THE ASTROPHYSICAL JOURNAL, 641:L33–L36, 2006 April 10

A MAGELLANIC ORIGIN FOR THE WARP OF THE GALAXY

MARTIN D. WEINBERG[1] AND LEO BLITZ[2]

Received 2005 November 7; accepted 2006 February 22; published 2006 March 17

ABSTRACT

We show that a Magellanic Cloud origin for the warp of the Milky Way can explain most quantitative features of the outer H I layer recently identified by Levine et al. We construct a model, similar to that presented by Weinberg in 1998, that produces distortions in the dark matter halo, and we calculate the combined effect of these dark halo distortions and the direct tidal forcing by the Magellanic Clouds on the disk warp in the linear regime. The interaction of the dark matter halo with the disk and resonances between the orbit of the Clouds and the disk account for the large amplitudes observed for the vertical $m = 0$, 1, and 2 harmonics. The observations lead to six constraints on warp-forcing mechanisms, and our model reasonably approximates all six. The disk is shown to be very dynamic, constantly changing its shape as the Clouds proceed along their orbit. We discuss the challenges to MOND posed by the observations.

Subject headings: Galaxy: disk — Galaxy: kinematics and dynamics — Galaxy: structure

Online material: color figures

1. INTRODUCTION

The warp of the outer Milky Way, known since 1957 (Kerr 1957), was quantitatively determined for the first time by Levine et al. (2006). It can be described as a superposition of three and only three of the lowest-order vertical harmonics of a disk: a *dish-shaped* $m = 0$, an *integral-sign–shaped* $m = 1$, and a *saddle-shaped* $m = 2$ harmonic. The lines of nodes for each are close to coincident and nearly radial. The amplitude of each reaches 7%–10% of the radius of the disk. A number of possible warp-producing mechanisms have been suggested, including long-lived eigenmodes, forcing by halo triaxiality, persistent accretion of cold gas, and tidal excitation. We show here that the origin of this warp can be well described as the tidal interaction of the Magellanic Clouds (MCs) with the disk and dark matter halo of the Milky Way. The interaction of the dark matter halo with the disk and resonances between the orbit of the MCs and the disk account for the large amplitudes of the three harmonics and their approximate shape and orientation.

2. OBSERVATIONS

Levine et al. (2006) found that a dynamical model for the warp must satisfy six observational constraints: (1) the three lowest-order harmonics, $m = 0$, 1, and 2, are necessary and sufficient to describe the global shape of the warp, with higher order global harmonics typically an order of magnitude or more weaker; (2) the $m = 1$ warp has the largest amplitude everywhere in the outer disk; (3) the $m = 0$ and $m = 2$ warps are comparable in amplitude to the $m = 1$ warp but are smaller at all radii; (4) the $m = 1$ warp has a measurable amplitude at the Galactocentric radius of the Sun, R_0, but the $m = 0$ and $m = 2$ warps begin near the edge of the stellar disk, at $R = 2R_0$; (5) all three harmonics grow approximately linearly with radius, reaching amplitudes of 1–2 kpc at about $R = 30$ kpc; and (6) the amplitude of each of the harmonics reaches 5%–10% of the radius of the disk. The lines of maximum descent of the $m = 1$ and $m = 2$ warps are coincident within about 12° in Galactocentric azimuth ϕ and show little evidence of precession. The lines are located near $\phi = 90°$.

3. METHODOLOGY

We use the procedure described in Weinberg (1998, hereafter W98) to couple the halo response to the tidal excitation theory presented by Hunter & Toomre (1969, hereafter HT). This assumes that the disk remains thin and that gas dissipation is unimportant for the dynamics. HT concluded that the direct excitation of the disk by the MCs produces a warp of only a few hundred parsecs, an order of magnitude less than the amplitude of the observed warp. W98 wedded the halo excitation presented in Weinberg (1989) to the HT approach, allowing the disk to feel both the tidal field from the Clouds directly and the force from the dark matter halo wake excited by the Clouds. The assumption of linearity limits the predictions to modest amplitudes.

