1. The formation of the Milky Way in the CDM paradigm

K. C. FREEMAN

1.1 Introduction

What does our Galaxy look like? We can compare the COBE image of our Galaxy, taken in the near-IR, with the visible image of the edge on spiral NGC 891. Our Galaxy would probably look much like NGC 891 if it were observed in visible light from far away (see Figure 1.1). The Milky Way is very clearly a disk galaxy: its disk is the primary component and is supported almost entirely by its rapid rotation. We also see a small central bulge which contributes about 20% of the total light. Some galaxies have much larger bulges. The small bulge of the Milky Way is a pointer to the events that occurred as it formed and evolved. We would like to understand how our Galaxy came to look like this.

Figure 1.2 shows schematically the five main components of the stellar galaxy. The thin disk and bulge are the main visible components. The thin disk is enveloped in a thicker thick disk which contributes only about 10% of the light of the disk. These thick disks are very common and their formation appears to be part of the formation process of disk galaxies. The stellar halo provides only about 1−2% of the total light but is very important for understanding how the Galaxy was assembled. The stars of the halo are metal-poor, mostly with abundances of $[Fe/H] < -1$. Unlike the disks and the bulge, the stellar halo is not rotating significantly: it is supported against gravity by the random motions of its stars. Currently we believe that the halo represents the debris of small metal-poor galaxies that were accreted by the Galaxy during its formation and evolution. Finally there is the dark halo. It appears to contribute at least 95% of the total mass of the Galaxy. Current opinion is that the dark halo does not contribute much to the gravitational field in the inner few kpc of the Galaxy, but it rapidly becomes the dominant contributor at larger radii. The dark halo appears to extend to a radius of at least 150 kpc.

Each of these components has something to tell us about the formation history of the Galaxy. Our task is to understand how the formation and evolution of the Milky Way took place and to evaluate how the Galaxy compares with the predictions of CDM simulations.

The thin disk is relatively metal-rich and its stars cover a wide range of ages. The other stellar components are all relatively old and more metal-poor. Figure 1.3 summarizes our current belief about the age-metallicity relation for the components of the Galaxy. The similarity of the [Fe/H] range for the thick disk and the globular clusters is worth noting.

The total mass of the Galaxy is about 2×10^{12}. The stellar mass in the bulge is about 1.2×10^{10} $M_\odot$, the disk is about 5×10^{10} $M_\odot$, and the stellar halo only about 1×10^9 $M_\odot$. The halo and its globular clusters have ages of about 10–12 Gyr and the thick disk stars appear to be older than 10 Gyr. Star formation in the thin disk started about 10 Gyr ago and has continued at a more or less constant rate to the present time.

How did the Galaxy come to be like this? To study the formation and evolution of galaxies observationally, we have a choice. We can observe distant galaxies at high redshift and see them directly as they were long ago at various stages of their formation and evolution. Distant galaxies are faint however, and not much detail can be measured about their chemical properties and motions of their stars. Also, we cannot follow the evolution of any individual galaxy. Alternatively, we can recognize that the main structures of our Galaxy formed long ago, at high redshift. For example, the halo formed at $z > 4$ and the

