dark matter, and introduce the hypothesis that luminous matter alone is responsible for the circular velocity in the inner region ('maximum-disc' hypothesis). In §3 we discuss constraints on the relative amounts of dark and luminous matter in the inner regions from various sources, and we conclude that there is strong support for the maximum-disc hypothesis. This conclusion implies that the more or less constant value of the rotation velocity in the outer region is in some way linked to the luminous matter in the inner region. Additional consequences of the maximum-disc hypothesis are discussed in §4. In §5 we briefly discuss self-consistent halo models to see whether the coupling of disc and halo is related to the gravitational interaction of the two components. Evidence on the shape of dark halos is reviewed in §6 and we summarize our main conclusions in §7. Several of the issues considered in this paper have also been addressed by Bosma (1978, 1981 a, b), Bosma & van der Kruit (1979), Faber & Gallagher (1979), Petrou (1981), Bahcall & Casertano (1985), Carignan & Freeman (1985), Rubin (1983, 1986), Rubin et al. (1986), van Albada et al. (1985), Sancisi & van Albada (1985) and Burstein & Rubin (1985).

2. Evidence for dark matter

(a) Observed and predicted rotation curves

The approximately flat nature of rotation curves outside the optical radius was first demonstrated convincingly by Bosma (1978, 1981 a, b) from 21 cm line aperture-synthesis observations. \( \text{HI} \) rotation curves obtained recently (Begeman 1986) reach out even farther than
those of Bosma and have higher precision. A few H I rotation curves for galaxies of different luminosities are shown in figure 2. After an initial rise, all curves are approximately flat out to the last measured point at 5 or more scalelengths (h). If the gravitational field derives from the distribution of luminous matter, with constant M/L, the curves should begin to decline beyond about 2.5 disc scalelengths. The discrepancy is attributed to the presence of dark matter.

![Figure 2. H I rotation curves for a number of spiral galaxies (Sanctis & van Albada 1986). Distances are based on \( H_0 = 75 \) km s\(^{-1}\) Mpc\(^{-1}\). The optical radius, \( R_{\text{opt}} \), and the number of disc scalelengths, h, at the last measured point are indicated. For the inner region of UGC 2885 optical velocities (Rubin et al. 1986) have been used. All curves remain approximately flat beyond the turnover radius of the disc (2.5 h).](image)

The amount and distribution of dark matter can be derived by calculating rotation curves from the distributions of light and by comparing these with the observed rotation curves. This approach presupposes that the distribution of matter in a galaxy can be described with a number of distinct components, e.g. a bulge, a disc and a dark halo. Bulges and discs are assumed to have M/L values that are independent of radius. The bulge is generally represented by an \( R^\alpha \) law spheroid (de Vaucouleurs 1948, 1953), for which the potential has been tabulated by Young (1976). The rotation curve of the disc can be calculated by using Bessel functions (Toomre 1963). The dark halo can be modelled with an isothermal sphere (Carignan & Freeman 1985) or with a suitable family of density distributions (see, for example, van Albada et al. 1985). (A problem is that often the separation of the light profile into a bulge and a disc cannot be done in a unique way.) It is clear that this 'building block' approach is a highly idealized one. There is no reason why the mass:light ratio of the disc could not depend on radius. In fact, radial variations in the rate of star formation make such a dependence seem likely. Yet, there appears to be no evidence for large radial gradients in the stellar population characteristics (e.g. large colour gradients are rare).
(b) The maximum-disc analysis

Because the mass:light ratios for bulge and disc are not known \textit{a priori} a first step is to maximize the contribution of luminous matter to the observed rotation curve (‘maximum-disc’ hypothesis). This yields a lower limit for the amount of dark matter. Note that such a maximum-disc model does not specify the \textit{nature} of the material in the disc. All the maximum-disc hypothesis purports to test is the constancy of $M/L$ with radius and to derive a lower limit for the amount of dark matter.

Results for two galaxies, NGC 2403 and 3198, showing that there is a discrepancy between observed and predicted rotation curves in the outer regions, are shown in figure 3. In both cases the light profile (upper panels) is close to a straight line, i.e. surface brightness decreases exponentially with radius. The background level of the sky is 21.05 mag arcsec$^{-2}$, F-band. Five magnitudes below the sky level the uncertainty in the light profiles is approximately 0.2 mag (Wevers 1984). In the lower panels observed and predicted rotation curves are compared. The observed rotation curves extend to 9.5 and 11 disc scalelengths for NGC 2403 and 3198, respectively. The predicted rotation curves contain an arbitrary scaling factor (the $M/L$ ratio), chosen in accordance with the maximum-disc hypothesis. The discrepancy between the two curves starts near the turnover radius of the disc. This occurs at a surface brightness level of about 22.6, i.e. about 1.5 mag below the sky level.