We use perturbation theory rather than *N*-body simulation because of the intrinsic difficulty and subtlety in obtaining accurate multiple-timescale results from a simulation. The excitation hierarchy of satellite orbit → halo wake → disk bending modes results in multiple interleaved timescales: the orbital periods of the MCs, the pattern speeds of the halo wakes, and the pattern speeds of the bending modes. We also have multiple spatial scales, from the MC orbital radius to the disk scale height. In addition, particle simulations can degrade the resonant dynamics as described by Weinberg & Katz (2005). With such difficulties, one should verify the dynamics of each component of the mechanism before putting everything together in an *N*-body simulation, although this is rarely done.

4. MODEL

The perturbation theory includes the force from an extended satellite. However, the halo and disk excitation depend only on the lowest-order harmonics, while the spatial extent changes the higher order harmonics, and thus the extent plays little role. Therefore, as long as the MCs remain bound, their masses may be added for our estimates. We use the disk profile from W98. Our halo is an NFW profile with $c = 15$ (Navarro et al. 1997) and virial mass 20 times the disk mass. The orbital plane was computed as described in W98. We used the radial velocity

[1] Department of Astronomy, University of Massachusetts Amherst, 710 North Pleasant Street, Amherst, MA 01003; weinberg@astro.umass.edu.

[2] Department of Astronomy, University of California, Berkeley, 601 Campbell Hall, Berkeley, CA 94720; blitz@berkeley.edu.