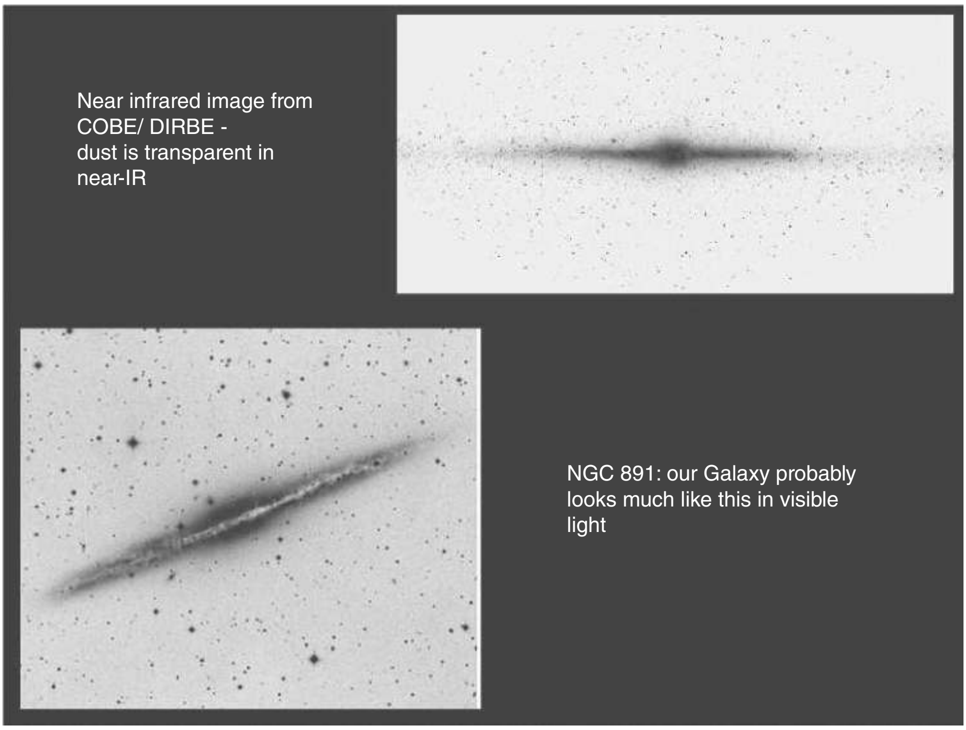

FIG. 1.1. Images of NGC 891 (visible light) and the Milky Way (NIR).

disk at $z \sim 2$. We can study the motions and chemical properties of stars in our Galaxy at a level that is impossible for other galaxies, and we can probe back into the formation epoch of the Galaxy. This approach is now called *near-field cosmology*.

The ages of the oldest stars in the Galaxy are similar to the lookback time for the most distant galaxies we can observe. Both give clues to the sequence of events that led to the formation of galaxies like the Milky Way.

The numerical simulation of galaxy formation (courtesy Sommer-Larsen, 2008) shown in this chapter summarizes our current view of how a disk galaxy like the Milky Way

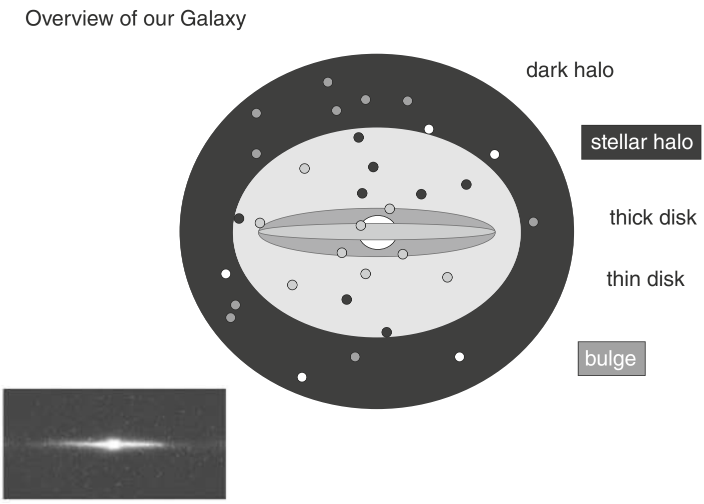

FIG. 1.2. The Galactic components.

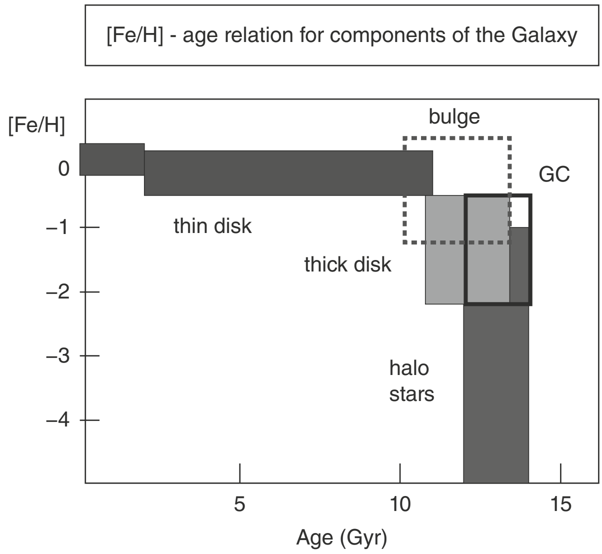

FIG. 1.3. The age-metallicity relation for the Galactic components.

