Stellar Populations and the Formation of the Milky Way

By STEVEN R. MAJEWSKI

Department of Astronomy, University of Virginia, Charlottesville, VA 22903-0818, USA

I review topics in the area of Galactic stellar populations and the formation and evolution of the Milky Way, with particular emphasis on the role of globular clusters in tracing stellar populations and unraveling the Galactic history. While clusters provide a means of determining some global properties of stellar populations, our understanding of stellar populations serves also to guide us to scenarios for the origin and evolution of the family of Milky Way globular clusters.

1. Introduction

1.1. Stellar Populations – General Concepts

The goal of stellar population studies is to understand the formation and evolution of galaxies through investigation of the detailed distributions of properties – including, but not limited to, stellar types, kinematics, chemical abundances, ages, and spatial distributions – of its constituent luminous parts. Inherent to this endeavor is the notion that stars in galaxies can be categorized into groups – populations with shared, or well-defined distributions of, properties. Since stars generally form in associations and clusters†, these families of sibling stars may represent the smallest viable population “unit” – the “simple stellar population” (“SSP”; Renzini & Buzzoni 1986) – of coeval, initially chemically homogeneous stars on similar orbits through a galaxy. In the context of this discussion, then, a galaxy is made up of populations, each which is assembled from particular combinations of SSPs:

\[
\text{GALAXY} = \sum_{i=1}^{m} a_i(\text{POPULATION}_i) \\
\text{POPULATION}_i = \sum_{j=1}^{n_m} b_j(\text{SSP}_j).
\]

The larger the relative sizes of homogenized SSPs, and the fewer the number of them that constitute a galaxy (i.e., the smaller are \(m\) and \(n_m\)), the easier it should be to unravel its history. In the other less tractable extreme, the SSPs might be extraordinarily small and multitudinous, at the level of individually forming stars.

In principle, a galaxy can be broken down into an expansion of smaller population units, and, by retrieving an age-dated collection of components of this expansion, one can put together a timeline of formation. More often than not, identification of individual SSPs is difficult in complex galaxies. In this circumstance, the hope of “stellar population” studies is that the SSPs have been strung together in somewhat simple patterns constituting what I will call principal component stellar populations‡ (Figure 1). In this

† Although this may not be true for the first stars that may have formed in the early universe, the so-called “Population III”.

‡ Use of this terminology here is intended as an allusion to principal component analysis.
way, the problem is reduced to the (hopefully simpler) task of identifying these patterns characteristic of each principal population. Examples of principal component populations in the Milky Way might be “the disk”, “the halo”, or “the bulge”, each constructed of smaller, simple stellar populations characteristic of specific sites and epochs of star formation (perhaps star clusters or associations).

Thus, when confronted with a galaxy having a complex mixture of stellar populations we seek correlations between various observable attributes, such as

- **spatial distributions** e.g., density laws
- **kinematics** velocities, velocity dispersions
- **chemistry** e.g., mean [Fe/H], chemical abundance patterns ([O/Fe], [Ca/Fe], [Zn/Fe], ...)
- **ages** reflected, e.g., in the types of stars seen (see Figure 2),

in order to find and characterize the principal component populations that will allow us to reconstruct a complete, physical, *galactic chemodynamical evolutionary model* of the entire system. This ultimate model of the system must necessarily incorporate all baryonic components, and for each principal stellar population might include the model attributes: $f = f(\mathbf{x}, \mathbf{v}, t)$, the evolution of the phase space distribution of stars; $g = g(\mathbf{x}, t)$, the evolution of the gas density; $X = X(\mathbf{x}, t, X_1, X_2, X_3, ...)$, the evolution of the detailed abundances of atomic species $X_i$ in the interstellar gas out of which stars form; $\text{SFR} = \text{SFR}(\mathbf{x}, t)$, the *star formation rate*, the number of stars formed per unit time interval; and $\text{IMF} = \text{IMF}(\mathbf{x}, t)$, the instantaneous *initial mass function* which denotes how the new stars are distributed by mass. In general, these model functions are not observable, at least not completely. The physics of the system drives their evolution, and, if our model is an accurate representation of the Galaxy, the model functions, in turn, will be able to predict the proportions of observed stellar types, i.e., the $a_i$ and $b_j$ in the schematic equations above.

Unfortunately, determining this array of model descriptors is impossible for any galaxy. In the first place there are limitations of physics, in that some aspects of the evolution of galaxies may well be hopelessly irrecoverable, erased by “the operation of contingent processes that cannot, even in principle, be inferred from observations of its present.
Figure 2. The luminosity-weighted relative contribution of various evolutionary stages to the integrated bolometric luminosity of a simple stellar population as a function of age (bottom scale) and mass at the turnoff of the main sequence (top scale). This example model is for a specific composition of helium, Y, and metals, Z. “MS” = main sequence, “SGB” = subgiant branch, “RGB” = red giant branch, “HB” = horizontal branch, “AGB” = asymptotic giant branch, and “P-AGB” = post-asymptotic branch stars. From Renzini & Buzzoni (1986).

state” (Searle 1993). For example, the processes leading to dynamical relaxation alter stochastically the kinematical attributes of stars in dense regions. Fortunately, as we shall see (section 3.4.3), at least some regions of galaxies avoid this regime.

Secondly, technological limitations in the observer’s domain prevent us from measuring detailed distributive properties of stellar systems in external galaxies, though great strides are being made in this area for Local Group galaxies. Whereas previously our knowledge of stellar populations in galaxies was limited to that which could be ascertained from their integrated light, now color-magnitude diagrams (“CMDs”) of resolved stars on images from the Hubble Space Telescope yield detailed mappings of the SFR histories of our neighbors (see below). This remarkable advance is a product of the unique time stamps afforded by the orderly progression of stellar evolution phases in SSPs. These various phases separate into branches of the CMD (see the discussion of stellar evolution and the branches of the CMD by Castellani and by King, this volume), and the presence or lack of representative stellar evolution phases, or the relative numbers of stars between phases, may be exploited to age date simple population systems (Figure 2). Because the age of a stellar population is tied directly to the mass of stars evolving off the main sequence, particularly useful are stars in stellar evolutionary phases specific to a narrow stellar mass range. An example of a logic chart for time stamping a stellar population that utilizes these well-defined age indicators is shown in Figure 3. Intrinsically bright age indicators are, of course, especially useful for the study of more distant galaxies.

By virtue of the Vogt-Russell theorem (Vogt 1926, Russell et al. 1927), which states that the mass and chemical composition of a star are sufficient to define uniquely its structure (which in turn determines, albeit not necessarily uniquely, its position in the CMD) at any given age, the shape of the CMD for an ensemble of stars in an SSP is a function of both its age and chemical composition (with relative densities modulated by the initial mass function). The metallicity/age dependence of the CMD for the SSPs represented by globular clusters is discussed by Castellani and by King (this volume).
While some combinations of age/abundance do yield similar morphologies in the CMD, the CMD is a powerful tool for ascertaining information on both the age and mean abundance of an SSP.

Clearly, the CMDs of more complex systems consisting of multiple stellar populations will be commensurately more complicated. CMDs for simple agglomerations of SSPs, as are now being found in some of the dwarf galaxies of the Local Group, can often be well represented by a superposition of a small number of single age/abundance CMDs. A good example is the Carina dwarf galaxy, a satellite of the Milky Way that appears to have formed stars in three bursts, each which leaves a characteristic hallmark in the galaxy’s CMD (Figure 4). From deconstruction of the CMDs of complex systems, we may begin to connect some of the chemical and star formation history of these systems.

A convenient way to visualize these connections is by way of the “Hodge population box” (Hodge 1989), a three dimensional representation with the axes of age, mean metal abundance ([Fe/H]), and star formation rate. The Hodge population box for a rather simple system like Carina, would look something like Figure 5. Typical galaxies, especially more massive ones, have more complicated star formation histories that include more extended periods of star formation and enrichment. An example of a Hodge population box for a galaxy with a more complicated star formation history is shown in Figure 6. A great achievement of the last decade is that with large aperture telescopes, the precision afforded by CCD detectors, and the ability to resolve stars in nearby galaxies, approximate star formation and abundance distributions are being worked out for a

† These degeneracies can often be broken by constructing CMDs with different filter systems.
‡ In principle, equivalent representations that chart the enrichment in other chemical species could be formulated. However, it is traditional to evaluate the overall level of enrichment via the ratio of iron to hydrogen. Moreover, we are not presently able to discern much more than the mean level of enrichment of stars in other galaxies, although the measurement of chemical abundance patterns for stars in the Milky Way, as well as in intergalactic gas clouds (via absorption lines by the gas in the spectra of background quasars) have become cottage industries.
majority of the Local Group (see summary of Hodge boxes for nearby galaxies by Grebel 1998).

Clearly, completing our chemodynamical models requires tying together the chemical enrichment history of galaxies with the dynamical evolution of the stars and gas within. One might imagine visualizing the dynamical history of a galaxy with a dynamical population box, such as that shown for a fictitious galaxy in Figure 7. In this case two axes of the box are given by the star formation rate and age, as before. The third axis is a measure of the relative distribution of the stars formed at an epoch in the disordered motion of velocity dispersions ($\sigma$) versus the ordered motion in rotation ($V_{rot}$): $\log (V_{rot}/\sigma)$.\*

\* It is common in Milky Way studies to refer to rotational velocities as measured with respect to a single axis of rotational symmetry in the galaxy. With a single defined axis it is possible to have populations with negative rotational velocities, when they are in retrograde motion compared to a defined axis. Because there is little evidence suggesting differing axes of symmetry for different populations in the Milky Way, and to preserve the commonly used sense of $V_{rot}$.
The quantity $V_{\text{rot}}/\sigma$ can be thought of as a measure of the relative importance of rotational support to pressure support for the stellar population. As an example, the stars found in a galaxy that formed similarly to the model described by Eggen, Lynden-Bell & Sandage (1962) – with a rapid collapse of more or less randomly moving gas followed by the formation of a disk with spin-up (see Section 3.1 below) – might have a dynamical population box like that shown in Figure 7.

Unfortunately radial velocities for stars in external galaxies are difficult enough to obtain; the hope of measuring complete internal kinematics with proper motions for stars in external galaxies must await the promise of new technologies. Projects as technologically advanced as the planned Space Interferometry Mission, which will have microarcsecond positioning capability, will deliver astrometric precision to measure transverse motions of bright stars in only the very nearest galaxies.

In the Milky Way at least we have the unique opportunity to measure a full array of properties for individual stars and clusters. Thus, the Milky Way is presently the only “laboratory” galaxy where we can get detailed information on the kinematics of stellar populations to tie together with the chemical, age, and spatial properties toward our goal of synthesizing a complete chemodynamical model of formation. However, it is common to assume (and there is little reason to think otherwise) that the Milky Way is representative of galaxies of similar Hubble type. The status of our knowledge of our “laboratory” in this context is the emphasis of the remainder of this discussion. In Section 4 we conclude by attempting to assemble “population box” representations for principal component populations of the Milky Way.