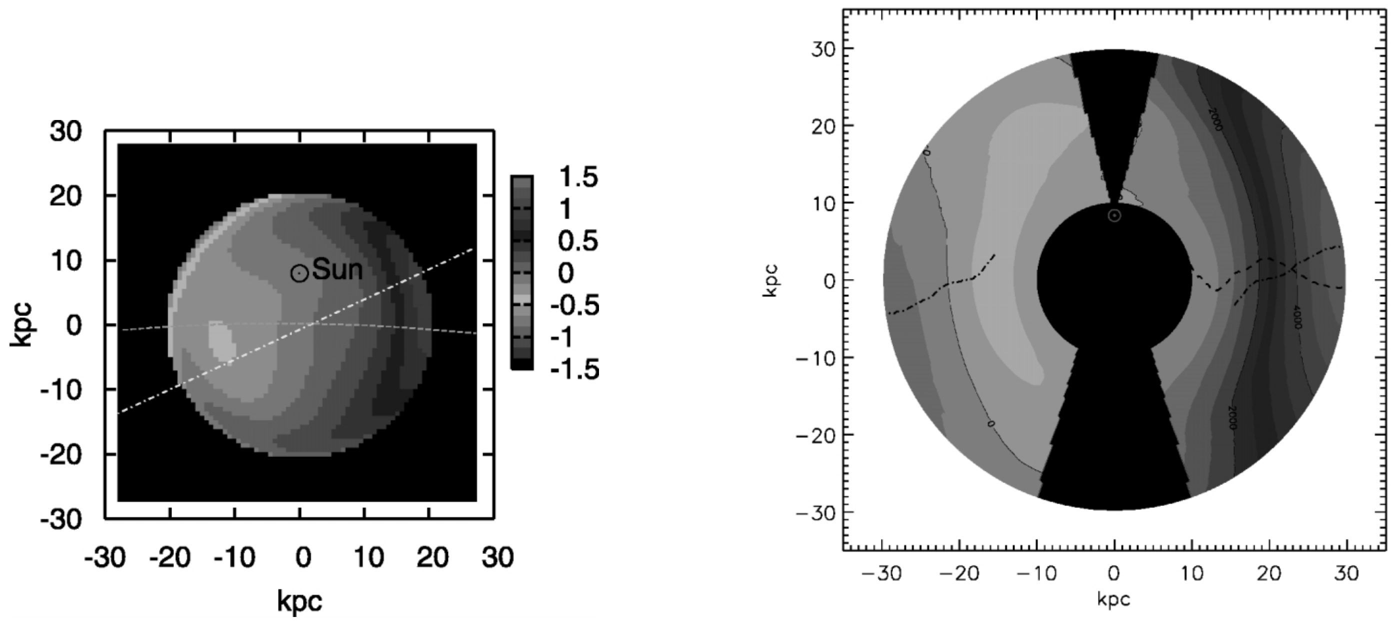

FIG. 1.—Lines of maximum descent (perpendicular to the line of nodes) for $m = 1$ and $m = 2$ from the model (*dashed and dot-dashed lines, respectively*). The lines are superposed on a contour plot of the deviation in z of the midplane of the disk from $b = 0°$. Compare with Fig. 11 in Levine et al. (2006). [*See the electronic edition of the Journal for a color version of this figure.*]

and proper motion from Kallivayalil et al. (2006) and the distance modulus from Freedman et al. (2001) to derive a space velocity, and we computed the resulting orbit in the spherical dark matter halo. We adopted the LMC mass from Westerlund (1997) of $2 \times 10^{10}\ M_{\odot}$. The current position of the LMC is shown, as is the current state of the warp. This orbit will carry it toward the north Galactic pole.

Although our calculation assumes a collisionless and vertically thin medium, it is more generally applicable for the following reasons: First, although the dispersion relations for a multicomponent and collisionless media differ at small scales (see, e.g., Jog 1996; Rafikov 2001), the large-scale warp will be governed by inertia and the gravitational restoring force, not by local pressure. In addition, the disk self-gravity is dominated by the inner stellar disk; the outer gas layer plays only a minor role in establishing the modes. Therefore, our collisionless results are likely to be similar to a multicomponent calculation. Secondly, the vertical restoring force depends very weakly on the thickening as long as the vertical degree of freedom does not couple to the bending (this may be demonstrated by straightforward but tedious algebra). Therefore, the modes will be largely unchanged by thickening over short timescales. However, the challenging problem of vertical coupling is important and remains to be investigated thoroughly.

The warp is a very dynamic structure, based on the temporal evolution of the model.[3] Rather than a static structure that might be expected for a warp in response to a triaxial halo, a warp that results from the MCs is continuously changing shape because of the varying amplitudes and phases of the various modes. The image looks rather like a flag flapping in the breeze as the MCs complete an orbit of the Milky Way.

[3] This can be seen in the AVI file of the simulations that can be found at http://www.astro.umass.edu/~weinberg/lmc. Also included at this site are comparisons of the $m = 0$, 1, and 2 evolution.

5. COMPARISON WITH OBSERVATIONS

Figure 1 shows the shape of the warp for both the model and the data. We plot only $m = 0$, 1, and 2 from the data, ignoring the weak but significant $m = 10$ and $m = 15$ terms. The overall agreement is quite good, although there are some differences. The model does not, for example, have a minimum that is as extensive in Galactic azimuth as the data. Lines of maximum descent for $m = 1$ and $m = 2$ in both the simulations and the H I data analysis are also shown in the figure. They are separated by about 20° in the model, a bit more than but close to the 12° in Levine et al. (2006). Both sets of lines are oriented in the same sense: close to $\phi = 90°$. Neither set of lines shows evidence of significant variation with Galactocentric radius.

Figure 2 shows the amplitudes of the first three harmonics, $m = 0$, 1, and 2, in both the model and the data. We increased the MC mass by 33% for a better fit. As in the data, the $m = 1$ in the simulations is the strongest and increases nearly linearly out to the edge of the disk. At large R, where the inertia is small, the linear theory is expected to overpredict the warp (Tsuchiya 2002). The simulations also show a weak response of $m = 0$ and $m = 2$ out to about 15 kpc, which then increases nearly linearly, but with approximately the same amplitude for both, providing a reasonable representation of the data.