came together from dark matter and baryons. The formation of the oldest halo stars begins at high $z \sim 13$, long before the Galaxy itself became visible. These oldest stars form in small overdensities that ultimately merge to become part of the Galaxy itself.

The formation of the thin rotating disk is a dissipative process that occurs after the dark halo itself has been assembled through the hierarchical merging of smaller structures. Throughout this assembly, much dynamical and chemical evolution is going on. By a redshift of $\sim$3, the Galaxy is partly assembled and is surrounded by hot gas which is cooling to form the disk. At $z \sim 2$, large lumps are falling in, and the Galaxy is already a well-defined rotating system.

The simulation showed the formation and evolution of a large spiral in a ΛCDM simulation. What does each component of the Milky Way contribute to our understanding of the formation and evolution of disk galaxies in the CDM context? Living inside the Milky Way has advantages and disadvantages. The Milky Way will be very good for assessing some CDM issues and not so good for others.

Here are some of the issues with galaxy formation for which CDM has so far not been fully successful. Our Galaxy may be able to contribute to understanding at least some of these difficulties, which include

- the density distribution of the inner dark halo: flat core (as mostly observed) or steep cusp (as predicted)
- the large number of predicted satellites
- the difficulty of forming disks with small bulges
- the related active accretion history within CDM
- the low predicted baryonic angular momentum
- the high predicted fraction of baryons that are converted into stars

1.1.1 *The structure of the inner dark halo*

Simulations consistently predict that the dark halo density distribution has an inner cusp, and observers equally consistently claim that the dark halos have flat inner cores. This argument has been going on for many years. For a density distribution parameterized as $\rho \sim r^{\alpha}$, the halos from simulations have $\alpha \sim -1$ (e.g., Navarro *et al.*, 1996) while rotation curve studies of low surface brightness galaxies (in which the stellar distributions are

Cambridge University Press
978-1-107-02380-2 — Local Group Cosmology
Edited by David Martínez-Delgado
Excerpt
More Information

unlikely to have much effect on the density distribution of the dark halos) typically have α close to zero (e.g., de Blok and Bosma, 2002).

1.1.2 *The number of predicted satellites*

From simulations (e.g., Moore *et al.*, 1999), we would expect a large disk galaxy like the Milky Way to have about 500 satellites with bound masses in excess of 10^8 $M_\odot$. This is much larger than the numbers of satellites detected around the Milky Way optically or in HI. Although new fainter satellites are being discovered, it seems unlikely that the number will approach 500. Are there large numbers of dark satellites? Are some (or all) of the Galactic globular clusters associated in some way with these missing satellites?

1.1.3 *Formation of disks with small or no bulges*

It is currently difficult for ΛCDM to generate galaxies with small or no bulges because of the high continuing merger rate inherent in the CDM context. Most galaxies produced in CDM simulations have very substantial bulges, unlike the Milky Way and the many other giant galaxies with small bulges. Understanding how the bulge of the Milky Way formed could be a contribution toward resolving this problem. Current belief is that the Galactic bulge formed through instabilities of the disk rather than through a merger process. If this turns out to be correct, then we need to understand why the Milky Way escaped the high expected merger rate.

1.1.4 *The high merger rate in CDM*

CDM predicts an active ongoing accretion history, leaving debris of accreted satellites in the stellar disk and halo. The Milky Way stellar halo provides direct evidence of such accretion, of small dense systems that formed before the Milky Way itself. Further evidence comes from the currently disrupting Sgr dwarf galaxy. A very active continuing accretion history of significant sub-halos as expected in CDM is probably inconsistent with the presence of a dominant thin disk. The epoch of last major merger is particularly important for disk survival. We are uniquely located in the Milky Way to evaluate the detailed accretion history of a large spiral and measure the distribution of its first stars.