Before proceeding, it is worth noting one possible ambiguity of the “population box”.

as a quantity measured with respect to the rotational axis of the young Milky Way disk, I will occasionally employ the quantity $\log (|V_{\text{rot}}|/\sigma)$ here. However, in the general case when populations may not share common axial symmetries, it may be more physical to think of $V_{\text{rot}}$ as the velocity pertaining to the angular momentum axis for each population. An illustration of the difference is had by imagining $\log (V_{\text{rot}}/\sigma)$ for a stellar population concentrated to an annulus on a polar orbit around the Milky Way, like what one might get from the tidal disruption of a satellite galaxy (Section 3.4.3); $V_{\text{rot}}$ is clearly different when measured with respect to the Galactic disk and with respect to the angular momentum axis of the population itself!
The intention is to demonstrate the properties of the star forming gas as presumably “stored” in the properties of the extant stars that we actually observe to derive these data. Unfortunately, several problems conspire to thwart a simple mapping from observed to original properties. For example, in their evolution past core hydrogen burning, stars can alter chemically their stellar atmospheres through convective mixing and dredge up of products of the nucleosynthetic yield from their fusion cores. To a lesser degree, the process of astration, the accumulation of enriched interstellar gas as a star passes through the disk (Clayton 1989) will also alter the observed abundance from the birth abundance. Similarly, a number of processes (some discussed in the next section) can alter the dynamics of stars. Generally these processes, like the scattering of disk stars off of giant molecular cloud complexes via the Spitzer-Schwarzschild (1953) mechanism, serve to increase the random motions of stars. Thus, one must always bear in mind the evolution from the birth values to the observed values when discussing a property of a stellar population, or visualizing it in a population box as we do here. In the present paper, we assume that the observed $[\text{Fe/H}]$ distributions are relatively unchanged from the observed ones, but we do concern ourselves with possible changes between the birth dynamics and observed dynamics of stars.

1.2. Globular Clusters as Stellar Population Tracers

Since globular clusters are the main topic of this Winter School, it is worth emphasizing their importance to stellar population studies. Clusters, open and globular, provide us with unique tracers of stellar populations. Obviously, because of their luminosity, they are easy to find and easy to see to great distances. More importantly, perhaps, clusters seem to have been formed in relatively brief star formation bursts with a resultant yield of stars having great chemical homogeneity that testifies to efficient mixing of cluster precursor gas. This is evidenced, for example, by the remarkably small spread in the mean metal abundance of stars within each cluster, which is typically given by $\sigma([\text{Fe/H}]) < 0.10$ (see
summary by Suntzeff 1993, and Gratton, this volume). This is to be contrasted with a large spread in abundances from cluster to cluster. Clusters may well be the closest thing we have to cohesive, “simple stellar populations” and may be paradigms for the building blocks out of which at least some of the stellar populations in galaxies are constructed (the possibility of specific populations formed directly from globular clusters is addressed below). In the Milky Way, globular clusters represent the oldest sites of star formation still cohesively bound; globular clusters therefore hold clues to the early chemodynamics of the Milky Way.

Through the color-magnitude diagram, the distances of clusters can be estimated fairly readily, while the age and mean metallicity are, in principle, ascertainable, at least in a relative sense. Moreover, through high resolution studies of the coeval stars in clusters, we are making some progress toward understanding sources of observed stellar variation in chemical abundance patterns, though it is still not clear to what degree these variations are a result of primordial inhomogeneities in the pre-cluster gas and how much they are the result of the mixing of stellar interiors during the evolution of the cluster stars (Kraft 1994, Sneden et al. 1997, Gratton, this volume). These variations are often seen among the elements of the CNO cycle, as well as other light elements like Na and Al.

Finally, compared to individual field stars at the same distance, the orbits of globular clusters (and dwarf satellite galaxies) can be determined with much better precision, since cluster mean proper motions can be determined as the average of a large number of member stars.

While globular clusters have played, and continue to play, a central role in stellar population studies, it is important to be cognizant of their limitations. Foremost among these shortcomings for Milky Way studies is the limited number of globular clusters. It is unlikely that the some 150 known Galactic globular clusters (Harris 1996) represent a complete census (e.g., some clusters probably live in highly dust obscured regions of the Galaxy), but it may well be close to complete. Nor is it likely that the present family of globulars represents the initial retinue of the Milky Way, as a number of processes act to destroy globular clusters (see Weinberg 1993; Elson covers these topics extensively in her contribution to this volume). The usual means is by expansion of the cluster “halo” to the point where stars on the outermost orbits are liberated from the gravitational hold of the cluster. Several processes will act in clusters regardless of outside forces:

- **Evaporation** operates as stars in the cluster interact through two body encounters. Gradually, through two body relaxation, the cluster approaches equipartition of kinetic energy and this produces a tendency toward higher velocities for the lower mass stars, which can escape if they overcome the cluster escape velocity (Ambartsumian 1938, Spitzer 1940, 1958). As low mass stars escape, the binding energy of the cluster decreases, allowing even more stars to reach the lowering escape velocity. While models of this process based on the properties of Galactic globular clusters indicate timescales for total evaporation that exceed the age of the Galaxy by a factor of $2 - 20$, the change in the cluster mass function is noticeable in a Hubble time (Johnstone 1993).

- **Core collapse** in clusters occurs because two-body relaxation preferentially removes the “hotter” stars near the core. Because the lost stars are preferentially the less massive, mass segregation hastens the collapse (Spitzer 1969). The core reacts to create dynamical energy to stave off further collapse, and two-body relaxation transports this energy outward, increasing again the number of stars that can achieve escape velocities. Evaporation rates in post collapse globulars may be several times greater than normal clusters (Lee & Goodman 1995). Rotation of the cluster apparently acts to enlarge the regimes of both normal and post-collapse evaporation (Longaretti & Lagoute 1997).
Mass loss by stellar evolution also decreases the binding energy of the cluster, assisting with evaporation. Mass loss occurs through supernovae explosions, ejection of planetary nebulae, and stellar winds. These processes, because they are mainly associated with high mass stars, are most important for the early evolution of clusters. Energetic disruption of the cluster by supernova explosions is also possible at early times.

Each of these mechanisms for shedding stars is enhanced in the presence of the Galactic potential in that the cluster becomes “tidally truncated” at the approximate point where the Galactic force exceeds the cluster force and effectively reduces the required velocity of escape. For clusters of mass $M$ on nearly circular orbits around the Galaxy, the tidal radius of the cluster is given by $R_t = R_G(M/3M_G)^{1/3}$, where $R_G$ and $M_G$ are the Galactocentric radius of the orbit and the mass of the Galaxy interior to this orbit.

The Galactic potential also works actively to destroy clusters in several ways:

- **Disk shocking** of the cluster occurs as it plunges through the disk. The transient perturbation acts to compress the cluster in the direction perpendicular to the disk, resulting in the acceleration of cluster stars. The kinematical “heating” of stars in the outer regions of the clusters from this “compressive gravitational shock” enhances the ability for some to escape (Chernoff et al. 1986, Spitzer 1987).

- **Bulge shocking** occurs when a cluster passes near the Galactic center. Both disk and bulge shocking were probably more important for the primordial cluster population because these processes selectively act on clusters with highly eccentric orbits (Aguilar et al. 1988). This also suggests the likelihood of differences in the distribution of orbital types between destroyed and extant clusters.

- **Bar shocking** may increase cluster destruction rates above those from bulge shocking. The effect, however, appears to be small and active only on clusters that pass within a few kiloparsecs of the Galactic center (Long et al. 1992).

- **Tidal gravitational shocking** is similar to the above processes, and occurs when the cluster passes another spherical mass (Spitzer 1987). The predominant affect is by passage of giant molecular cloud complexes in the Galactic disk, which, for less dense clusters (not globulars) can destroy them in one encounter. Globular clusters can be destroyed by repeated shocks from molecular cloud complexes, but this is limited to particular orbits, specifically, prograde orbits of low eccentricity and low inclination to the disk (Surdin 1997). The effect of globulars on each other is minimal (with timescales $3 - 4$ orders of magnitude longer than their lifetimes), though cluster-cluster shocking may have been important for large clusters at early times (see Elson chapter).

- **Galactic tides** will strip stars that reach the cluster tidal radius with sufficient energy. The stars that come off with negative relative orbital energy with respect to the cluster should form a leading débris stream at a slightly smaller orbital radius, and those that come off with positive relative orbital energy will form a trailing débris stream at a slightly larger orbital radius. This process is seen to occur in at least some of the dwarf satellites in the Milky Way (see Section 3.4.3 below). More recently, Grillmair (1998), collaborators, and others (see summary in Grillmair 1998) have enumerated 20 Galactic and M31 globular clusters showing evidence of tidal tails.

- **Dynamical friction** occurs as a cluster continuously loses orbital energy to stars (and dark matter) in the field through which it passes (Chandrasekhar 1943a,b; see also Saslaw 1985). Gradually, the cluster spirals towards the center of the potential. Both tidal stripping and evaporation of stars from a cluster will increase as the cluster orbit decays by dynamical friction. This process would act preferentially to deplete the interior parts of galaxies of its most massive clusters (Tremaine et al. 1975).

- **Field star diffusion** through globular clusters results in the star’s deceleration
Several conclusions follow from an analysis like that resulting in Figure 8. The first is that the initial globular cluster population of the Milky Way may have been quite...
larger than it is today. Those clusters no longer with us have dispersed their stars into the field populations of the Galaxy. A second conclusion relates to the Galactocentric radial dependency of the survivability of a cluster, whence there are particularly strong effects on the innermost clusters. The influence on inner clusters, combined with the effects of bulge and disk shocking on objects of larger mean orbital radius but highly eccentric orbits, raises important questions regarding the relationship of the presently observed clusters both to the primordial cluster population as well as to the field star population that the observed clusters may, or may not, trace. I return to the question of a contribution of destructed globular clusters to the formation of the halo in Section 3.4.3 and to the formation of the bulge in Section 3.5.

2. The Size and Shape of the Milky Way and its Stellar Populations

In my view, interpretation of the chemodynamical properties of stars and clusters in the various populations of the Milky Way is highly dependent on an understanding of the spatial structure of these populations. Assigning stars and clusters to one or another of the Galactic populations is made an especially troubling task because of the significant overlap in chemistry, kinematics and spatial distributions between populations. Untangling them would be aided if we could at least determine where in the Galaxy certain populations dominate.

2.1. Assessing Spatial Distributions of Populations With Stars

The earliest advances in understanding the “sidereal universe” were based in the science of star counting. For example, the idea that the Milky Way was a disk-like structure (as proposed by Thomas Wright and Immanuel Kant in the 1700s) became evident with the basic telescopic observation (first done by Galileo) of resolving the “via lactea” into a dense sea of stars, much more populated than other parts of the sky. However, many ideas concerning the structure of the Galaxy were based more on mere speculation rather than hard-core data until William Herschel began a systematic approach to determining the distribution of stars. Herschel may be said to have begun the science of statistical astronomy†. By counting stars in defined areas of the sky (“star-gaging”), and assuming that stars had the same intrinsic luminosity and were distributed equally throughout the Galaxy, and moreover, assuming that his telescope could see to the edge of the Galaxy, Herschel (1785) produced a three dimensional model of the universe. Naturally, the Sun was found to be near the center of this distribution, which was shaped like a flattened (axis ratio 5:1) American football. By 1817, Herschel recognized various problems with the assumptions used to construct his model (for example, by noting himself that stars in binary systems were often of different luminosities) and concluded that (1) the number of stars in a field related not only to the size of the Galaxy in that direction but to the stellar density, and (2) even his great telescopes might not be able to “fathom the Profundity of the milky way.” It became evident that constructing a model of the stellar system without knowledge of the intrinsic brightnesses (distances) of stars beset any direct approach.