The amplitude ordering of these vertical harmonics has a natural physical explanation. The force from the halo wake and the satellite may described by three-dimensional harmonics, and these affect the warp height as follows: A spherically symmetric $m = 0$ halo distortion does nothing to the warp. An $m = 1$ halo distortion will tend to accelerate the disk in the vertical direction; the differential acceleration of the disk results in a vertical $m = 0$ “dishing.” An $m = 2$ halo distortion will attract the disk upward and downward in a reflection-symmetric way, causing the classic integral-sign warp. Higher order symmetries may be deduced from similar geometric considerations. The power in the halo excitation drops off as an inverse power

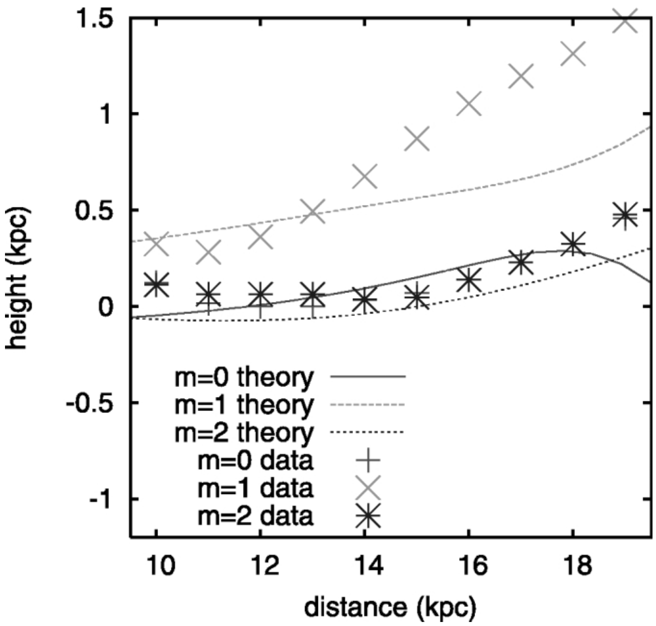

FIG. 2.—Amplitudes of the $m = 0$, 1, and 2 harmonics as a function of Galactic radius from the simulations (*curves*) and from the H I Milky Way data (*points*; Levine et al. 2006). [*See the electronic edition of the Journal for a color version of this figure.*]

of the harmonic order, and only the lowest-order terms have features well inside the satellite orbit. Conversely, the existence of these higher order harmonics with a power-law dropoff is a natural consequence of this tidal theory and is consistent with the data.

We fixed the disk and the halo mass inside of the virial radius while adjusting the satellite orbit and halo concentration and found the following trends: First, the halo wake and its pattern speed are determined by the halo concentration. The disk bending modes have a natural set of frequencies for a given halo. These will be maximally excited when forced by the halo wake at or near harmonics of this natural frequency. The ratio of $m = 2$ amplitude to $m = 1$ amplitude is maximized for a halo concentration $c \approx 10$. For an NFW profile, $\rho \propto r^{-1}(r + r_s)^{-2}$, this yields $r_s \approx 30$ kpc. This is very close to the ΛCDM estimates for the Milky Way concentration. Secondly, the orientation of the response depends on whether the nearest resonance is larger or smaller than the natural frequency. Therefore, changing the satellite orbit, which changes the forcing frequency, affects both the amplitude and the orientation of the warp response. In short, the warp depends on a “clockwork” of frequency relationships that depends on the satellite orbit, the dark matter halo, and the disk. A pericenter larger than the current 49 kpc estimate shifts the position angle so that the warp peaks closer to P.A. = 180°. Similarly, a smaller pericenter increases the amplitude and also shifts the position angle. We conclude that our model “prefers” our current fiducial MC model. We are not claiming that our fiducial model is the most probable among the distributions of allowed values, but that a plausible choice of parameters corresponds to many features of the observed data.

6. DISCUSSION

6.1. *Comparison with N-Body Simulations*

The predictions of W98 were checked by several groups using *N*-body simulations. García-Ruiz et al. (2002) used a hybrid *N*-body particle-ring code and did not find the warp predicted in W98. They offer the failure of the linear theory as the culprit but did not investigate a variety of models. Similarly, Mastropietro et al. (2005) performed a simulation including both the gaseous stellar and that dark components of the LMC and the Milky Way, and they remark that the effect on the Milky Way is negligible. Conversely, Tsuchiya (2002), using a hybrid code that includes both a potential expansion and a tree code, obtained amplitudes of $m = 0$, 1, and 2 warps in the larger halo model that show very good agreement with the observed amplitudes in the Milky Way. He also shows that the amplitude of the warp depends strongly on the dark halo mass model.