A detailed disc-halo model for NGC 3198 has been published by van Albada \textit{et al.} (1985). A similar mass model for NGC 2403 is shown in figure 4. The observed rotation curve has been
decomposed into a (maximum) contribution by the 'luminous' material, calculated from the light profile and from the distribution of H I gas (including a correction for helium), and a contribution by the dark halo. For this model the ratio $M_{\text{dark}}/M_{\text{lum}}$ at 2.5 times $R_2$ is 0.6, at $R_2$ it is 1.1, and at the H I radius 3.8. For NGC 3198 the latter two numbers are, respectively, 0.8 and 3.9. Similar values are given by Sancisi & van Albada (1986) for slightly different disc–halo decompositions (as in figure 3). The models for NGC 3198 and 2403 show that a good fit to the rotation curve is obtained by choosing the maximum disc and a halo. It is not possible, however, to get a lower limit for the $M/L$ ratio of the disc; equally good fits can be obtained for any assumed disc mass less than the maximum value (see figure 5). This question is discussed further in §3.

HI rotation curves extending to at least 10 disc scalelengths have been measured for four additional galaxies. In all cases the rotation curve is approximately flat out to the last measured point, whereas the light profile is close to an exponential. Thus the results are very similar to those for NGC 2403 and 3198. Parameters for these six systems are given in table 1. (Values of mass $M_T$ are based on spherical symmetry.) The $M/L$ values at the last measured points are among the largest known for spiral galaxies. Note that they are still an order of magnitude smaller than those for clusters of galaxies.

Of particular interest is the question of whether or not low-mass galaxies are also surrounded by dark halos, and, if so, whether there is a dependence on luminosity. Mass distributions in four late-type spirals of low luminosity have been studied by Carignan & Freeman (1985). These systems have maximum circular velocities in the range 60–120 km s$^{-1}$. For three galaxies the rotation curves as measured from single-disc HI observations continue to rise or remain approximately flat beyond the turnover radius of the disc. Thus in these low luminosity galaxies...
Table 1. Parameters for six spiral galaxies with exponential light profiles and extended HI discs

<table>
<thead>
<tr>
<th>NGC</th>
<th>type</th>
<th>distance</th>
<th>$h$</th>
<th>$R_{25}$</th>
<th>$R_{\text{max}}(\text{H}I)$</th>
<th>$\frac{R_{\text{max}}(\text{H}I)}{h}$</th>
<th>$v_{\text{Bmax}}$</th>
<th>$M_T$</th>
<th>$L_B$</th>
<th>$M_T/L_B$</th>
<th>$R &lt; R_{\text{max}}(\text{H}I)$</th>
<th>references</th>
<th>rotation curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>2403</td>
<td>Sc</td>
<td>3.2</td>
<td>2.1</td>
<td>8.5</td>
<td>20</td>
<td>9.5</td>
<td>134 ± 4</td>
<td>8</td>
<td>0.8</td>
<td>10</td>
<td></td>
<td>Wevers</td>
<td>(1984) Begeman (1986)</td>
</tr>
<tr>
<td>2841</td>
<td>Sb</td>
<td>9.0</td>
<td>3.1</td>
<td>10.7</td>
<td>36</td>
<td>11.6</td>
<td>280 ± 10</td>
<td>65</td>
<td>2.5</td>
<td>26</td>
<td></td>
<td>Boroson</td>
<td>(1981) Bosma (1981 a)</td>
</tr>
<tr>
<td>5055</td>
<td>Sbc</td>
<td>8.0</td>
<td>4.1</td>
<td>14.4</td>
<td>41</td>
<td>10.0</td>
<td>180 ± 10</td>
<td>31</td>
<td>2.8</td>
<td>11</td>
<td></td>
<td>Wevers</td>
<td>(1984) Begeman (1986)</td>
</tr>
<tr>
<td>6503</td>
<td>Sc</td>
<td>7.0</td>
<td>2.0</td>
<td>6.2</td>
<td>25</td>
<td>12.5</td>
<td>115 ± 5</td>
<td>8</td>
<td>0.7</td>
<td>11</td>
<td></td>
<td>Wevers</td>
<td>(1984) Begeman (1986)</td>
</tr>
</tbody>
</table>
there is a similar discrepancy as found for the more massive systems discussed above. \( \frac{M_{\text{dark}}}{M_{\text{lum}}} \) values based on the maximum-disc fit do not vary much with total luminosity (as expected from the similarity in shapes of rotation curves; see §4). Over a range of 100 in luminosity the variation may not be more than a factor 2 or 3, with no clear systematic trend (cf. Bahcall & Casertano 1985).

(c) Discrepancies in the inner region

In general the discrepancy between observed and predicted curves starts at the point where the predicted curve reaches its maximum. For a small number of galaxies, however, the measured rotation curve differs considerably from that predicted by the light profile also in the inner region. In these cases the discrepancy is, as far as we know, apparent only, and can be traced to a problem with the light profile or rotation curve. These problems can be illustrated with NGC 5907 and the Sombrero nebula.