As if lack of distances were not enough of a problem for gauging the Milky Way, the possibility of the absorption of starlight in space (first postulated by H.W.M. Olbers in 1823

† Students inclined to the history of science will find fascinating reading in E. R. Paul’s book *The Milky Way Galaxy and Statistical Cosmology 1890–1924*, from which some of the following (highly selective) discussion is based. Parts of Paul’s book are also presented in articles in the *Journal of the History of Astronomy*. Reid (1993) provides additional historical context.
and confirmed by Trumpler in this century) complicated matters more, although postulation of this process did provide apparent help with certain other problems. Friedrich Struve (1847), building on Herschel’s work and concerned with the inconsistencies in his assumptions, applied the first statistical approaches to understanding the density laws of stars. In so doing, and as a means also to avoid Olbers’ paradox‡, Struve became convinced that interstellar absorption did indeed occur¶.

Even with the development of extensive catalogues of hundreds of thousands of stars, like the Bonner Durchmusterung, and the promise of even more systematic work with photographic star counts to fainter magnitudes, e.g. the Carte du Ciel and the Cape Photographic Durchmusterung of David Gill (see Reid 1993), most workers in “Galactic structure studies” at the end of last century were still relying on the traditional “direct” analyses of the data, a method introduced and developed by Struve, William Herschel and Herschel’s son, John. No doubt frustrated by his own attempts over a decade and a half with this same approach, Hugo von Seeliger (steeped in a mathematical physics training at Leipzig, and heavily influenced by Neumann and Gauss) introduced a powerful new mathematical approach to star counts around 1900. Von Seeliger’s “Fundamental Equation of Stellar Statistics” got around the “vicious circle”† of the interdependence of stellar density laws and the stellar luminosity function – i.e. the fact that full knowledge of one requires full knowledge of the other – that had plagued all previous work. A quarter century earlier, the Swedish astronomer Gyld` en suggested that the brightnesses of stars might be distributed according to an analytical frequency function (he suggested a Gaussian form), but this idea had lain dormant (and apparently hidden from the non-Swedish community) until von Seeliger because the theory of integral equations was still so new that Gyld` en himself could not solve his integral expression! Now von Seeliger adopted the idea of a luminosity function (of unknown shape, of course, but constrained by \( \int \Phi(M)dM = 1 \)) and combined it with the density law (also unknown, of course) \( D(r) \), for an expression in terms of the observable star counts per magnitude, \( A(m) \), in a given direction of the sky:

\[
A(m) = \int \Phi(M)dM \int D(r)dr
\]

Several points are worth noting. First, the simple elegance of the Fundamental Equation belies the fact that it is not uniquely invertible. At some level, we remain in the “vicious circle” of before. However, von Seeliger provides us with a statistical means by which we may compare combinations of functional form for the luminosity function and density law against the observations. So, for example, if one assumes a standard form for the luminosity function, then a global form of the density law may be derived (more accurately, tested) by fitting counts in various directions of the sky. On the other hand, if one has a sense of the density law, then we may arrive at a functional form of the luminosity function. These examples suggest the idea of an iterative approach to the solution of von Seeliger’s equation. I address this and other approaches to the problem in the next section.

Von Seeliger understood well that application of his formula was limited by the quality of the data, and he devoted the next decades to refining his solutions as new catalogues

‡ Olbers pointed out that if space were isotropically filled with stars and were also infinite, then the sky should be uniformly as bright as the surface of a star, since any line of sight should intercept a star.
¶ Alas, while absorption is a key factor in Galactic studies, Struve’s suggestion that it was the solution to Olber’s paradox was incorrect.
† Paul (1993), p.76
of data were produced. This problem was not lost on the other industrious statistical astronomer of the time, Jacobus Kapteyn, in Groningen. The director of an astronomical institution without a telescope, Kapteyn first volunteered to measure and reduce Gill’s plates from the southern hemisphere Cape Durchmusterung. With a voracious appetite for data, and a knack for convincing directors of major observatories to contribute large amounts of telescope time for his (aptly named) “attack” to address systematically the problem of understanding “the sidereal world”, Kapteyn (1906) devised the Plan of Selected Areas (“SAs”). The SA’s are a regularly spaced grid of survey regions around the sky. Apart from starcounts, Kapteyn was extremely motivated to create this Plan by his extensive efforts to understand Galactic kinematics. There ensued a period of great activity, whereby substantial amounts of effort the world over were devoted to contributing photometry, astrometry, and spectroscopy of stars in Kapteyn’s 206 SAs. The grand scope and initially perceived importance of the Plan was such that coordination was essential, and this prompted the eventual creation of IAU Commission 32: Selected Areas as well as the Subcommittee on Selected Areas of IAU Commission 33: Structure and Dynamics of the Galactic System. Coordination of the Plan was also the subject of two of the earliest IAU Symposia (Nos. 1 and 7).

Since the mid-part of this century – when it was discovered that the spiral nebulae were extragalactic systems, and coincident with the rise in emphasis on star clusters as a tool for Galactic astronomy (see Paul 1981) – activity on the SAs has unfortunately strongly declined (IAU Commission 32 no longer exists). However, the wisdom and value of a systematic and coordinated astrometric, photometric and spectroscopic approach to studying the Milky Way is now obvious. There is growing evidence that various subsystems of the Galaxy (e.g., the bulge with its bar, the disk with its warp, and the apparently dynamically unrelaxed halo with its gaseous and stellar, tidal streams; see below) are highly asymmetric, and therefore not described adequately by global models derived from only a few lines of sight (as is common practice). Kapteyn’s original vision of a fully integrated photometric, astrometric and spectroscopic survey has never been fully realized. Ironically, the decline in SA activity has overlapped with the development of modern instrumentation that might be brought to bear on the program with far more efficiency, precision and depth than Kapteyn could have imagined. Fortunately, with the development of high speed photographic plate scanners able to produce huge imaging databases like the Digitized Sky Survey and the APM catalogue, and astrometric databases soon to be produced by the US Naval Observatory and the Minnesota groups, as well as the imminent production of the Sloan Digital Sky Survey, aspects of the Kapteyn vision may soon be revisited with a vengeance.

2.2. Modern Starcount Analyses and Galactic Structure

In modern usage, the von Seeliger equation is given by

$$A(m, S) = \sum_i A_i(m, S) = \Omega \sum_i \int \Phi_i(M, S) D_i(\vec{r}) r^2 dr,$$

where $A(m, S)$ refers to the differential counts of stars of type $S$, $\Omega$ is the area of the sky surveyed in steradians, $D_i(\vec{r})$ is the density law as a function of position in the Galaxy, and $\Phi_i(M, S)$ is the luminosity function, in this case presented as the relative number of stars per magnitude of type $S$. As usual, the absolute magnitude of the star, $M$, is

‡ Today an astronomer (especially one so highly placed in the astronomical community) making such an offer would make a prized collaborator!

¶ “Kapteyn presented the unique figure of an astronomer without a telescope. More accurately, all the telescopes of the world were his.” Frederick Seares (1922).
related to the apparent magnitude of the star via

\[ m = 5 \log(r) + M - 5 + a(\vec{r}), \]

which makes the integral over \( r \) calculable. The function \( a(\vec{r}) \) accounts for the three dimensional distribution of absorbing material along the line of sight. Because we recognize that the Galaxy consists of separate principal populations (disk, halo, etc.), which may (or may not) be discrete and follow their own density laws, the starcounts are given as the sum of \( i \) individual populations.

The introduction of the stellar type, \( S \), is meant to address the problem of degeneracy of contributors to a given apparent magnitude bin (e.g., a nearby, faint star and a distant, but bright star). That is, it is a means by which to simplify the convolution of luminosity function and density law that makes the equation uninvertible. As demonstrated by the work last century, the starcounts problem becomes intractable if you do not have some assessment of the absolute magnitude, and therefore distance, of each star. Alternatively, if one knew in advance that stars of a certain type (however defined) had the same magnitude, and one could identify stars of that type \( a \) priori, then the problem simplifies to the star-gauging of the past. For example, if one only sought stars of spectral type G2V (stars like the Sun) then one could obtain the density law for that stellar type readily through measuring distances by apparent magnitudes relative to the Sun (sans the effects of absorption). Clearly, the narrower the definition of \( S \), the more accurate the solution becomes.

In practice, spectral types would not be easy to obtain for lots of stars. But fortunately, unlike our predecessors of last century, we can gain ready access to some information about stars to practical magnitude limits that allow us to improve our approach to the luminosity function portion of the von Seeliger integral. For example, with a few photographic exposures, one might identify certain classes of variable stars, like RR Lyrae, that have more narrowly limited mean magnitudes.

More commonly, starcounts are measured in pairs of filters, so that \( S \) may be restricted to stars of certain colors. An example of such a data set is the field star CMD shown in Figure 9 (left panel). Note that restricting starcounts to stars of specific colors still yields degeneracies in absolute magnitudes, but in general, these degeneracies are bimodal at most colors. For example, at \( B - V = 1.0 \), the color of K type stars, we will see contribution from K giants (\( M_V \sim 0 \)) and K dwarfs (\( M_V \sim 6 \)). Depending on the density law and its interplay with the growing volume element along the line of sight, it is often the case that the contribution from one or the other of these luminosity classes will dominate at certain apparent magnitudes.

Thus, a common approach to the starcount problem is the following:

a) Measure starcounts \( A(m, \text{color}) \).

b) Assume a given luminosity class for stars of this color, which then narrows considerably the range of absolute magnitudes for the stars. For example, when looking at high Galactic latitudes and faint (\( V \gtrsim 17 \)) magnitudes, it is often safe to assume that one is dealing entirely with dwarf stars. This particular assumption works best for red stars, where the absolute magnitude separation of dwarfs and giants is greatest, but it becomes unreliable for colors near the main sequence turn off.

c) From an established form of the color-magnitude relation for the appropriate luminosities.

\( \dagger \) An often used mapping of reddening is that given by Burstein & Heiles (1978, 1982), but improved, higher resolution maps based on COBE/DIRBE and IRAS/SISSA far infrared data have been derived by Schlegel et al. (1998).

\( \ddagger \) If the density law for the population from which the giants are seen falls faster than the volume element grows (i.e., as \( r^{-3} \)), as might be the case for the halo, then the counts for the giants falls off. At faint enough magnitudes, presumably you run out of Galaxy.
Figure 9. Left: Counts in the $(V, B - V)$ plane for 0.3 deg$^2$ at the North Galactic Pole (SA57) from the photographic survey of Majewski (1992). The field star CMD shows a characteristic two-ridge feature (first discussed by Kron [1980]) in this particular color plane. The left ridge is caused by the build up of stars at the color of the main sequence turnoff for populations of the age of the upper disk and halo. The right ridge is caused by the “saturation” of $B - V$ colors for late type stellar spectra due to the growth of TiO absorption in both the $B$ and $V$ bands. Right: Starcount model fit to the data on the left. Crosses are for model halo stars, squares are for IPII stars, and triangles are for thin disk stars. From Reid & Majewski (1993).

**nosity class, assign an absolute magnitude and, from the apparent magnitude for each star, derive a photometric parallax, $r(m, M)$.**

d) Fit model functions to the count density, $A(r)$.

This approach is particularly well suited for study of the disk because of its expected primary dependence on height from the Galactic plane (other populations of the Milky Way show a stronger dependence on the radius from the Galactic center). An application of the technique is shown in Figure 10.