1.1.5 *The low predicted baryonic angular momentum*

Baryonic angular momentum is lost to the dark halo via hydrodynamical and gravitational effects. This is an old problem, that baryons have been predicted to have less angular momentum than observed. The observational consequences are that disk galaxies are smaller and more rapidly rotating than observed. There is some evidence now that this is less of a problem with higher resolution simulations (e.g., Governato *et al.*, 2007; Kaufmann *et al.*, 2007). This problem is observationally probably better studied in other galaxies.

1.1.6 *The high predicted fraction of baryons converted into stars*

Disk galaxies like the Milky Way appear to have only a small percentage of their baryons in the form of stars. This is not yet seen in most of the simulations, in which most of the baryons are rapidly converted into stars. This problem is probably related to the problem of continuing baryon acquisition needed to fuel ongoing star formation. Without such fueling, the observed star-formation rate would exhaust the current gas supply on a timescale of a few Gyr. Current belief is that there is a substantial reservoir of baryons in the Galactic hot halo, which may come from baryons ejected from the disk or virialized into this hot halo during early baryon infall. The details are poorly understood. Is the baryon acquisition related to the high velocity clouds? Does it come from gas that was previously ejected from the disk, or is gas from the hot halo being entrained by gas lost from the disk and now returning (e.g., Marinacci *et al.*, 2010)? The Milky Way is potentially well suited for investigating these problems of baryon content and acquisition.

1.1.7 *Reconstructing Galaxy formation*

We would like to observationally reconstruct the whole process of galaxy formation as the Galaxy comes together from the CDM hierarchy. What do we mean by the reconstruction of Galaxy formation? We want to understand the sequence of events that led to the Milky Way as it is now. Ideally, we would like to tag or associate the visible components of the Galaxy to parts of the proto-galactic hierarchy: i.e., to the baryon reservoir that fueled the stars in the Galaxy. This seems too difficult. In the process of galaxy formation and evolution from the CDM hierarchy, a lot of information about the proto-galactic hierarchy is lost.

Information about the proto-hierarchy is lost at several phases in the Galaxy formation process:

- as the dark matter virializes
- as baryons dissipate within the dark halo to form the disk
- in the bulge-forming process, whether the bulge forms by mergers or by disk instabilities
- during the subsequent accretion of objects from the environment: information is lost, though some traces remain
- during the evolution of the stellar disk, as orbits are scattered by dynamical processes, including interaction with transient spiral waves and molecular clouds

At each phase, information is lost but some remains. What does the Galaxy remember? What can we hope to discover with Galactic archaeology?

1.1.8 *Signatures remembered from each phase*

We can classify the kinds of information lost as zero order (since dark matter virialized), first-order (since the main epoch of baryon dissipation), and second-order losses (subsequent evolution). Each phase leaves some signatures. In reality, of course, galactic evolution is an ongoing process without such distinct phases. In later chapters, we will look at (1) ways in which we can derive information about the early Galaxy, and (2) some of the processes that cause loss of information or provide bogus information for us to misinterpret.

Zero-order signatures

The virialization phase is dominated by merging and violent relaxation. Early stars form in small elements of the hierarchy, long before the main body of the Galaxy has come together. Some of these stars will become part of the metal-poor halo. The total binding energy E, mass M, and angular momentum parameter $\lambda = J|E|^{1/2}G^{-1}M^{-5/2}$ where J is the angular momentum are more or less established at this phase, although they continue to evolve slowly: E, M, and J determine the gross nature of the galaxy.

The globular cluster system formed around this time: its underlying structure has evolved mainly through the destruction of clusters by evolutionary processes and the changes in the Galactic potential since the clusters formed. Note that the old globular clusters in the Milky Way, LMC, and nearby Fornax dwarf spheroidal galaxy have almost identical ages, within 1 Gyr. Globular clusters in interacting systems like the Antennae (NGC4038/4039) indicate that globular cluster formation is associated with interaction, as in this very early phase.