In NGC 5907 the observed rotation curve rises fairly rapidly near the centre (in accordance with the systematics of Rubin et al. (1986); see §3b). The rotation curve predicted from the light profile, as measured by van der Kruit & Searle (1981, 1982), rises much less steeply (see figure 7 in Sancisi & van Albada 1986). Thus there is a clear need for an additional component of matter. However, in edge-on galaxies a large fraction of the light is obscured by dust. Because of this the light profile measured in a strip at a fixed distance from the symmetry plane need not be representative for the overall distribution of luminous matter, and we believe that the ‘missing material’ may well be luminous material obscured by the dust.

In the Sombrero nebula (NGC 4594) the measured rotation curve in the inner region \((r < 45'' = 4.0 \text{ kpc})\) rises nearly linearly with modest slope (Schweizer 1978; Rubin et al. 1986) and falls below that expected from the well-defined \(R^1\) law bulge (photometry by Boroson 1981). It clearly deviates from the expected behaviour for a high luminosity Sa galaxy. Here, too, it seems likely that the problem is caused by the near edge-on orientation of the system \((i \approx 84^\circ)\). Gas in Sa galaxies is not abundant and often seems to be concentrated in a central disc with radius less than 1 kpc and a number of rings. In the Sombrero nebula the linearly rising part of the rotation curve might be caused by such a nearly edge-on ring with radius ca. 45". A more general model with a slightly warped gas layer (vertical scale height 150 pc), with a relative depletion of gas between 0.5 and 3.5 kpc, has been made (van Albada, unpublished results). Spectral line shapes have been calculated using the rotation curve predicted by the \(R^1\)-law bulge, scaled to the observed rotation curve at 45", and taking finite slit width, seeing and turbulent motions into account. With an ad hoc model of this type it is easy to explain both the linear rise of the observed rotation curve inside 45" and the spike near the centre (which is attributed to the central disc).

3. Constraints on the relative amount of luminous and dark matter

In the preceding section we have been concerned with a determination of the minimum amount of dark matter associated with spiral galaxies on the assumption that \(M/L\) of luminous matter is constant (separately for bulge and disc), and equal to the maximum value allowed by the rotation curve. In order to test the validity of this maximum-disc hypothesis we review below information from various sources on the \(M/L\) ratio in galaxies. We argue that the amount of matter in the disc is likely to be close to the maximum allowed value.
(a) Detailed shape of rotation curves

With few exceptions, the shapes of predicted and observed rotation curves are very similar over a large fraction of the visible part of galaxies for which mass models have been made (in total about 50 systems, including work in progress by S. M. Kent (personal communication)). In several cases, even relatively small features in the observed rotation curves are reproduced by the rotation curves calculated from the light profile. We regard this as strong support for the maximum-(bulge)disc hypothesis.

(b) Systematics of rotation curve shapes

The study of rotation curves by Rubin et al. (1986 and references therein) for approximately 60 galaxies with a large range in luminosity, and with Hubble types ranging from Sa to Sc, shows that the shapes of rotation curves depend on luminosity. Within a given Hubble type, low-luminosity galaxies have rotation velocities that rise gradually from the nucleus and reach a maximum at the isophotal radius, whereas high luminosity galaxies have velocities which rise steeply from the nucleus and become approximately flat at a smaller fraction of the isophotal radius. Velocity amplitudes increase with increasing luminosity. This pattern is observed for Sa, Sb and Sc galaxies separately. Rubin et al. (1986) and Rubin (1985) use this progression of rotation curve shapes with luminosity, and the absence of a clear correlation with Hubble type, to argue in favour of the presence of dark matter also in the inner region of galaxies. Luminous and dark matter together, in some way coupled to each other but with different spatial distributions, determine the circular velocity at all radii in a galaxy. However, the aforementioned analysis by Kent for nearly all Sb and Sc galaxies of Rubin et al., shows that there is no problem with the idea that luminous matter alone explains the rotation curve inside 2.5 h. It would seem that the pattern found by Rubin et al. implies that the character of the mass distribution in the disc, or perhaps the bulge:disc ratio, depends on luminosity. (The latter possibility is suggested by a comparison of the Sc galaxies NGC 3198, of intermediate luminosity and without a bulge, and UGC 2885, of high luminosity and with a conspicuous bulge.)

(c) The Tully–Fisher relation

A strong correlation exists between total luminosity and amplitude of the rotation curve, this is the Tully–Fisher relation (Tully & Fisher 1977); the higher the luminosity the higher the maximum circular velocity. The small scatter in this relation suggests that the amount of visible matter determines the maximum rotation velocity in galaxies (in agreement with point (a) above). If this were not the case the amount of dark matter inside, say, 2.5 disc scalelengths must be related in a unique way to the amount of visible matter. For a more complete discussion see van Albada et al. (1985).