As noted in the first careful studies with this approach in the 1980’s (Brooks 1981, Yoshii 1982, Gilmore & Reid 1983), fitting the disk density law with an analytical function requires at least two components, to account for an apparent break around 1 kpc from the Galactic plane. The expected “old disk” or “thin disk” population dominates the first kiloparsec or so. Beyond this dominates the Intermediate Population II, a population first formerly characterized at the historic 1957 Vatican Conference on Stellar Populations (O’Connell 1958). It is common to refer to this component as the “thick disk” after the parlance of Gilmore & Reid, who noted similarities of this Galactic population to thick disks seen in some edge-on spiral galaxies a few years earlier (Burstein 1979, Tsikoudi 1979).

It is convenient to fit these two apparent populations with exponentials, such that
Figure 10. Density distribution for stars in 1.85 deg² towards the south Galactic Pole field SA141 from the author’s CCD starcount collaboration with I.N. Reid, M. Siegel & I. Thompson. Distances are derived from \((R-I)\) colors, which are chosen because of the minimal metallicity effects in the \(M_I(R-I)\) relation. These colors are also ideal for sensitivity to late type stars (although bluer stars are shown in the example here as they probe to greater distances). The fits shown are for an old, thin disk of scaleheight 350 pc (dotted line), an Intermediate Population II/thick disk with scaleheight 1300 pc and 2.2% relative normalization to the thin disk (dashed line), and a halo, in this case modeled with an exponential with scaleheight 3500 pc and 0.15% relative normalization (dot-dash line). The solid line is the sum of the models.

\[
A(Z)/A(0) = \rho_{\text{IP}II}e^{-Z/h_{\text{IP}II}} + (1 - \rho_{\text{IP}II})e^{-Z/h_{\text{thin}}}
\]

where \(A(0)\) is the number of stars seen locally, \(h_{\text{IP}II}\) and \(h_{\text{thin}}\) are the exponential scale-height of the IPII and thin disks respectively, and \(\rho_{\text{IP}II}\) is the local fractional contribution of IPII stars to the old thin disk stars. Note that while exponentials are convenient functions, they are not necessarily physical. Increasingly, researchers are turning to \(sech^2Z\) functional forms, which (1) are not singular at \(Z = 0\), (2) have some physical motivation from dynamical analyses of isothermal disks, and (3) approach the functional form of exponentials at large \(Z\).

Several words of caution are in order. The first relates to the assumption that the counts are not affected by contamination from stars of other luminosity classes. Indeed, the question of contamination by giants was one of the early main criticisms of the photometric parallax count studies that required an extra component, when good fits to starcount data using another approach (discussed below) were found to be adequate without this extra component (Bahcall & Soneira 1980; Figure 11). While the veracity of the two component fit has now been borne out (cf. Casertano et al. 1990), ultimately, contamination by giants does have some effect that, while small in many cases, ought to be properly assessed in all starcount attempts. Second, metallicity variations certainly affect color-magnitude relations, but, it is not typical to have abundance information on a star by star basis. In this case, metallicity corrections may only be applied in a statistical way. Typically, this means assuming a \(d[Fe/H]/dZ\) dependence, which then implies a \(dM/dZ\) dependence. However, this is not a worry-free procedure when attempting to find \(A(Z)\), because the scale of the abundance gradient determines the degree of \(Z\)-compression in stellar densities (as stars are made increasingly subluminous, their derived photometric parallaxes decrease). Alternatively, one may attempt to do starcounts in a bandpass
system that minimizes metallicity effects in the color-magnitude relation (as is done in Figure 10).

Third, the discussion to this point has completely ignored density (and other) variations as a function of other spatial dimensions of the Galaxy. In more general studies, it is common to assume that the disk components have exponential dependences in the Galactocentric radial direction as well. In addition, we have completely left out the important consideration one must give to the Malmquist bias, an effect of the intrinsic absolute magnitude spreads we will encounter in our starcount sample, however restricted in stellar type we attempt to make our sample. A thorough description of the bias is beyond the scope of this lecture, but is discussed in the lectures by Sandage (1995). In brief, if $A(m)$ is increasing, then the Malmquist bias says that the mean absolute magnitude of stars in our sample will be systematically higher than one would expect in a volume limited sample. This occurs because at a given apparent magnitude the volume element occupied by intrinsically brighter stars contributing at that apparent magnitude is larger than the volume element for intrinsically fainter stars contributing at that apparent magnitude. Thus, more intrinsically bright stars can find their way into a specific $A(m)$ bin than can intrinsically fainter stars.

Finally, the overall problem of contamination of star counts by extragalactic objects (QSOs and compact galaxies) becomes a problem at faint magnitudes (see Reid & Majewski 1993).

A second approach to the starcounts problem is to use computers to integrate the von Seeliger equation directly (after assuming functional forms for the luminosity function, density law, and color-magnitude relation for each expected stellar population) and then numerically generate model $A(m,S)$ to compare to data (Bahcall & Soneira 1980, Reid & Majewski 1993). While it is possible to construct luminosity functions of disk stars from analysis of the solar neighborhood, the lower local density of individual stars from other Galactic components make generation of an observed luminosity function much more problematical. However, globular clusters are especially convenient to determining metal-poor luminosity function templates since, with all stars at the same distance, one need only count the numbers of stars as a function of apparent magnitude to get $\Phi(M)$, once the cluster distance is known (see the discussion of cluster luminosity functions by King in this volume). Because of the problem of the degeneracy of the $M(color)$ relation across luminosity classes, it is more physical to replace a single $M(color)$ relation used in combination with $\Phi(M)$, and instead employ a more sophisticated $\Phi(M,color)$ array, called a Hess diagram (see models by Robin & Crézé 1986, Méndez 1995, for example). In general, the Hess diagram varies greatly with age and abundance (and possibly the $r'$ distribution) of the stellar populations. The results of a computer modeling approach to the starcount problem are shown in Figure 9 (right panel).

The lack of a systematic starcount attack along many lines of sight, as envisioned and attempted by Kapteyn and von Seeliger, has not prevented a number of researchers (including the author!) from declaring “best” model parameters for the various Galactic components. Table 1 gives the parameters from a sampling of many published starcount analyses, to give the flavor of the general form and ranges of parameters adopted for density laws, scaleheights, etc. The density law typically adopted for the Galactic halo is based on observations of the spheroids of external galaxies, which seem to show surface brightnesses, $I(R)$ (in solar luminosities pc$^{-2}$), that fall off as a “de Vaucouleurs (1948) $R^{1/4}$ law”:

$$I(R) = I_e 10^{-3.33[(r/r_e)^{1/4} - 1]} = I_e \exp \left\{ -7.669[(r/r_e)^{1/4} - 1] \right\}$$

where $r_e$ is the effective radius, or the radius that encloses one half of the spheroidal
Figure 11. Luminosity functions for clusters of various metallicities. M3 is often adopted as representative of the Galactic halo, while the luminosity function for the intermediate metallicity cluster 47 Tucanae is often adopted for the Intermediate Population II thick disk. The local peak near $M_V = 1$ is due to the horizontal branch. Also shown is the Wielen et al. (1983) luminosity function for disk stars, and the halo luminosity function adopted by Bahcall (1986) and collaborators. Note that the larger number of giants in the latter luminosity function compared to, say, that of the M3 function shown, will imply a larger contribution of halo starcounts at magnitudes and colors normally associated with the Intermediate Population II; thus Bahcall’s model with this halo luminosity function gave reasonable fits to the data without inclusion of an IPII component. From Méndez (1995). See the model of Méndez & van Altena (1996).

$ light, and $I_e$ is the brightness at $r_e$. This equation is derived from noting that for most spheroids, a plot of surface brightness (in mag arcsec$^{-2}$) versus $R^{1/4}$ is fairly linear. Note that with this profile, the peak surface brightness is given by

$$I(0) = 10^{3.33}I_e \sim 2000I_e.$$  

Applying this two-dimensional brightness law of external spheroids as projected on the sky to a three-dimensional density distribution useful for studies in our own galaxy requires the Young (1976) deprojection:

$$\rho(R) = \rho_o \exp[-7.669(R/R_e)^{1/4}]/(R/R_e)^{0.875}.$$  

All studies presented in Table 1 use the latter density law for the halo, and universally accept an effective radius $R_e = 2700$ pc. The local normalization of the halo with respect to the thin disk, $\rho_o$, is typically found to be around 1 star out of 700 in the solar neighborhood.

An alternative density law often used for spheroidal components is the power law relation, given by
Table 1. A sampling of starcount models from the literature, showing the relative density normalization locally and axial ratio of the halo; the local normalization, scaleheight and color-magnitude diagram (where “47 Tuc” is from the cluster, “Mid” is an intermediate aged CMD, and “OD” is an old disk CMD) used for the IPII thick disk; and the scaleheight of the thin disk broken into bins by luminosity, as $M_V < 9$ and $M_V \geq 9$, respectively, to account for likely differences in mean age. The “standard” model is based on a review of literature by Reid & Majewski (1993).

\[
\rho(R) = \rho_0 \left( \frac{a_0^n + R_0^n}{a_0^n + R^n} \right)
\]

where $R_0$ is the Galactocentric radius of the Sun, $a_0$ is a core radius (typically about 1 kpc for the halo), and $\rho_0$ is as before. The power, $n$, of the power law is usually found to fit well the distribution of halo tracers, like RR Lyrae stars and blue horizontal branch stars, when $3 < n < 4$. For the inner bulge (< 1 kpc), a power of something like $n \sim 1.8$ seems to apply.

Whether a power law or de Vaucouleurs law is adopted, it is common to account for the fact that the halo of the Galaxy (and other spheroids) is not perfectly spherical, but show some flattening with minor to major axis ratio ($c/a$). Thus, a correction to the density law is applied in the $Z$-direction as

\[
Z \rightarrow (c/a)Z.
\]

Typical flattenings for the halo are found to be something like $(c/a) \sim 0.8$.

The examples of starcount models shown in Table 1 show relatively good agreement, but this is partly a result of some data sets being in common. It is important to note that several studies (two of them are shown in Table 1) indicate the need for two separate halo components. These dual halo models are discussed at greater length in Section 3.4.2.

Unfortunately, the general agreement of the starcount models breaks down with consideration of the disk. A contentious issue is the density law of the IPII, with a variety of relative normalizations and exponential scaleheights derived using a large range of possible tracers and starcount analyses (Figure 12). Part of the discrepancy derives from the problem that the IPII size is intermediate between that of the thin disk and the halo: While the thin disk and halo dominate the starcounts at bright and faint magnitudes, respectively, the IPII-dominant regime overlaps considerably with each of these two other populations, and this makes it difficult to “extract” from the mix (see model example in Figure 9). Moreover, fitting the exponential density law described above to stars predominantly in the $Z$-distance regime that the IPII dominates (wherever that may be) apparently yields some amount of degeneracy, in that the values of $\rho_{\text{IPII}}$ and
$h_{\text{IPII}}$ tend to play off of each other. (This seems especially a problem when starcounts have a rather bright magnitude limit that restricts the range of accessible extent of IPII dominance.) Figure 12 demonstrates the range of scaleheights and normalizations derived for the IPII from (top panel) starcount models and photometric starcount analyses, and (bottom panel) various objects thought to trace the IPII population. Also shown are lines indicating the surface mass density (mass per pc$^2$) of the IPII (as projected on the Galactic plane) relative to that of the thin disk, as a function of the density law parameters. As can be seen, even though the derived normalization and scaleheight ranges are large (an order of magnitude for the normalization), the derived mass density of the IPII is rather more constrained to be within $5 - 20\%$ that of the thin disk, and generally clustered around $10\%$. In summary, while there is agreement that by surface mass the IPII clearly represents a significant component of the Galaxy (by comparison, the mass of the halo is another factor of 10 less), it is not clear how this mass is distributed (how “thick” is the thick disk?).