It is difficult to reconcile these contradictory findings, but three possibilities obviously occur: (1) the linear theory does not apply; (2) the simulations do not apply because of numerical difficulties; and (3) the two simulations with null warps have chosen unlucky sets of parameters. Both García-Ruiz et al. (2002) and Tsuchiya (2002) chose methods that explicitly treat the multiple scales. García-Ruiz et al. (2002) used tilted rings to represent the disk and a tree code to represent the halo. Although the use of rings limits the investigation to the $m = 1$ response, it should be sufficiently sensitive to the halo excitation without strong particle-number issues. Tsuchiya (2002) used an expansion algorithm to represent the halo gravitational field and a tree code to represent the disk, to better represent the multiple scales. We feel that the discrepancy between the results of these two groups is most like item 3: the García-Ruiz et al. model is not particularly warp-producing. The good qualitative correspondence between W98 and Tsuchiya (2002) suggests that the linear theory captures the underlying physics, although it is likely to differ in detail. For example, the linear theory overpredicts the height in the edge of the disk (Fig. 2), and this motivates our truncation of the predictions in the outer disk. Mastropietro et al. (2005) emphasize the effect of the Milky Way tides on the LMC and do not tailor their approach to treat multiple scales. We suggest that their report of no disk warp results from items 2 and 3.

6.2. *Other Explanations*

Most warp theories depend on the bending response of the disk, and this response is strongly affected by the existence of a self-gravitating dark matter halo. The theories may be roughly grouped as follows: (1) bending modes may be persistent (e.g., Sparke & Casertano 1988); (2) the disk is responding to the nonaxisymmetric shape of the halo; (3) warps are tidally excited (as we have discussed here); and (4) the warp may be produced by the response of the disk to cosmic infall (e.g., Jiang & Binney 1999). This topic was nicely reviewed by Binney (1992). Subsequently, Nelson & Tremaine (1995) argued against long-lived modes. The triaxiality of the Milky Way halo remains uncertain. Helmi (2004) reports a prolate halo ($q = 1.25$), while Johnston et al. (2005) finds an oblate halo ($q = 0.8$–0.9) using Sgr dwarf constraints. Such modest triaxiality seems unlikely to produce the observed vertical $m = 2$ feature. In addition, cosmic infall more naturally produces tilted rings, an $m = 1$ feature; the observed vertical $m = 2$ may require a conspiracy of several inflow directions.

6.3. *Relevance for Modified Gravity*

Although the success of our model favors the existence of a dark matter halo in nature, many find it seductive to modify gravity to produce the observed rotation velocities in the Galaxy without dark matter (modified Newtonian dynamics, or MOND; Milgrom 1983a, 1983b; Bekenstein & Milgrom 1984). Might the tidal theory also apply in MOND? It is beyond the scope of this Letter to repeat our calculations using MOND, but it seems plausible that direct forcing of the Galactic disk by the MCs in MOND may provide warp amplitudes in excess of the original predictions without a dark halo (HT). Similarly, we would expect the disk modes and frequencies to be qualitatively similar, the excess restoring force of the halo being produced by the MOND force. However, (1) MOND would have to admit bending modes with similar morphology to those in the Newtonian theory, and (2) these modes would have to conspire to frequencies that couple them to the direct forcing by the MCs in such a way that they assumed the same orientation as in the "clockwork" described in § 5. Because this clockwork and the $m = 2$ to $m = 1$ amplitude ratio depends on the simultaneous halo and disk excitation, agreement with the observations described in § 2 seems rather unlikely and thereby disfavors MOND. We encourage detailed predictions.