(d) Presence of two-armed spiral structure in galaxies

Although no comprehensive theory of the formation and maintenance of spiral structure in disc galaxies is available yet, the conditions needed for the amplification of perturbations are fairly well understood, and can be worked out analytically (Goldreich & Lynden-Bell 1965; Toomre 1981), or can be experimentally obtained from N-body calculations (Sellwood 1985). A detailed model for the case of NGC 3198 has recently been made by J. A. Sellwood (personal communication). He finds that the maximum-disc case with $v_{\text{max}} = 140$ km s$^{-1}$
allows the formation of two-armed spiral structure, as observed. If the disc mass is lowered to 50% of the maximum disc mass (and the remaining mass is unresponsive), the growth of two-armed spiral structure is definitely inhibited and only multiple-armed features develop. Well-developed two-armed spiral structure appears possible only if the disc mass is at least 70% of the maximum disc mass. The corresponding value of $M_{\text{halo}}/M_{\text{disc}}$ inside 2.5 $h$ is 1.0 (see, however, the note added in proof). Arguments of this type based on Toomre’s local theory have been used by Athanassoula et al. (1986) to constrain the $M/L$ ratio of luminous matter in several galaxies.

(e) Amplitude of streaming motions

Streaming motions in the gaseous component, associated with grand-design spiral structure, have been observed in several galaxies, e.g. M81 (Rots & Shane 1975; Rots 1975; Visser 1980). No quantitative models that include a dark halo contribution to the gravitational field have been worked out, but one would expect that the velocity amplitude of streaming motions depends not only on the amplitude of the spiral perturbation in the disc (which can in principle be measured), but also on the relative contributions of the disc and of the (smooth) dark halo to the total gravitational field.

(f) Truncation drop in rotation curves

If a stellar disc has a sharp boundary, as seems to be the case for some galaxies (van der Kruit & Searle 1981, 1982), the rotation curve of the disc is expected to drop steeply beyond the truncation radius provided that the truncation occurs inside about four scalelengths (van der Kruit & Searle 1982; Casertano 1983; Bahcall 1983). If luminous matter dominates the gravitational field this drop-off should be observable in the $H_1$ rotation curve. In principle the velocity gradient provides a direct measure for the $M/L$ ratio of the luminous disc. NGC 5907 has for some time been regarded as a good example. On both sides of this edge-on galaxy the rotation curve shows a drop-off at 6' from the centre. Casertano (1983) identified this drop-off with the truncation effect described above, basing his conclusion on a preliminary model for the light distribution with a sharp truncation. The decline of the actual light profile (van der Kruit & Searle 1981, 1982), is rather smooth however and the rotation curve for this light profile, after rectification to face-on geometry, does not show a truncation feature (see figure 7 of Sancisi & van Albada 1986).

(g) The mass:light ratio in the solar neighbourhood

The mass:light ratio $M/L_\nu$ in the vicinity of the sun is about 3 (Bahcall 1984). This is of the same order as the $M/L$ value for the maximum-disc case for NGC 2403 and 3198, indicating that the maximum-disc solution gives plausible $M/L$ ratios.

(h) Mass:light ratio from stellar evolution theory

Rubin et al. (1986) found mean mass:light ratio’s $M/L_B$ within $R_{25}$ of 6.2, 4.5 and 2.6 for respectively Sa, Sb and Sc galaxies ($H_0 = 50$ km s$^{-1}$ Mpc$^{-1}$). Because there is already a discrepancy between observed and predicted rotation curves at $R_{25}$, the above $M/L$ values include a contribution by the dark halo. Allowing for the minimum amount of dark matter present (maximum disc) these $M/L$ values should be multiplied by about 0.6 (estimated from available mass models for a small number of galaxies), yielding $M/L_B = 3.7, 2.7$ and 1.6 for Sa, Sb and Sc galaxies. Mean B–V colours for these types are 0.75, 0.64 and 0.52. Models of
evolving discs (Larson & Tinsley 1978) also show that early type spirals, with their redder colours, should have highest $M/L$ ratios: $M/L_B = 3.1$, 2.0 and 1.0 for the B–V values corresponding to Sa, Sb and Sc galaxies; see Rubin et al. (1985). Although precise normalization is difficult because of the uncertainty in the Hubble constant and the unknown contribution to the mass density by low luminosity stars, the agreement between observed and predicted trends, and the fair overall agreement of the $M/L$ values, provides support for the maximum-disc hypothesis.

(i) Conclusions: the disc–halo conspiracy

We conclude from the above discussion that the amount of dark-halo material in the inner region ($R < 2.5 h$) is small with respect to the amount of luminous matter. Unfortunately formal limits on $M_{\text{dark}}/M_{\text{lum}}$ are not very strict. The possibility to have well developed two-armed spiral structure (see point (d) above) requires $M_{\text{dark}}/M_{\text{lum}} < 1.0$, but the detailed agreement between observed and predicted shapes of rotation curves (point (a)), in combination with the Tully–Fisher relation (point (c)), indicates that this upper limit is too generous.