In my opinion, the importance of solving the problem of how the IPII mass is distributed cannot be understated, especially when the appreciable kinematical and chemical overlap of the IPII with other populations is considered. Figure 13, which shows the relative fraction and actual count contribution of the IPII as a function of height above the Galactic plane for a variety of IPII density laws against constant halo and thin disk density distributions, demonstrates the essence of the problem. If one wants to select stars to characterize the IPII for, say, an assessment of its chemical and kinematical properties, from where does one take the representative sample? Without knowledge of the density law, it is not clear at what heights the IPII dominates. Just as troubling is the problem that for practically none of the examples shown in Figure 13 does the IPII suffer minimal contamination of stars from either the thin disk or halo (in contrast to the situation for either of the latter two components). Unless one is able to find some tracer object that is assuredly only found in the IPII, one is forced to make certain assumptions about the properties of the IPII (and the other populations) in order to sort its stars out. A certain amount of circularity then follows: E.g., if one sorts stars on the basis of abundance, then one may not address abundance questions but might be able to address kinematics, as long as there are no correlations between abundance and kinematics. However, it is often the case that such correlations do exist. Moreover, the distributions of abundance and kinematics of the halo and disk are known to have large dispersions that likely overlap the properties of the IPII significantly.

Figure 13 also illustrates how easy it is to generate artificial gradients in properties when they may not exist. Imagine, for example, single-valued parameters (e.g., [Fe/H], rotational velocity) attributed to each the thin disk, IPII and halo. Without knowing a priori the population assignments of any particular star, a tally of the mean observed value of this parameter as a function of height above the Galactic plane will yield relatively smooth gradients. A more complex, and even more difficult to deconvolve, example would be if some or all of the populations themselves showed gradients as a function of $Z$. Finally, with the ability to generate parameter trends and gradients with the combinations of homogeneous populations (see Section 1.1), how can one be assured of even the number of discrete components that might be required (cf. Lindblad 1927)? The issue of a discrete versus continuous disk/IPII is addressed again in Section 3.3.3.

Thus we find ourselves in a highly disagreeable situation. Separation of populations in the range $1 < Z < 10$ kpc is generally an extremely difficult, and risky, proposition. It is common practice to make certain assumptions about the stars in populations that overlap in order to ascertain certain other properties; but it is appropriate always to bear in mind that in science it is often the case that initial predilections drive an experiment
Figure 12. Intermediate Population II/thick disk scaleheights and local normalizations as a fraction of the thin disk population, as derived from a number of starcount analyses (top panel) and for various types of tracer objects (bottom panel). Dotted lines show loci of IPII mass density relative to the thin disk, when a scaleheight of 325 pc is adopted for the latter.

to reinforce those predilections, whether they are correct or not. This conundrum has a certain resonance with the state of the starcounting endeavor before von Seeliger’s great insight and statistical prowess were brought to bear on that problem. Overcoming the present impasse may require a similar advance in the level of statistical sophistication. Progress in this direction may well be in the direction of multidimensional, univariate mixture models discussed by Nemec & Nemec (1991, 1993) and others.

2.3. The Role of Globular Clusters in Assessing Spatial Structure

One of the first great successes in the use of globular clusters for stellar population studies was as a probe of the size and shape of the Milky Way by Harlow Shapley. Shapley’s work altered the Kapteyn and von Seeliger starcount view of the “universe” which, in spite of its great successes, still found the Sun to be near the center of the stars in the Milky Way system. Shapley, like John Herschel before him, noted an excess of clusters toward
the direction of the sky near Sagittarius. With the assumption that the globular clusters traced the true shape and extent of the Galaxy, Shapley suggested that the center of the Galaxy was in the direction of the apparent center of the globular cluster system (Figure 14).

Globular clusters were also central to what has now become known as the “Great Debate” between Heber Curtis and Harlow Shapley before the National Academy of Sciences (Washington, D.C.) in 1920 (Curtis 1921, Shapley 1921). The center of the debate, “The Scale of the Universe”, turned in large measure on the question of the distances to globular clusters, as each protagonist identified the globular cluster system as a part of the Milky Way with remote members that defined its extent.

The outcome of the debate is commonly synopsized as “Curtis was right for the wrong reasons and Shapley was wrong for the right reasons”. Shapley’s position was that the Milky Way’s diameter was likely to be at least 300,000 light years (90 kpcs), and that the Sun was some 50,000 light years (15 kpc) from the center (modern studies are obtaining a solar Galactocentric distance half of this; cf. Olling & Merrifield 1998). At the time, this was considered an immensely vast scale, and it seemed likely that the resultant volume defined by such a Galaxy contained all objects in the observed universe. It should be noted that the debate predated the revolution in understanding of “spiral nebulae” brought on in the next decade by groundbreaking work in the decade before. This included Slipher’s work (beginning in 1912, cf. Slipher 1913) at Lowell Observatory to measure the redshifts of the spiral nebulae and Hubble’s discovery of Cepheid variables in them at Mt. Wilson. Indeed, van Maanen’s (1916, et seq.) claims for astrometrically measurable rotations in spiral nebulae (later shown by Hubble [1935] to be artifacts of systematic error) greatly influenced many astronomers, including Shapley, in their belief that the spiral nebulae were much more local, star forming clouds, rather than “island
Figure 14. Shapley, working at the Mt. Wilson Observatory, was involved in a survey to estimate distances to globular clusters. These results, and those of others, he compiled into this representation of the cluster system projected on the $XZ$-plane, where $X$ is the line containing the solar neighborhood and the Galactic center, and $Z$ is perpendicular to the Galactic plane. From this distribution, he estimated both the size of the Milky Way and the distance of the Sun from the Galactic center. The units shown are 100 pc, as measured from the Sun (location of the "$x$"). From Shapley (1918).

universes" like the Milky Way, as postulated (with little physical foundation) long before by Thomas Wright, Heinrich Lambert and Immanuel Kant. As one line of support that his measurements of rotation in the spiral nebulae were real, van Maanen pointed out the contrast of the magnitude of his spiral arm proper motions to his finding (van Maanen 1925, 1927) of minute internal motions in globular clusters!†

In hindsight, Shapley's major "failing" in the debate was in assuming (quite understandably, even as admitted by Curtis, who was generally more cautious) that the Cepheids and blue stars in clusters were similar to those found near the Sun; thus Shapley overestimated distances in his cluster system. The problem, of course, is that he confused much more luminous, Population I, main sequence B stars with fainter horizontal branch stars. Moreover, in the case of the Cepheids, Shapley could hardly have known (although he acknowledged the possibility and Curtis certainly highlighted it) that Population I Cepheids (the classical Cepheids, which are high mass supergiant stars) near the Sun, and Population II Cepheids (low mass, metal-poor W Virginis stars), the type found in globular clusters, follow different period-luminosity relations. The W Virginis stars are

† To be fair, van Maanen's claims for rotational motion in spiral nebulae were not inconsistent with a number of related findings (including Slipher's [1914] and Wolf's [1914] findings of internal motions via radial velocities) and theories by predecessors, nor even similar measurements by other astronomers in what was a rather active field of endeavor (see Berendzen & Hart 1973, Heatherington 1975). However, van Maanen, by his reputation as a meticulous observer, may well have been the most trusted of the astrometrists working on the problem. He was at least the most vocal and most published in this research field. Surprisingly, Curtis' (1915) own work in this area on some of the same plate material yielded no detectable motions in the spiral nebulae, but Curtis' results seem largely to have been ignored.
fainter than the classical Cepheid prototypes adopted as standard candles by Shapley. Thus, while Shapley’s ultimate estimate for the size of the Milky Way as defined by clusters, i.e. probably greater than about 100 kpc, is approximately correct, his distance scale for individual clusters was exaggerated. The new outer cluster limit of about 100 kpc comes mainly from globular clusters found after Shapley’s work (a number of them, the “Pal” clusters, were found during the first Palomar Observatory Sky Survey). A modern view of the distribution of the Milky Way globulars is shown in Figure 15.

In contrast to Shapley, Curtis maintained allegiance to the island universe theory. This was motivated, in part, by his own observations of globular clusters and analysis of their distance, which severely underestimated the size of the Galaxy as being at most 30,000 light years (9 kpc) in diameter, a size that could easily exclude spiral nebulae by almost any accounting of their distance. The basis of Curtis’s underestimate of distance was that he had assumed, on the basis that the mean spectral type of globular clusters looks like that of the Sun, that the average stars seen in clusters were dwarfs of luminosity comparable to the Sun. As well argued by Shapley, Curtis ignored a serious systematic effect: “in a distant external system we naturally first observe its giant stars ... the comparison of averages means practically nothing because of the obvious and vital selection of brighter stars in the cluster”. In the end, the globular cluster distance scale is now established between the scales of Shapley and Curtis (see discussions by Feast and by King in this volume).

2.4. Modern Descriptions of the Galactic Globular Cluster System

In his contribution to this Winter School, Harris discusses the presently known spatial, chemical and kinematical distributions of the Milky Way globular cluster system,
Figure 16. (a) The familiar division of globular clusters into disk and halo systems, adapted from Figure 1 of Zinn (1985), and including his metallicities and \(Z\) distances. Panels (b) and (c) are discussed in Section 3.4.1. (b) Calculated orbital \(Z_{\text{max}}\) from Luis Aguilar (UNAM) for those old halo and disk globulars having proper motions (see Table 2 of Majewski 1994a, Dinescu 1998) and connected by vertical lines to the present \(Z\) distances for these systems. In many cases, \(Z \sim Z_{\text{max}}\) so that the pairs of points overlay one another. (c) Data points show the RHB “young halo” (open circles), BHB “old halo” (filled circles) and disk (solid squares) globular clusters, with updated \(Z\) and [Fe/H] taken from Harris (1996). Dashed lines show the range of metallicity spanned by the RHB “young halo” globulars, while the solid, curved line shows, schematically, a cluster paradigm (Zinn 1993a) in which the BHB “old halo” and disk globulars represent one system, perhaps connected through a dissipational collapse of the disk.

based on his very useful compilation of cluster data (Harris 1996, also available on-line at http://physun.physics.mcmaster.ca/Globular.html). Therefore, I briefly concentrate here (and in later sections) on only those aspects of the distributions that are germane to this discussion, and some for which I present a different interpretation than Harris.

Studies of the globular cluster system in the past few decades have made evident the existence of at least two subsystems. A disk system of globulars (Zinn 1985, Armandroff 1989) is found concentrated towards the plane of the Galaxy (the darkened patch of points in Figure 15), while the halo system of globulars is represented by the more extended distribution of points in Figure 15. A conventional description of the two systems is given by the [Fe/H]-|\(Z\)| distribution, as presented by Zinn (1985), reproduced here in Figure 16(a). Zinn recommended a division between the disk and halo systems at [Fe/H]=−0.8 (and a division near −1.0 is commonly used today), so that, by definition, the disk globulars are more metal rich than those in the halo. The range of distances from the Galactic plane, |\(Z\)|, is much wider for the metal-poor, “halo” globulars than the metal-rich, “disk” clusters. More recently, Zinn (1993a) has proposed a different division of the globular cluster system into three components after taking into account not only the metallicities of the clusters, but the character of their horizontal branches. This new description is addressed in Section 3.4.1.