Some of the properties of the metal-poor stellar halo were probably established in this epoch, as small satellites that had already formed stars were accreted by the virializing halo (more later). The Tully-Fisher law, which relates the rotational velocity and the baryon mass of galaxies, may have been established at this phase, or maybe in the next phase if the loss of baryons associated with star formation was significant.

First-order signatures

What information remains from the epoch when baryons dissipated to form the disk and the bulge? The scale length of the disk may be roughly constant since then. The mass of the disk continues to grow as gas falls in and stars form. Chemical gradients in old components like the thick disk may be conserved but could be affected by radial mixing of stars by transient spiral arms. The vertical scale height of the old disk evolves with disk heating, at least for a few Gyr, but appears to be roughly constant after about 3 Gyr from birth, so we probably see the old disk as it was about 7 Gyr ago.

The old thick disk appears to be a ubiquitous component of disk galaxies, but its formation is not yet well understood. It may represent the early thin disk, dynamically heated by accretion of satellites long ago, or it may have formed much earlier in a gas-rich merger (Brook *et al.*, 2007), or perhaps it is the debris of accreted satellites (Abadi *et al.*, 2003). It is probably now much as it was after it formed and is one of the most important of the Galactic fossils.

The bulge has also probably not changed much since its formation. If it formed by disk instability, then it is probably much as it was about 7 Gyr ago (except for the effects of stellar evolution). The shape of the dark halo may have been affected by the growth of the baryonic component within it, but is probably more or less as it was after the disk began to form stars, except perhaps near the disk plane where the effects of adiabatic compression may flatten and concentrate the dark matter distribution.

Second order signatures

What information remains from the subsequent evolution after the disk began to form? Objects like the Sgr dwarf that are accreted by the Galaxy are tidally disrupted and break up to become part of the stellar halo. Their debris gradually mixes away structurally but their stars conserve some dynamical properties that can in principle be detected in phase space or integral space.

Star-forming events in the disk mostly dissolve and phase-mix around the Galaxy. Some maintain their kinematical identity as moving stellar groups, at least for a few Gyr (e.g., the HR 1614 moving group: De Silva *et al.* (2007). A few survive as open clusters: some of these old open clusters are almost as old as the disk. Old clusters are seen in the Milky Way out to at least 15 kpc in radius. The debris of all of these star-forming events in the disk will maintain their chemical signatures, whether or not the stars stay together in configuration space (old clusters) or phase space (moving groups). Figure 1.4 shows the distribution of [Fe/H], [Mg/Fe], and [Ba/Fe] for the Hyades and Collinder 261 open clusters and for the HR 1614 moving group. These three systems have clearly maintained their chemical identity.

1.1.9 *The metal-poor stellar halo*

Figures 1.5 and 1.6 from Carney *et al.* (1996) show the orbital accentricity and azimuthal velocity against metallicity for a sample of high proper motion stars. These are more recent versions of the famous diagrams from Eggen *et al.* (1962). It shows how the more metal-rich stars that lie in the rapidly rotating thin and thick disks are mostly in orbits of low eccentricity, whereas the metal-poor stars that define the non-rotating stellar halo are in highly eccentric orbits.

Figure 1.7 compares the metallicity distribution of halo stars and globular clusters, as it was known in 1996; now a few halo stars are known with abundances of [Fe/H] < -5. The metallicity distribution of the halo globular clusters is narrower than for the halo stars and does not extend below about -2.5.