The dominance of luminous matter inside 2.5 disc scalelengths implies that luminous matter also controls the approximately constant value of the circular velocity in the outer regions. (We recall that the term luminous matter is used in this paper to describe matter lying in bulge or disc and characterized by a constant $M/L$ value. The nature of this material is not specified.) This coupling of disc and halo can be illustrated with different mass models for NGC 3198, one based on the maximum disc and the others based on discs with masses which are fractions of the maximum disc mass: 0.75, 0.50 and 0.25 respectively. These different decompositions of the observed rotation curve are shown in figure 5. It is clear that if the disc mass exceeds 70% of the maximum disc mass the flat behaviour of the observed rotation curve between 3 and 11 disc scalelengths is due to a falling rotation curve for the disc and a rising one for the dark halo. It seems impossible that two supposedly independent components of a galaxy, a flat disc and a more or less spherical halo could produce such a result as a rule. In some way the distributions of luminous and dark matter must be closely coupled through their formation (cf. Petrou 1981; Bahcall & Casertano 1985). At present this coupling, which may also be called the disc–halo conspiracy, is not understood.

This problem appears to be more general and complex than just a coupling between disc and halo. Consider again the rotation curve of UGC 2885 (figure 1). Between 2 and 50 kpc from the centre a flat rotation curve results from the combination of a declining curve for the bulge and a rising one for the disc. Beyond 60 kpc dark matter is needed with a rising rotation curve to balance the declining curve of the disc. Clearly, this rather odd situation raises the question why seemingly independent components, bulge, disc and halo should conspire to produce a circular velocity that is constant within about 10%. It would seem that some process during galaxy formation requires $v_{\text{cir}}$ to be nearly independent of radius.

A way to avoid the disc–halo conspiracy, and still keep luminous matter dominant in the inner region, would be to release the initial assumption of constant $M/L$ for the disc. The $M/L$ ratio could be approximately constant in the inner regions and increase gradually in the outer parts with a functional form $M/L(r/h)$ similar for all galaxies. Probably the $z$-thickness of such a disc would also have to increase with radius (see §6). In alternative, less conventional, explanations of the flatness of rotation curves far beyond the optical image an inadequacy of Newtonian dynamics or gravity has been invoked (Milgrom 1983; Sanders 1984).
4. Some Consequences of the Maximum-Disc Hypothesis

The maximum-disc hypothesis provides a convenient way to derive some simple analytical relations illustrating the coupling between light and matter in galaxies. To do so we consider two extreme cases: (i) dark matter, like luminous matter, resides in a disc and (ii) dark matter forms a more or less spherical halo around the galaxy. We assume that the light profile is exponential and that the observed rotation curve is flat beyond the turnover radius of the disc (see also Petrou 1981).

(a) A dark disc

In a disc, the rotation velocity at radius $r$ depends on the distribution of matter both inside and outside this radius. For a galaxy with a rotation curve that remains flat far beyond radius $r$ the surface mass density is

$$\sigma_{\text{tot}}(r) \approx v^2/(\pi^2 G r);$$

the total mass inside $r$ follows from:

$$M_{\text{tot}}(r) \approx 2v^2r/(\pi G).$$

By comparison, for an exponential disc with scale length $h$ (de Vaucouleurs 1959; Freeman 1970), ($M/L$ constant):

$$\sigma_{\text{lum}}(r) = \sigma_{\text{lum}}(0) e^{-r/h},$$

with

$$\sigma_{\text{lum}}(0) = M_{\text{lum}}/(2\pi h^2) = v_{\text{max}}^2/(0.78 \pi G h).$$
Equation (5) indicates that the local mass:light ratio in a galaxy with an exponential light profile must rise steeply with radius to keep the rotation curve flat; the same conclusion was reached by Bosma & van der Kruit (1979) and Petrou (1981). Application to NGC 3198, which has a flat rotation curve out to at least 11 $h$, shows that the local $M/L$ value at the boundary would have to be about 1400 times larger than the $M/L$ ratio in the inner region, or $M/L_B (11 \ h) \approx 6000 \ M_\odot /L_\odot$. This result would exclude the possibility that the material far out in the disc consists of low mass hydrogen-burning stars.

(b) A dark halo

For a spherical halo the amount of dark matter inside a given radius, needed to keep the rotation curve flat, is given by:

$$M_{\text{dark}}(r)/M_{\text{lum}} = 0.39 \ r/h - 1 \ (r > 3 \ h). \tag{6}$$

Here we have used the point-mass approximation to calculate the circular velocity of the exponential disc at large radii (maximum error ca. 8% beyond 3 $h$). Thus, for galaxies like NGC 2403 and 3198 there is at least three times as much dark matter as luminous matter inside the HI radius. Note that this result is independent of the amplitude of the rotation curve. All that one is doing is comparing observed and expected shapes of rotation curves. For flat rotation curves the maximum-disc hypothesis will therefore, in first approximation, produce a ratio $M_{\text{dark}}/M_{\text{lum}}$ (at fixed $r/h$) that is independent of galaxy luminosity (cf. Bahcall & Casertano 1985).