2.5. Pushing the Envelope

The presently known halo globular clusters show a rather spherical density distribution, with a majority of clusters within 30 kpc of the Galactic center (Figure 15). After this radius, the density apparently drops, so much so that the region between about 40 < \(R_{\text{GC}}\) < 80 kpc has been characterized as a globular cluster “gap” between the inner
and outer halo systems (there being one known cluster in the gap, according to Harris’ 1996 compilation). The five outer (past the gap) globulars inhabit $80 < R_{GC} < 120$ kpc, with the most distant known outlier, AM-1, at $R_{GC} = 120$ kpc.

By the earlier definition of the extent of the Galaxy as given by globular clusters, this would define the limit of the Milky Way. However, beginning in the 1930s with discoveries of the Fornax and Sculptor systems by Shapley, a number of dwarf satellite galaxies of the Milky Way have been found. The most recently found dwarf satellite is, ironically, also the closest one – Sagittarius, which is at $R_{GC} = 16$ kpc (Ibata et al. 1995). The present count of satellite galaxies of the Milky Way is eleven, with Leo I the most distant at 280 kpc from the Galactic center (Lee et al. 1993). The recently discovered (van de Rydt et al. 1991) Phoenix system is at $R_{GC} = 417$ kpc, but it is not yet certain whether this system is bound to the Milky Way. Among the satellite galaxies, the most massive members, the Large and Small Magellanic Clouds, Fornax, and Sagittarius all have globular cluster systems – an important clue to the formation of some of the Milky Way globulars (see Section 3.4.1 below). Note that the globulars of Fornax, located at $R_{GC} = 143$ kpc, are technically the most distant globular clusters associated with the Milky Way system. Depending on whether or not Phoenix is included, the known extent of the Milky Way system is therefore established to be at least 280 kpc, but perhaps more than 400 kpc, in radius. It is not unreasonable to expect that more dwarf satellites of the Milky Way may remain to be found. Because of their very low surface brightnesses, dwarf galaxies are hard to spot, but several groups are now using sophisticated algorithms to search for faint brightness enhancements in scans of POSS-II plates and ESO-SERC plates and other wide field data. This work has activated a new round of discoveries of Local Group dwarfs (e.g., the Antlia dwarf, Cas dSph, And V and And VI (=Peg dSph); Whiting et al. 1997, Armandroff et al. 1998, Jacoby et al. 1998, Karachentsev & Karachentseva 1999).

Note that the $R_{GC}$ of these outer Galactic satellites are in the same regime as the extent of Lyman $\alpha$ absorbers found in QSO absorption line studies (Lanzetta et al. 1995). If the Milky Way extends to $> 300 – 400$ kpc in distance, and M31 and M33 have similarly extended outer parts, then we begin to reach the regime where the outer parts of these individual systems begin to overlap. Considering that the dark matter halos of galaxies are likely to be more extended than the luminous tracers, it may be that the primary galaxies of the Local Group are not really isolated, but, at the least, floating in a shared dark matter soup.

What about the extent of the Milky Way as gauged by individual field stars? This is not well studied and may be difficult to establish without having full velocity information to check whether outliers are bound to the Milky Way. The possibility that some dwarf satellites are tidally stripped of stellar and cluster débris (see Section 3.4.3 below) begs the question of whether there may be extremely distant, now isolated stars bound not to the Milky Way or other galaxies, but to the Local Group. Stars released on hyperbolic orbits are a characteristic of galaxy interactions (Toomre & Toomre 1972) and could lead, perhaps, to “intergalactic tramps”. Intergalactic planetary nebulae have been identified in the Fornax (Theuns & Warren 1997) and Virgo (Méndez et al. 1997) clusters, but very little work has been done on finding very distant stars bound to the Milky Way, let alone “extra-Galactic” stars in the Local Group. The latter would be especially useful for gauging the extent of the Local Group dark matter. Gould et al. (1992) report finding dwarf stars at a distance of about 100 kpc, near the Sextans dwarf galaxy, while the author and collaborators have recently found giant stars more distant than 140 kpc. Richstone et al. (1992) sought distant main sequence turn off (MSTO) stars in small area ($0.002$ deg$^2$), but very deep ($B \approx 27$) CCD frames that reach potentially to 500 kpc.
for stars at main sequence luminosities. While Richstone et al. found no excess signal above the expectations for contamination of their sample from compact galaxies, from statistical arguments these authors place a lower limit on \((M/L)_{\text{Local Group}}\) of 400.

3. Survey of Age, Kinematical and Chemical Distributions in Stellar Populations

Our goal is to recognize chemodynamical patterns in stellar populations that will give some idea of the progression of events during the evolution of the Galaxy, so that we may synthesize models to explain the origin and distribution of stellar populations in the Milky Way, and, by analogy, similar spiral type galaxies. In the interests of the limited time in these lectures, and influenced by the predilections and expertise of the author, here the focus will be on the observer’s role of seeking patterns among age, chemistry and kinematics, rather than the host of chemodynamical modeling efforts being conducted at a number of centers around the world.

A great deal of work has been done in the last decade or so to establish the chemodynamical and age characteristics of the Galactic stellar populations. My approach to covering this wealth of material is to survey various approaches utilized to establish population properties and summarize results of these approaches in an attempt to assemble a first-order picture (via chemical and dynamical “population box” representations) of the evolution of the Galaxy. A number of the observational techniques, while discussed specifically in certain contexts, lend themselves to broader use in population studies.

3.1. Brief History of Kinematical and Chemical Studies of the Milky Way

The modern era of Galactic structure studies may be identified with a shift in preoccupation from the simple clarification of the order and shape of stellar systems within the Galaxy to the application of the properties of these systems toward ascertaining the evolutionary history of the Milky Way. This change in mindset was technically feasible only after great theoretical and observational advances in the 1950s that made possible the joining of Galactic structure, kinematics, chemistry, and age into a unified evolutionary context. The development of a theory of stellar evolution not only made possible a means for dating the ages of stellar clusters through use of the color-magnitude diagram (Sandage & Schwarzschild 1952) but also led to an understanding of the process of nucleosynthesis (Burbidge et al. 1957) and the idea of the progressive chemical enrichment of the Galaxy. Once it was appreciated that stellar spectra held key information on the level of stellar enrichment via the strength of absorption lines (Chamberlain & Aller 1951), classification of individual field stars into relative age groups became possible. These absorption lines are not evenly distributed across the visible spectrum, and it was found that some broad band filters are particularly sensitive to variations in abundance. For example, weak-lined stars show an ultraviolet excess of flux (normally measured as a particularly blue \(U - B\) color for stars of a given \(B - V\) color, Wildey et al. 1962) due to suppression of line absorption normally concentrated in the near ultraviolet. The relative age scale of field stars could be calibrated to that of star clusters, once it had been shown (Sandage & Walker 1955) via ultraviolet excess measurements that the peculiar, weak-lined spectra exhibited by some stars were shared by globular cluster stars.

Decades earlier, through the work of Kapteyn, Stromberg, Oort, Lindblad, and others, great advances had been made in the understanding of stellar kinematics and in the identification of kinematic subsystems within the Galaxy. Through Baade’s (1944) insight, the association of these kinematic groups with specific structural components in galaxies was made: Population I were the metal-rich stars in disks moving with rotational veloci-
ties like the Sun, while Population II objects Baade associated with the spheroidal parts of galaxies (bulges and halos), which he thought to be primarily metal-poor. The final link between the structural-kinematic stellar populations and the age-metallicity groups came with the discovery by Roman (1954) that metal deficiency in stars is typically correlated with high velocities with respect to the Sun, while higher metallicity stars move with velocities similar to the Sun and other disk stars.

With loose connections established between kinematical, chemical, age, and structural groupings of stars within the Galaxy, the stage was set for the development of the modern evolutionary pictures that accounted for the existence and properties of the various stellar populations. The first major breakthrough along these lines, the landmark paper by Eggen, Lynden-Bell & Sandage (1962, ELS hereafter), remains the foundation for modern discussions of Galaxy formation. It also set the precedent for the construction of modern Galactic structure surveys. ELS compiled ultraviolet excesses, radial velocities, and proper motions for nearby stars, which ELS selected from a catalogue of large proper motion stars (i.e., typically stars with large velocities with respect to the solar neighborhood) combined with a catalogue of well studied, bright stars without kinematical biases (the latter sample would be dominated by the overwhelming number of stars in the local disk population). From their combined sample, ELS discovered smooth correlations between ultraviolet excess and (a) the orbital eccentricity of the stars, (b) their angular momentum, and (c) the $|W|$ velocity (the velocity perpendicular to the Galactic disk).

Based on these correlations ELS constructed a formation model that incorporated the new age-dating techniques from stellar evolutionary theory as well as a dynamical analysis of stellar orbits. In this picture, the Galaxy formed from the rapid collapse out of the general Hubble expansion of a metal-poor, roughly spherical, primordial density fluctuation. During this collapse, condensations of gas were created and star formation turned on. The orbits of these stars today were presumed to reflect the kinematical state of the gas from which they formed, while the stellar abundances were a function of the level of chemical enrichment of that gas (via the fusion of “metals” in the cores of massive stars which belched forth these processed elements into the interstellar medium in supernova explosions). Thus, the earliest stars formed had the weakest metal abundances, and were born into highly radial orbits with the momentum of the initial collapse. Radial collapse was unable to continue as far as that in the direction perpendicular to the Galactic plane because of the increase in rotational velocity and centrifugal acceleration due to angular momentum conservation. Dissipation of energy via cloud-cloud collisions enabled the gas eventually to settle into a flattened, rapidly-rotating disk. Continued star formation during the contraction allowed progressive enrichment of the interstellar medium. It is from the flattened, metal-rich disk of gas that new, circularly-orbiting stars are forming today (Figure 10 shows the relative flattenings of various Galactic components as viewed near the Sun).

Based on the apparent age of the globular clusters, which were attributed as members of the first formed stellar population, ELS assumed the earliest star formation to have started 10 Gyr ago. From the $|W|$ velocities, the maximum distance from the Galactic plane, $Z_{\text{max}}$, can be calculated under an assumed Galactic potential. From the range in $Z_{\text{max}}$, ELS estimated the vertical collapse to have been a factor of 25. From the

† The true age of the globular clusters is still much debated, and derived values have encompassed a range up to a factor of two higher than ELS assumed (Chaboyer et al. 1996). However, some recent age determinations have come almost back down to the ELS value, near 10–12 Gyr (e.g., Salaris & Weiss 1997, Reid & Gizis 1998). Throughout the present contribution I’ve used an age scale for the Galaxy and globular clusters about in the middle of the range typically discussed.

†
ratio of apogalactica of high eccentricity to circular orbits for stars with similar angular momenta, the radial collapse was estimated to have been a factor of 10. Finally, the rate of the collapse was determined from the dynamical constraint that in order to form the old stars (i.e. the low metallicity stars) on highly radial (i.e. high eccentricity) orbits, the radial velocity of the initially collapsing gas from which they formed must have been less than the rotational velocity. In other words, the rate of the collapse must have been more rapid than the Galactic rotation period, or \( \sim 0.2 \) Gyr. Because this is approximately the timescale for freefall collapse from the estimated initial gas distances, ELS concluded that the halo of the Galaxy formed during a collapse without pressure support. As support for this idea, ELS noted: (a) if the gas had been hot enough to provide pressure support to slow the collapse, it would have also prevented the small-scale collapses which form stars, and (b) the five halo globular clusters with accurate photometry at that time all had nearly the same age, which suggests a brief formation epoch for the halo globular system.