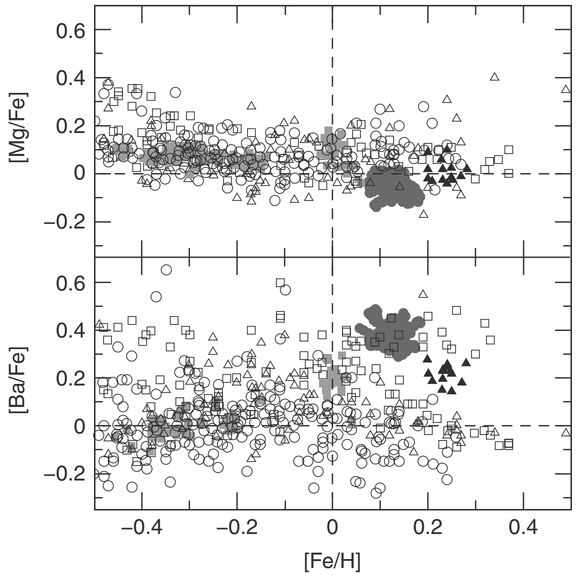

FIG. 1.4. Mg and Ba abundances for stars in the Hyades cluster (full circles), the Collinder 261 cluster (full squares), and the HR 1614 moving group (full triangles), compared with stars in the Galactic disk (open symbols), from De Silva *et al.* (2009).

We discuss the thick disk in more detail later. Most of its stars have abundances between about -0.5 and -1.0, but some of the metal-poor stars ([Fe/H] < -1) have disk-like kinematics and form a metal-poor tail of the thick disk extending down to [Fe/H] $= -2$. About 25% of the stars with [Fe/H] $= -1.5$ near the sun belong to the thick disk: this is not so apparent in Figure 1.7 because this sample of stars is kinematically selected and favors stars of the halo. The important point here is that the stellar halo and thick disk are ancient structures and are very significant for galactic archaeology.

Halo streams

The halo extends out beyond a radius of 100 kpc, and its long orbital timescales allow the survival of identifiable debris from accretion events. The Sgr tidal stream, originally discovered behind the Galactic bulge, appears to extend at least twice around the Galaxy. The stars of the stream include some that are younger than those of the bulge and extend out to much redder colors. The Sgr stream is delineated in longitude by M giants.

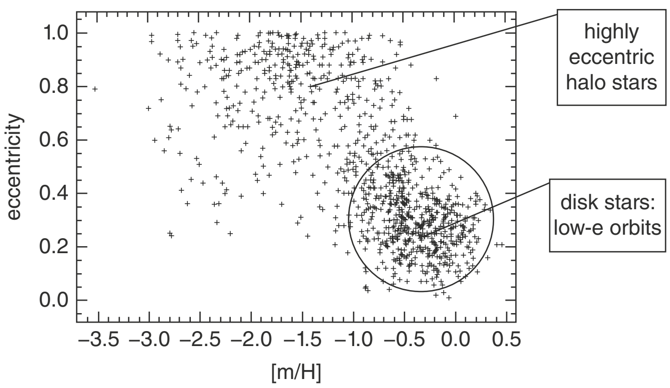

FIG. 1.5. The eccentricity-metallicity relation for stars of the Galactic halo and disk, adapted from Carney *et al.* (1996).

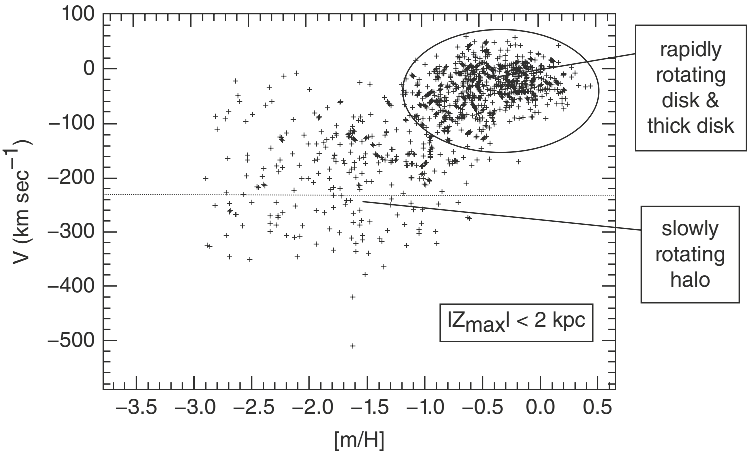

FIG. 1.6. The relation between azimuthal velocity V and metallicity for stars within 2 kpc of the Galactic plane, adapted from Carney *et al.* (1996).