It is also easy to relate halo and disc densities. If, for convenience, we model the halo with an isothermal sphere with velocity dispersion $\sigma_v$ and core radius $r_c$ we have:

$$\sigma_v = v_{\text{max}} / \sqrt{2} \tag{7}$$

and

$$r_c = 4.7 \ h; \tag{8}$$

$r_c$ is the radius at which the density of the halo is $2^{-1}$ times its central density. The constant 4.7 in (8) has been determined from the fit for NGC 3198. For the central density of the halo we then find:

$$\rho_{\text{halo}}(0) = 8.73 \ \sigma_v^2 / (4\pi G r_c^2) = 0.0061 \ M_{\text{disc}}/h^3. \tag{9}$$

For a family of discs with constant central mass density $M_{\text{disc}} \propto h^2$, thus, the central density of the halo is expected to be inversely proportional to the scalelength of the disc. Values of $\rho_{\text{halo}}(0)$ calculated with (9) agree to within 10–20% with those given by Carignan & Freeman (1985) for their low luminosity galaxies.

Bahcall & Casertano (1985), hereafter B.C., determine disc and halo densities at the optical radius $R_{\text{out}}$ (i.e. the isophotal radius corresponding to 26.6 mag arcsec$^{-2}$, B-band). This radius corresponds typically to 4.6 $h$ (mean central surface brightness 21.5 mag arcsec$^{-2}$, B-band), which is just the core radius of the isothermal halo. Thus one expects:

$$\rho_{\text{halo}}(R_{\text{out}}) \approx 0.0022 \ M_{\text{disc}}/h^3. \tag{10}$$
**DARK MATTER IN SPIRAL GALAXIES**

In the mean we find for the 8 galaxies in the B.C. sample: 0.0017 ± 0.0009 (1σ) $M_\odot$/pc$^3$, in good agreement with their value of 0.0015 $M_\odot$/pc$^3$. (Individual values differ typically by a factor of two.) Evaluating the disc density (in the plane $z = 0$) at $R_{\text{out}}$ in the same way as B.C. gives:

$$\rho_{\text{disc}}(R_{\text{out}}) = 0.0040 \, M_{\odot}/h^3$$  \hspace{1cm} (11)

Thus,

$$\rho_{\text{disc}}/\rho_{\text{halo}} = 0.5,$$

also in rather good agreement with B.C.'s value of 0.3. Bahcall & Casertano note that somewhat different assumptions have been made in the mass modelling procedures for the galaxies in their paper. It seems safe to conclude however that the regularities found by them are essentially equivalent to the use of the maximum-disc hypothesis.

5. **Disc–halo coupling in self-consistent halo models**

The mass models discussed in §2 give static decompositions of the rotation curve into the contributions by luminous and dark matter. They do not provide information on questions like: what would the halo look like without a disc; is the halo substantially flattened by the disc and is there a coupling between disc and halo due to the gravitational interaction of the two systems? The latter question in particular is important because a self-consistent treatment of the interaction between disc and halo might shed some light on the disc–halo conspiracy.

To study these questions, one of us (T.v.A.) has performed $N$-body simulations in which the halo is modelled by a collection (5000) of interacting particles which are exposed to the gravitational field of an embedded disc (for details on method see van Albada 1982). A truncated isothermal sphere, with a fixed boundary at 200 kpc, is used to represent the dark halo. The boundary of the sphere acts like a reflecting surface, i.e. a particle leaving the system through the boundary surface is replaced by another particle entering the system. Such a configuration is in dynamical equilibrium. In this halo we place a rigid exponential disc with fixed scalelength $h$ and an exponential density distribution in $z$, with scalelength $z_0 = 0.2 \, h$.

Initially the mass of the disc is zero, then the disc mass is allowed to grow on a time scale of $10^9$ a to a maximum value; thereafter the evolution of the system is followed until it again reaches an equilibrium.

First the case of NGC 3198 has been modelled. The two parameters characterizing the halo, initial core radius and velocity dispersion (or mass inside the boundary), have been varied until a good fit to the observed rotation curve is obtained. Results are shown in figure 6. The best fit is obtained for $r_c = 27$ kpc (dimensionless truncation radius $\xi_1 = 22$; for notation see Chandrasekhar (1939)) and $\sigma_v = 85 \, \text{km s}^{-1}$ (figure 6a). The density distribution in the inner part of the halo is altered substantially due to the presence of the disc; the final core radius ($r_c \approx 12$ kpc) is about 40% of the initial core radius. The halo is not perceptibly flattened by the disc. If the initial core radius is chosen 10% smaller ($\xi_1 = 24$) a smaller value of $\sigma_v$ is needed to fit the observed amplitude of the rotation curve at 8 kpc (figure 6b). The halo mass inside 200 kpc is also smaller and the halo is more responsive to the disc. Note that the rotation curve in this case is clearly declining. A rotation curve with a depression at 15 kpc is found for a larger halo core radius (figure 6c, $\xi_1 = 20$, $\sigma_v = 100 \, \text{km s}^{-1}$). These results show that the halo parameters are fixed to within narrow limits, $r_c = 27 \pm 2$ kpc and $\sigma_v = 85 \pm 5 \, \text{km s}^{-1}$ (1σ), to get a good fit to the observed rotation curve. In other words, for the disc of NGC 3198 there is a uniquely defined unperturbed halo.
Next we have chosen the best fitting halo for NGC 3198 in figure 6a as a standard halo, and discs with different masses and scalelengths have been embedded in this halo. All discs belong to a family with constant central surface density, i.e. $M_{\text{disc}}/h^2$ is constant. Results are shown in figure 7. The best fit for NGC 3198 is repeated in figure 7a. Already for a 20% more massive disc (figure 7b), the final rotation curve of the combined system is clearly declining and inconsistent with the shape of observed rotation curves. A decrease in disc mass is somewhat