6.4. *Relevance for Other Galactic Systems*

A large fraction of other warped galaxies also show warps in their H I layers, and a significant fraction of these are asymmetric. Warps in these systems are generally analyzed with a program such as ROTCUR (Begeman 1989) or one of its derivatives but invariably are forced to fit the $m = 1$ warp only. This Letter and the work of Levine et al. (2006) show that at least three harmonics ought to be fitted to the warps of external galaxies, especially those with asymmetric warps, which can be caused by a superposition of these harmonics. The degree to which various harmonics are present in a warp can produce important constraints on whether the warp is due to a satellite, a triaxial halo, cold gas inflow, or some primordial excitation. All warps may not be alike.

6.5. *Summary*

We have demonstrated a plausible mechanism for the excitation of the warp that explains all of its general features. It is possible that different mechanisms may act in different galaxies and possibly in concert in a single galaxy. Nonetheless, our simple model nicely reproduces the observed features in the Milky Way H I gas layer. The existence of massive companions, the Magellanic Clouds, and the prediction from linear perturbation theory and at least one corroborating *N*-body simulation suggest that a tidal explanation is viable. Our model depends on the gravitational response of the halo and thereby suggests that the dark matter is not an artifact of modifying the laws of gravity. Conversely, given a dark halo, we argue that the tide from the MCs must be affecting the Milky Way disk, and given the quality of the agreement, it seems to be the dominant mechanism. This model then promises an additional constraint on the distribution of dark matter. Although we have emphasized the gas-layer response beyond the stellar disk, the effect of other satellite encounters such as the recent accretion of the Sagittarius dwarf may be detectable in future high-resolution surveys and may help determine the properties of the inner halo. Warp observations are important because they promise to reveal aspects of the dark matter distribution that are otherwise observationally inaccessible. Further analysis of warps may provide more precise constraints on the profile of dark matter in the Milky Way and nearby external galaxies.

L. B. would like to acknowledge partial support from NSF grant 02-28963. M. D. W. has been partially supported by NASA grant NAG 5-12038 and NSF grant 02-05969. We thank Arend Sluis and the anonymous referee for comments on the manuscript.

REFERENCES

Begeman, K. G. 1989, A&A, 223, 47
Bekenstein, J., & Milgrom, M. 1984, ApJ, 286, 7
Binney, J. 1992, ARA&A, 30, 51
Freedman, W. L., et al. 2001, ApJ, 553, 47
García-Ruiz, I., Kuijken, K., & Dubinski, J. 2002, MNRAS, 337, 459
Helmi, A. 2004, MNRAS, 351, 643
Hunter, C., & Toomre, A. 1969, ApJ, 155, 747 (HT)
Jiang, I.-G., & Binney, J. 1999, MNRAS, 303, L7
Jog, C. J. 1996, MNRAS, 278, 209
Johnston, K. V., Law, D. R., & Majewski, S. R. 2005, ApJ, 619, 800
Kallivayalil, N., van der Marel, R. P., Alcock, C., Axelrod, T., Cook, K. H., Drake, A. J., & Geha, M. 2006, ApJ, 638, 772
Kerr, F. J., Hindman, J. V., & Carpenter, M. S. 1957, Nature, 180, 677
Levine, E. S., Blitz, L., & Heiles, C. 2006, ApJ, in press (astro-ph/0601697)
Mastropietro, C., Moore, B., Mayer, L., Wadsley, J., & Stadel, J. 2005, MNRAS, 363, 509
Milgrom, M. 1983a, ApJ, 270, 365
———. 1983b, ApJ, 270, 371
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, ApJ, 490, 493
Nelson, R. W., & Tremaine, S. 1995, MNRAS, 275, 897
Rafikov, R. R. 2001, MNRAS, 323, 445
Sparke, L. S., & Casertano, S. 1988, MNRAS, 234, 873
Tsuchiya, T. 2002, NewA, 7, 293
Weinberg, M. D. 1989, MNRAS, 239, 549
———. 1998, MNRAS, 299, 499 (W98)
Weinberg, M. D., & Katz, N. 2005, MNRAS, submitted (astro-ph/0508166)
Westerlund, B. E. 1997, The Magellanic Clouds (Cambridge: Cambridge Univ. Press)