These tidal streams from the currently disrupting Sgr dwarf are interesting for Galactic archaeology, as are the ancient streams from small objects accreted long ago into the halo. The long orbital periods allow these ancient streams to survive in phase space, so the metal-poor halo is the best place to attempt reconstruction of such accretion events. Some are visible in the projection of configuration space on the sky. Even if they are too faint to see in configuration space, they may be visible in projections of phase space, such as position and radial velocity relative to the Galactic center (R_G, V_G), or in the space of integrals of the motion for stellar orbits, like energy and angular momentum (E, L_z).

Accretion is important for building the stellar halo, but it is not clear yet how much of the halo comes from discrete accreted objects (debris of star formation at high z) versus star formation during the baryonic collapse of the Galaxy. At one extreme, simulations of pure dissipative collapse (e.g., Samland and Gerhard, 2003) suggest that the halo may have formed mainly through a lumpy collapse, with only ∼10% of its stars coming

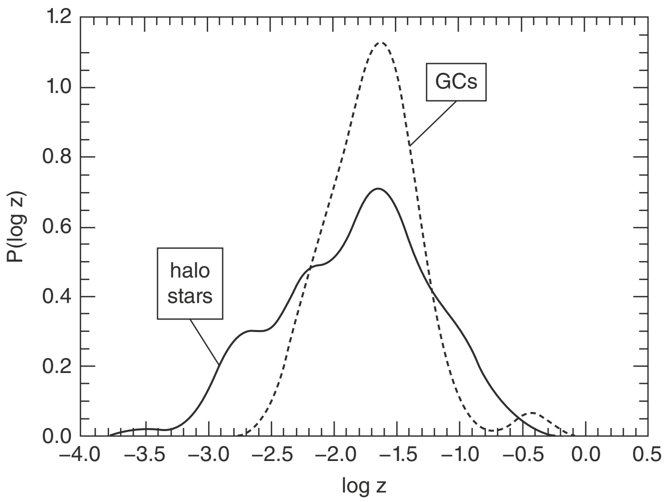

FIG. 1.7. The metallicity distributions of halo stars and globular clusters, from Carney *et al.* (1996). Some halo stars are now known with metallicities below −5.

from accreted satellites. In any case, we can hope to trace the debris of these lumps and accreted satellites from their phase space structure.

The first stars

Where are the first stars now? Simulation indicates that the metal-free stars formed until redshift $z \sim 4$ in chemically isolated sub-halos, far away from the largest progenitor. If these stars survive, they would be spread through the Galactic halo. If they are not found, then their lifetimes are less than a Hubble time, which would imply a truncated IMF for these early stars.

The oldest stars form in the early rare density peaks in the region of the final system. Now they are concentrated to the central region of the Galaxy (Diemand *et al.*, 2005; Brook *et al.*, 2007). The first stars have orbits of fairly high eccentricity (Scannapieco *et al.*, 2006), like the stars of the metal-poor stellar halo.

1.1.10 *General references*

Here are some general references on Galactic dynamics and astronomy.

Galactic Dynamics: Binney and Tremaine (1987, 2008).

Galactic Astronomy: Binney and Merrifield (1998). A more descriptive book, well worth reading for background.

Galaxies in the Universe: Sparke and Gallagher (2007). A good descriptive book about galaxies, including some essential basic theory.

Galactic Populations, Chemistry and Dynamics: Turon *et al.* (2008). This is a very useful and up-to-date compendium of Galactic knowledge, problems, techniques, and surveys.

1.2 Dynamical processes which lose information

In this section, we discuss some of the dynamical processes that lose information or generate potentially misleading information for Galactic archaeology. These processes include the accretion of satellites, dynamical resonances with the bar and spiral structure, heating of the disk by various processes, and radial mixing.

1.2.1 *Accretion and destruction of satellites*

This is an important part of CDM theory: the merging of smaller objects of the hierarchy to form larger objects. Small galaxies are accreted and destroyed by larger galaxies. The debris of the small ones becomes part of the halo, bulge or disk of the larger one.

The orbital energy and angular momentum of the smaller galaxy is absorbed by the dark halo and disk of the larger one. The existence of thin disks constrains the merger history since the disk formed, because disks can be puffed up or destroyed by significant mergers.