<table>
<thead>
<tr>
<th></th>
<th>$M_{\text{disc}}$</th>
<th>$h$</th>
<th>$M_{\text{halo}}$</th>
<th>$\xi_1$</th>
<th>$\sigma_v$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a</td>
<td>0.35</td>
<td>0.268</td>
<td>7.5</td>
<td>22.0</td>
<td>85.3</td>
</tr>
<tr>
<td>6b</td>
<td>0.35</td>
<td>0.268</td>
<td>5.5</td>
<td>24.0</td>
<td>73.7</td>
</tr>
<tr>
<td>6c</td>
<td>0.35</td>
<td>0.268</td>
<td>10.5</td>
<td>20.0</td>
<td>99.9</td>
</tr>
<tr>
<td>7a</td>
<td>0.35</td>
<td>0.268</td>
<td>7.5</td>
<td>22.0</td>
<td>85.3</td>
</tr>
<tr>
<td>7b</td>
<td>0.42</td>
<td>0.293</td>
<td>7.5</td>
<td>22.0</td>
<td>85.3</td>
</tr>
<tr>
<td>7c</td>
<td>0.09</td>
<td>0.134</td>
<td>7.5</td>
<td>22.0</td>
<td>85.3</td>
</tr>
</tbody>
</table>

**Figure 6** (left). Mass models for NGC 3198 with self-consistent halo. Dots with error bars give the observed rotation curve and diamonds ($\bigodot$) the rotation curve for the unperturbed halo. Solid lines give the rotation curves for disc, perturbed halo and for these two combined. The three panels show the same disc model with a best fitting halo with core radius $r_c = 27$ kpc and $\sigma_v = 85$ km s$^{-1}$ (a), and with halos with 10% smaller (b) and larger (c) core radius. Masses are expressed in $10^{11} M_\odot$.

**Figure 7** (right). Mass models for a family of discs embedded in the best fitting halo for NGC 3198 (see figure 6a). Diamonds ($\bigodot$) indicate the rotation curve for the unperturbed halo and solid lines the rotation curves for the disc, the perturbed halo and these two combined. Panel (a) gives the best fit for NGC 3198, as in figure 6a. In (b) the disc is 20% more massive and in (c) a factor four less massive than in (a). Disc scalelength, $h$, is expressed in units of 10 kpc, and mass in $10^{11} M_\odot$. 

[ 28 ]
more promising. For the case shown in figure 7c, the disc mass has been reduced by a factor of four. In first approximation the combined rotation curve is rather flat out to 12 disc scalelengths, suggesting that an observed flat rotation curve could be a shoulder on a rising curve for the halo. However, the dip at 10 kpc in figure 7c is a clear signature of the presence of two separate components, and such a feature is not present in observed rotation curves (see also figure 6c).

We conclude from these experiments that the gravitational interaction of disc and halo provides hardly any relief for the problem that disc and halo seem to conspire to produce an approximately flat rotation curve.

6. Shape of dark halos

Measurement of the shape of dark halos is important because it may tell us about their origin. Several methods have been proposed but results are rather inconclusive.

The strongest constraint, at present, on the z-distribution of dark matter is provided by a measurement of the half-thickness of the H I layer in edge-on galaxies as a function of radius. Observations show a strong increase of z with radius. Information on the vertical velocity dispersion of H I is needed to derive the character of the force field perpendicular to the plane from the run of z; this quantity must be taken from observations of nearly face-on galaxies. Fortunately the variation of \( \langle v_z^2 \rangle \) within a galaxy, and between galaxies, is small (van der Kruit & Shostak 1984). For his analysis of NGC 891, van der Kruit (1981) uses a value of 10 km s\(^{-1}\), and finds that a model with all matter in the disc with vertical scaleheight 1.0 kpc is not consistent with the data (predicted flaring too small), whereas a model with the luminous matter in a disc and the dark matter in a spherical halo agrees well with the observations. It is not clear whether a flaring disc containing both luminous and dark matter is also at variance with the data, or, to what extent one can constrain the axial ratio c/a of a spheroidal dark halo.

Monet et al. (1981) have proposed to use the flattening of bulges in external galaxies and the population II component in our galaxy as a constraint on the flattening of the dark halo. If the velocity distribution of population II stars depends on energy alone (a rather implausible assumption, but the anisotropy can in principle be measured in our galaxy), they will fill equipotential surfaces. Thus, the flattening of the population II component may give an indication of the axial ratio of the equipotentials of the total gravitational field. Equipotentials are quite round however, even for flat discs, and the flattening of the stellar distribution must be measured with high precision to provide interesting limits. For our galaxy Monet et al. find that no more than half the mass within the solar circle lies in a disc. Separation of bulge and disc components is necessary to translate this into a result on the distribution of dark matter.