While the following decades brought criticisms and refinements of the ELS model, and the rise of a competing picture for the formation of at least some part of the halo (see Section 3.4), the ELS picture incorporated the basic elements that are the foundation of modern chemodynamical models, and the general ideas they presented are likely relevant for at least some part of the formation of the Galaxy (see Section 3.3.3, for example). It has been a primary endeavor of Galactic astronomy since ELS to refine and elaborate on their themes with ever improving data to constrain Galactic formation models.

### 3.2. The Thin Disk

#### 3.2.1. Age and Star Formation History of the Disk

A number of means have been used to determine the distribution of ages of stars in the disk of the Milky Way. In this subsection I concentrate on age dating the Baade Population I thin disk, but remind the reader that the problem of truly separating tracers of the thin disk from those of the IP II – indeed, whether this is even possible if they form a single, contiguous population – remains a nagging concern for understanding the elder part of the thin disk. Because of remaining uncertainties in the age scale, I will try whenever possible, to discuss ages in relative terms; whenever absolute values for ages are given, the reader should bear in mind that the absolute age scale still varies among researchers by up to a factor of two.

Traditionally, a maximum age of the thin disk of about 10 Gyr was adopted from an estimated age (now questioned) of the open cluster NGC 188 (VandenBerg 1985). In the last decade, a number of new techniques have been used to age date the disk, with results from slightly less than 10 Gyr to several Gyr more:

- **Nucleocosmochronology** utilizes measures of ratios of long-lived isotopes to estimate ages. To first order, when a star is formed, its outer atmosphere provides an unchanging sample of the composition of the Galactic gas at that time, modified only by the decay of radioactive species. For example, one may use as a chronometer variations in the line strengths of the radioactive thorium isotope \(^{232}\)Th, which has a half-life of 20 Gyr (i.e., close to a Hubble time), to that of a stable element, like Nd (which has an absorption line very near the Th line at 4019 Å, a useful coincidence for high resolution spectroscopic studies). Butcher (1987) attempted the experiment with nearby G dwarfs of very different assumed ages (postulated using a variety of means) and found no variation in \(^{232}\)Th/Nd over the entire range of ages. This means that either (1) all of the \(^{232}\)Th was formed in a single event before all of the stars were formed, (2) all of the stars were in actuality formed only a short while ago, so that no significant decay has occurred, or (3) the synthesis
of $^{232}\text{Th}/\text{Nd}$ is increasing fractionally with time. From measures of other transuranic elements in meteorites, we know that the duration of nucleosynthesis before formation of the solar system (itself at least 4.5 Gyr old) is about $5.4\pm1.5$ Gyr (Fowler 1987), so (2) is excluded, but so too is (1) if the solar system measure gives a maximum age. Analysis of option (3) yields a maximum age for the disk of about 10 Gyr. Unfortunately, this type of analysis is subject to a number of uncertainties, especially over the amount of infall of gas to the disk from outside – something sure to affect isotope ratios.

• The **white dwarf luminosity function** provides a useful limit on the age of the Galactic disk, because, for all but the youngest populations, stars evolving off the main sequence have masses ($<8M_\odot$) that result in the formation of white dwarfs as their final evolutionary stage. White dwarfs are essentially cooling embers, for which the thermodynamical physics, in principle, can be estimated. Since the age of the disk is finite, the allowed cooling time for the oldest stars is constrained, and this should lead to a sudden cutoff at the faint end of the luminosity function corresponding to the limit of white dwarf cooling in this time (D’Antona & Mazzitelli 1978). Indeed, such a cutoff was found by Liebert et al. (1979); a recent compilation of the data is presented by Isern et al. (1997; Figure 17). A significant amount of theoretical work has been devoted to understanding what the cutoff means in terms of age limits.

Modeling the luminosity function requires an understanding of the various stages of white dwarf cooling, and the details of each type of cooling dependent on the white dwarf mass and chemical composition (see Isern et al. 1997 for a more complete description): (1) *neutrino cooling* from p–p chains in the hydrogen layer; (2) *fluid cooling* by gravothermal shrinking; (3) *crystallization*, which takes advantage of two new sources of energy, the latent heat which makes a small (5%) contribution to the luminosity of the star, and the sedimentation of heavy elements towards the center of the star, which results in the release of gravitational energy; and (4) when the star is almost completely solidified,
Figure 18. Disk star formation rate given by stellar chromospheric ages, as corrected for sampling effects by Rocha-Pinto et al. (1998). The various star formation maxima, labeled as "bursts A", "B", and "C", correspond roughly to those found by Barry (1988) with the same technique, and roughly correspond to the age distribution by Twarog (1980) from Strömgren photometry of disk stars. The data are incomplete at the oldest ages, where there may have been an additional burst of star formation.

Debye cooling takes over. With a model for cooling time that incorporates these various stages, one may estimate the expected luminosity function of white dwarfs from

\[ n(\log (L/L_o)) = \int_{M_{\text{min}}}^{M_{\text{max}}} \Phi(M) \Psi(T - t_{\text{cool}} - t_{\text{MS}}) \tau_{\text{cool}} dM \]

where \( M \) is the main sequence mass of the white dwarf progenitor stars which have a luminosity function \( \Phi(M) \). \( \tau_{\text{cool}} = dt/dM_{\text{bol}} \) is the characteristic cooling time at each white dwarf bolometric luminosity, \( M_{\text{min}} \) and \( M_{\text{max}} \) give the range of masses of main sequence stars able to produce white dwarfs of luminosity \( \log (L/L_o) \), \( t_{\text{cool}} \) is the time required to cool to this luminosity, \( t_{\text{MS}} \) is the main sequence lifetime of a star of mass \( M \), and \( T \) is the age of the white dwarf population.

Note that because the observed luminosity function is also dependent on the star formation rate, \( \Psi(t) \), which is creating stars according to the luminosity function \( \Phi(M) \) at any given time \( t \), the shape of the white dwarf luminosity function can also give information about the history of star formation. According to Isern et al. (1997), the best fitting star formation rate to the disk white dwarf luminosity function shows an age of 12 Gyr for the first star formation (but at a relatively low rate) followed by a rise in activity after 1 – 3 Gyr and reaching a peak 4 – 6 Gyr after that. The star formation rate has remained constant or slowly declined since this time. Compare this disk star formation history with that presented next (Figure 18). Note that the limit to the disk age as set by this technique is still controversial at the level of about 20%, partly due to the data (which are heavily dependent on only a small number of known white dwarfs at the turnover point of the luminosity function; e.g., Wood & Oswalt 1998), and partly to differences in the adopted cooling rates and white dwarf core compositions (yielding some age estimates as low as 8 ± 1.5 Gyr; Leggett et al. 1998).
Stellar chromospheric activity declines with stellar age, and the amount of activity can be measured via the strength of the CaII H and K lines in F and G dwarf stars (Soderblom 1985, Barry 1988, Henry et al. 1996†). Results for a sample of 730 stars selected from the literature and recalibrated with a new calibration of the chromospheric line strength-age relation are given by Rocha-Pinto et al. (1998, Figure 18). Interestingly, the results of this work imply a rather variable star formation history for the Galactic disk, with a variation of perhaps more than an order of magnitude in the star formation rate. For example, a rather strong SFR is indicated at intermediate ages, and a rather low, but non-negligible, SFR is indicated some $10 - 12$ Gyr ago, consistent with the white dwarf luminosity function results. A major decline in the SFR several Gyr ago can also be seen and has been long established (Vaughan & Preston 1980, Henry et al. 1996). A somewhat “bursty” star formation history for the disk begs the question of possible triggers to set off the cycles. One trigger that has been studied several times relates to the passage of the Magellanic Clouds through the disk; however, recent work to check the timing of the Magellanic orbit (still somewhat uncertain, of course) against the timing of the disk star formation rate maxima yields no satisfactory connection (Rocha-Pinto et al. 1998).

Strömgren photometry is a very useful technique for age-dating stars, as the intermediate band filters of the Strömgren system are extremely sensitive to changes due to the evolution of stars away from the zero-age main sequence. A $uvby\beta$ photometric study by Twarog (1980) shows the same increase in the number of stars of age $\sim 4$ Gyr old (Meusinger 1991) indicated by the chromospheric age data (those shown as “burst B” in Figure 18). The highest age found among the stars in the Edvardsson et al. (1993) Strömgren photometry and spectroscopic sample is 12 Gyr.

The red edge of the red HB clump is well defined for MSTO stars with $0.8 M_\odot < M < 1.3 M_\odot$ (i.e., ages between 2 and 16 Gyr) and is only a function of metallicity. Using a CMD of disk stars from HIPPARCOS, Jimenez et al. (1998) were able to identify the abundance of the most metal-rich red clump stars as no more than $[\text{Fe/H}] = +0.3$, and from this assessment an appropriate isochrone can be selected to match the red envelope of HIPPARCOS subgiants. Under the conservative assumption that the most metal-rich subgiants have $[\text{Fe/H}] = +0.3$, Jimenez et al. derive a minimum age for the disk of 8 Gyr. However, it is unlikely that the oldest disk stars are this metal rich, and therefore the age of the red envelope subgiants is likely to be higher than this limit.

Open clusters are commonly regarded as a purely thin disk population, especially as the lifetime of the typical open cluster was thought to be relatively brief, $\sim 200$ Myr or less (Janes et al. 1988). These notions have no doubt been influenced by the concentration of work on relatively nearby examples. Indeed, of the some 1200 open clusters known (Lynga 1987), less than one quarter have received more than superficial attention, and fewer still have quality CMDs. However, new surveys (e.g., Phelps et al. 1994) of open clusters reveal a substantial number of very old open clusters (age $\geq 5$ Gyr), and the age of the oldest known open cluster, Be 17, suggests that the age of the disk is at least $\sim 12$ Gyr old (Kaháň 1997, Phelps 1997). That long-lived population open clusters exist is unexpected given ideas of cluster dynamics (e.g., Wielen 1977), which predict that clusters moving near the Galactic plane will have short lifetimes. What seems to permit the long-lived open cluster population to exist is that they are located mostly in the outer disk, where disruptive encounters with molecular clouds (e.g., Spitzer 1958) are less frequent. Moreover, from their observed distances from the plane, many old open clusters

† Henry et al. (1996) note the possibility of some age ambiguity depending on what part of the “solar minimum-maximum” cycle a star is in when observed.
clusters must have $Z$-motions that take them away from the plane for much of their orbit, so that their survivability is greatly enhanced relative to the rest of the open clusters. Curiously, some of these old open clusters are several kpc away from the plane — too far to have been “heated” to this distance by secular dynamical processes. These clusters had to have been formed far from the Galactic midplane — which would seem to be an important clue to the history of the old disk.