The goal here is to describe some of the essential dynamics of merging, accretion and disruption.

Galaxy mergers

The interaction and merging of galaxies and pregalactic fragments is a major element in their formation and evolution. Galaxies are believed to be built up by the merging of a hierarchy of sub-galactic fragments. Mergers of fullyformed galaxies and groups of galaxies are commonly observed, and the end products are believed to be giant elliptical galaxies or large early-type disk galaxies. Accretion of smaller satellites by disk galaxies are believed to contribute to the thickening of the disk, as the disk absorbs energy and angular momentum from the orbit of the satellite. In this section, we discuss the dynamics of the merging and accretion process. Merging usually means the merging of systems of comparable mass, while accretion means the accretion of a small galaxy by a larger one. A nice example of a small galaxy being tidally disrupted by a larger one is seen in the APOD image of the galaxy NGC 5907. The "field of streams" seen in the SDSS star counts for the halo of our Galaxy shows halo streams that are most likely related to one or more accretions of small galaxies into the halo of the Galaxy.

As disk galaxies undergo a close approach, they can interact tidally and merge. The merging stimulates star formation, generates tidal arms and bridges, and disrupts the galaxies. NGC 4038/4039 is a nice example of an ongoing interaction of a pair of spiral galaxies. A fine image of this system can be found on the STScI website.

Much of the discussion follows Binney and Tremaine (1987). When two galaxies interact, direct hits of the stars are unlikely, because the fraction of the area of the galactic disk that is filled by stars is quite small. For example, in the solar neighborhood, the number density of stars is about 20 pc^{-3}, and the radius of a typical star is about $10^{-0.2}$ $R_\odot$, so the fractional area covered is about 10^{-14}. But encounters of galaxies do change the dynamical state of stars in the encountering galaxies. Orbital energy is converted into internal energy within the galaxies, and this can lead to merging.

Some encounters lead to mergers and some don't. Consider two interacting galaxies, A and B. A star in orbit about the center (O_A) of galaxy A gains energy at a rate $\mathbf{v} \cdot \mathbf{g}(\mathbf{r})$, where $\mathbf{g}$ is the (gravitational attraction at the position $\mathbf{r}$ of the star) − (the gravitational attraction at O_A) and $\mathbf{v}$ is the stellar velocity relative to O_A. Let the initial relative velocity of galaxy B relative to A be $\mathbf{v}_\infty$. As $\mathbf{v}_\infty$ increases, the time $t_\circ$ to closest approach of the two galaxies decreases, and the total change of energy of our star

$$\Delta E = \int_0^{t_\circ} \mathbf{v} \cdot \mathbf{g}(\mathbf{r}) dt \tag{1.1}$$

decreases, and the star takes less energy from the orbit. There is a critical velocity v_f such that $v_\infty > v_f$ means that the galaxies can escape to infinity after the closest approach. On the other hand, if $v_\infty < v_f$ then the systems will merge. For $v_\infty > v_f$, the orbits and internal structure of the galaxies are relatively weakly affected. However, for galaxies that lie in the tidal field of a cluster of galaxies, even these fast encounters can be quite damaging. They increase the internal energy of the victim which then becomes more loosely bound and prone to disruption by the tidal field of the cluster.

For the Milky Way, with its prominent thin disk and small bulge, accretion of small galaxies is more important for its evolution than major mergers. We look now at the two main processes involved in accretion.

Dynamical friction

Dynamical friction is the frictional effect on a mass M moving through a sea of stars of mass m. Assume that the smaller masses m are uniformly distributed, and adopt the "Jeans Swindle" (i.e., ignore the potential of the uniform distribution of the m objects). Then the motion is determined only by the force of M and the disturbances that M produces to the distribution of m objects.

M raises a response in the sea of smaller objects, and this response acts back on M itself. Summing the effects of the individual encounters of M and m, we see that M suffers a steady *deceleration* parallel to its velocity $\mathbf{v}$. If the velocity distribution of m is Maxwellian

$$f = \frac{n_\circ}{(2\pi\sigma^2)^{3/2}} \exp(-v^2/2\sigma^2) \tag{1.2}$$