An interesting method proposed by Schweizer et al. (1983) is the comparison of the velocities in the disc and polar ring of spindle galaxies. Good data are available for A0136–0801. This method yields in essence also a measurement of the flattening of the equipotential surface of the gravitational field. Schweizer et al. conclude that the dark halo is 'more nearly spherical than flat'.

The main weakness of these methods is that assumptions must be made that are hard to verify. Moreover, highly accurate data are needed before one can discriminate between different possibilities. Detailed models for many disc–halo combinations, exploring also a range
of assumptions, are needed before quantitative statements on the flattening of the halo can be made.

Attempts to find information on the shape of dark halos have also been made by using theoretical arguments on the persistence of warped HI discs. Warps will slowly dissolve through differential precession of the orbits of the gas clouds. From a test particle calculation for NGC 5907, Tubbs & Sanders (1979) find that the persistence of the warp for $5 \times 10^9$ years requires a massive spheroidal halo with an axial ratio not exceeding 1.5. On the other hand, Petrou (1980) finds that a spheroidal halo for which the flattening lies beyond the region of the warp helps to reduce differential precession; see also Sparke (1984).

7. Summary

We have studied the discrepancy between dynamical and luminous mass in spiral galaxies (in the context of Newtonian gravity). The amount and distribution of dark matter are derived on the assumption that the mass:light ratio of luminous matter is constant (separately for bulge and disc) and equal to the maximum value allowed by the rotation curve. This 'maximum-disc' hypothesis is supported by several observational and theoretical arguments. The main conclusions are:

1. There is clear evidence for dark matter in the outer regions of galaxies. The most convincing cases are those of six spiral galaxies with exponential light profiles and flat rotation curves measured out to about 10 disc scalelengths (ca. three times the optical radius $R_{25}$).

2. The amount of dark matter inside 10 disc scalelengths (20–40 kpc) is three to five times the amount of luminous matter.

3. In the inner region, i.e. inside ca. 2.5 disc scalelengths, luminous matter dominates.

4. The distributions of luminous and dark matter are closely coupled: bulge, disc and halo ‘conspire’ to produce flat rotation curves.

5. The ratio $M_{\text{dark}}/M_{\text{lum}}$ inside the optical radius, derived with the maximum-disc hypothesis, appears to be independent of galaxy luminosity.

6. Self-consistent halo models show that the large-scale gravitational interaction of disc and dark halo does not alleviate the problem of disc–halo coupling: for a given (maximum) disc there is a unique halo.

7. No convincing information is available yet on the shape of the distribution of dark matter. It is not possible to discriminate between a spherical or flattened halo, or a flaring disc, but a flat disc seems unlikely.

Many colleagues contributed to this paper through fruitful discussions and comments, in particular: J. N. Bahcall, S. Casertano, S. M. Kent, R. H. Sanders and J. A. Sellwood. We are grateful to S. M. Kent for letting us use his photometry of UGC 2885, and for sending us inspiring information on his work on dark matter in spiral galaxies. We also thank K. Begeman for the use of his rotation curve data before publication and assistance with the preparation of figures, S. Casertano for the use of his software, and J. Sellwood for providing us with the results of the stability analysis for disc–halo models for NGC 3198.

Note added in proof (1 September 1986). In Sellwood’s calculations described in §3d it has been assumed that the $Q$-value for the disc is equal to 1.5. Different $Q$-values of this order do not
affect the result. In galactic discs with an effective $Q$-value close to one, two-armed spiral structure could be present for much lower disc densities, and the conclusion reached would not be valid. The possible existence of such cold discs is a matter of dispute however. We thank G. Bertin, C. C. Lin and A. Toomre for correspondence and discussions on this point.

References

Begeman, K. 1986 (In preparation.)
Discussion

MARIA PETROU (Department of Theoretical Physics, 1 Keble Road, Oxford, OX1 3NP). I have two comments to make. (1) If one plots the local mass:light ratio against the local light, one will find that all spiral galaxies follow the same curve, irrespective of their type. This is an indication that there is a strong correlation between dark and luminous matter (Petrou 1981). (2) Professor van Albada's last point was that one can stabilize warped discs by spherical halos. Actually, one can stabilize the warps much better with oblate spheroidal halos, provided that the flattening of the halo is outside the disc (Petrou 1980).

T. S. VAN ALBADA. I thank Maria Petrou for pointing out the omission with regard to the stabilization of warps. It has been remedied in the published paper. As regards the first point, we are in agreement there (cf. §4a).

R. J. TAYLER (Astronomy Centre, University of Sussex). Can I just raise a small point of clarification? Professor van Albada referred to the inner part of the rotation curve being determined by the luminous matter and to the outer part being flat because of the hidden matter. Of course the mass in the inner regions includes the Oort limit hidden mass.

T. S. VAN ALBADA. That is correct. The analysis presented here only tests the constancy of $M/L$ with radius; it does not specify the nature of the objects in the disc.