Before the recent work to uncover new examples of ancient open clusters, it was commonly believed that there was a fairly long pause between the epochs of globular cluster and open cluster formation. However, incorporating the recent discoveries into a comparison of the relative ages of the open clusters and globular clusters as determined in a homogeneous way (using a “morphological age index” in the color-magnitude diagram of the cluster stars; Phelps et al. 1994) reveals that a gap no longer exists (Figure 19). Moreover, the ages of the oldest known open clusters are now beginning to overlap the ages of the youngest globular clusters. Thus, the disk open cluster system appears to have begun formation even while the halo globular cluster system (which contains the young globular clusters shown in the Figure) was still forming. While the age distribution shown for the open clusters in Figure 19 is influenced heavily by selection effects (research emphasis has been predominantly on the youngest and oldest open clusters), it is interesting to note that present population of Galactic open clusters appears to have formed over the lifetime of the disk.
Curiously, while the oldest open clusters seem to be located outside the solar circle, with no age gradient found for radii larger than the solar circle, this is in direct contradiction to the findings for solitary disk stars. Edvardsson et al. (1993) conclude from their well studied sample of F and G dwarfs that the disk formed “inside out”, and question whether stars more than $10-12$ Gyr formed in the disk at the radius of the solar circle. From the radial dependence of the distribution of $\alpha$/Fe ratios to [Fe/H] (see below), which is a function of the enrichment history of the gas, and, in turn, the star formation history of the disk, Edvardsson et al. claim that star formation in the disk proceeded quickly at early times near the center of the Galaxy, while star formation was increasingly drawn out as a function of radius. This “inside-out” formation is a characteristic of recent chemodynamical models of the disk (e.g., Burkert et al. 1992, Chiappini et al. 1997).

Perhaps the resolution of this apparent contradiction lies in the discussion of cluster survivability in Section 1.2 above. In the inner parts of the disk, clusters are subject to far more disruption; thus we would expect a larger contribution of old cluster stars to the field at smaller radii than we would farther out in the disk.

### 3.2.2. The Age-Metallicity Relation

As discussed in Section 3.1, historically it was believed that the metallicity of a star correlates to its age according to some age-metallicity relation (AMR). However, the observed age-metallicity distribution for open clusters (Geisler et al. 1992, Friel 1995) and disk stars (Marsakov et al. 1990, Edvardsson et al. 1993) almost completely dispels this notion. As can be seen in Figure 20, at any given age of a wide range of [Fe/H] can form, and, moreover, this is also the case in the abundance patterns, such as the $\alpha$/H and neutron capture elements, like barium. The mean increase in these abundances over the life of the disk has been only a few 0.1 dex. This great inhomogeneity in element abundances at any age suggests that chemical mixing in the disk is not very efficient.

Closer inspection of Figure 20 shows, however, that there do appear to be radial abundance gradients. For [Fe/H] in the open clusters, the gradient is found to be $-0.091 \pm 0.014$ dex kpc$^{-1}$ (Friel 1995). This is similar to the gradient found in the Edvardsson et al. stellar sample. Based on the observed radial abundance gradient, it appears that over the history of the disk it was position in the disk that played a dominant role in determining a star/cluster metallicity, so that, remarkably, at almost any given disk abundance we can find both old and young disk stars.

It should be pointed out that the sample used to produce Figure 20 is a highly selected one, and not representative of the relative metallicity distribution in the disk. Several groups have updated the thin disk metallicity distribution via surveys of G dwarfs in the solar neighborhood. The results of these studies are shown in Figure 21. Note that the lack of a significant number of very metal poor stars in the solar neighborhood (the “G-dwarf Problem”) has long been a problem for simple chemical evolution models of the Galaxy that follow the enrichment of gas in “closed-box scenarios”. That simple model scenario is unlikely to be a good one, given the likely possibilities for pre-enrichment of disk gas or infall of enriched gas onto the disk, both explanations for the paucity of extremely metal poor disk stars (see Gratton’s discussion in this volume).

If we now simplistically adopt the large scatter in the metallicity distribution of disk stars as shown in Figure 21, the age-metallicity relation as shown in Figure 20, and the chromospheric age distribution shown in Figure 18, we may attempt construction of an approximate Hodge population box for the Galactic disk. The result is shown in Figure 22. In this Figure, we also anticipate the addition of the IPH, discussed next.
3.3. Intermediate Population II, Thick Disk

3.3.1. Age

The general consensus is that the IP II component is old—approximately the age of the halo. The precise relative timing of the first star formation in the halo and IP II bears critically on some formation scenarios, but is still somewhat uncertain. As usual, complications in age dating the IP II extend from problems in separating pure IP II samples reliably from those of the halo (particularly the “low” halo) and the old, thin disk, both populations that have significant overlap with the IP II spatially, kinematically, and chemically.

- Disk globular clusters are commonly attributed as tracers of the IP II on the basis of their vertical scaleheight (see Figure 16), velocity dispersion, and the known orbit of the prototype disk globular 47 Tuc. Unfortunately, few disk globulars have been age-dated relative to other globulars or in an absolute sense, but the numbers are growing. 47 Tuc appears to be some 1−2 Gyr younger than the oldest halo globulars, when measured on the same age scale (Chaboyer et al. 1992), but perhaps several Gyr older than some
Figure 21. The observed G dwarf metallicity distribution for stars in the solar vicinity, as derived by three groups. From Chiappini et al. (1997).

Figure 22. The SFR-abundance population box for the Galactic disk, synthesized from our survey of results for disk stars and clusters. The figure shows the three SFR maxima ("bursts" A, B, and C) from Figure 18, and one additional burst (D?) near 13 Gyr that corresponds to the formation of the IP II thick disk (discussed in Section 3.3). The abundance spreads are constrained by the metallicity distributions for thin disk G dwarfs presented in Figure 21, and the metallicity distribution for the IP II presented in Figure 23. The overall spread in abundance with age is adopted from the Edvardsson et al. (1993) data presented in Figure 20. A possible IP II metallicity spread to a "metal-weak thick disk" for the oldest "burst" is hidden from view in this representation.

of the youngest halo globulars (see below). From a variety of studies, it appears that many of the disk globulars studied – 47 Tuc, NGC 6352, NGC 6760 and M71 – are coeval (Hodder et al. 1992, Fullton 1995, Grundahl 1996, Salaris & Weiss 1998), and, in the age dating studies with the shortest cluster timescales, the same age as the oldest disk white dwarfs (Reid 1998, Salaris & Weiss 1998). However, one disk globular, the more metal rich NGC 5927, exhibits a younger age by 2 – 3 Gyr than these other disk clusters.
Another disk cluster, NGC 6553, has yielded a range of ages less than or equal to the old halo clusters (Demarque & Lee 1992, Ortolani et al. 1995). On the basis of their globular cluster populations, it would appear that the IPII began forming soon after the halo began forming but before the halo stopped making globular clusters. Note that the question of the population assignment of metal-rich clusters within a few kiloparsecs of the Galactic center has become a matter of debate (see Section 3.5): It has been proposed that these metal-rich clusters, including those with younger ages just mentioned, may be members of a bulge population of clusters (Minniti 1995, 1996).

- Tidal circularization of binary stars has yielded similar ages for halo and IPII binaries (Latham et al. 1992, Carney 1993). The basis of the technique is that short period binaries have their orbits circularized by tidal friction mechanisms, e.g., turbulent viscosity in the convective envelopes of stars acting on the equilibrium tide (Zahn 1977). Thus the orbital period at which the transition from elliptical to circular orbits due to the tidal interaction is a function of time.

- The existence of an apparent disk-like RR Lyrae population has been attributed to the IPII. More metal rich members, if IPII, set the age of the IPII to $>11\text{Gyr}$ (see Figure 3), via comparison to the age of onset of an RR Lyrae population in the Magellanic Cloud clusters (Rodgers 1991).

- The apparent main sequence turn-off for IPII field stars could be used to establish a formation age. The color of the turnoff for stars of intermediate metallicity is consistent with the IPII stars being as old as 47 Tuc (Carney et al. 1989, Rose & Agostinho 1991, Gilmore et al. 1995). Moreover, if there were a younger component to the IPII, then one would expect to find stars bluer than $B-V = 0.5$ at the intermediate metallicities expected to be most representative for the IPII (Norris & Green 1989). Few such stars are found in the distance range $1< Z < 5\text{kpc}$ in the magnitude complete surveys of Croswell et al. (1991) and Majewski (1992). Thus, it would appear that the bulk of the IPII stars were formed at the same time, unless the mean abundance for the IPII is significantly higher than that of 47 Tuc, as has been suggested recently by Reid (1998).

- From Strömgren photometry a more specific timeline emerges. For a reasonably large sample of stars from the halo and “high velocity disk stars” (IPII) it has been established that the IPII is 1-2 Gyr younger than the inner halo, but perhaps a little older, or as old as, the outer halo (Marquez & Schuster 1994). One explanation is that the formation of the outer halo and IPII are related, perhaps in a puffing of a primeval thin disk to produce an old thick disk at precisely the time the halo was apparently accreting its outer halo. Perhaps the same “fragments” (Searle & Zinn 1978) that formed the external halo were responsible for dynamically heating the thin disk at early times. In this case, the IPII, thick disk stars could represent the population of primeval thin disk stars that were dynamically heated, whereas the thin disk we see today represents those disk stars formed after this event, as well as any primeval thin disk stars left relatively unaffected by the event. Presumably, the thick disk could be composed, partly or wholly, from the remains of the shredded, merged satellite galaxy or “fragment”.

From the present data, the most consistent time sequence that emerges is that the bulk of the IPII stars and globular clusters appear to have formed at about the same time, which was after the inner halo formed but at the same time or before the outer halo formed. Such a timeline has important implications for the formation of the IPII. For example, if the IPII formed from a merger event puffing up a previously formed disk, this event must have occurred quite early in the history of the Galaxy. However, the age sequence with the thin disk is also consistent with a formation scenario wherein the IPII may have been the less dissipated precursor to the present thin disk.
3.3.2. Abundance

It has been considered well established that the mean metallicity and peak of the metallicity distribution for IPII stars is near [Fe/H] \(-0.6\) (see summary in Majewski 1993; Gilmore et al. 1995). However, one of the surprising results of work on presumed IPII stars in the past decade is the possibility of a metal poor tail to the IPII (Norris 1986, Morrison et al. 1990, Majewski 1992, Morrison 1993, Beers & Sommer-Larsen 1995). Clearly there is some controversy over this “metal weak thick disk”. For RR Lyrae stars with kinematics like those expected for the IPII, Martin & Morrison (1998) report stars as metal poor as [Fe/H] \(-2.05\), yet, from a very similar data set, Chiba & Yoshii (1998) report essentially no thick disk stars with [Fe/H] \(<-1.6\). Moreover, recent work with HIPPARCOS data by Reid (1998, 1999) argues that commonly used abundance scales for disk stars may be substantially underestimated, by as much as 0.4 dex. If so, such a change would affect the distributions shown in both Figures 21 and 23.

Apart from new issues regarding the metallicity scale, there are the longstanding problems associated with difficulties in separating IPII stars from other components. Nevertheless, there have been some attempts to determine a metallicity distribution for IPII stars; the abundance distribution for a sample of IPII stars at \(Z = 1.0 - 1.5\) kpc is shown in Figure 23. Though the sample size is relatively small in this case, and whether or not the metallicity scale adopted needs to be shifted or compressed, an apparent tail to low abundances makes the distribution decidedly non-Gaussian.

There seems to be some consensus that there is almost no vertical abundance gradient in the IPII (Yoss et al. 1987, Carney et al. 1989, Majewski 1992, Gilmore et al. 1995). However, Chiba & Yoshii (1998) find some evidence for an abundance gradient in the metal weak portion of the IPII, which they argue implies a dissipative IPII formation (Section 3.3.4) after major parts of the halo formed.

Combining the age timeline from the last section with a smoothed version of Figure 23, we might expect the age–abundance distribution of the IPII to look something like that shown in Figure 31 at the end of this article. Note the asymmetric tail to lower abundance in that representation. A schematic Hodge population box of the combined thin and IPII disk was shown in Figure 22, earlier, with the IPII as part of the possible “burst D” in the star formation rate history as shown in Figure 18.
3.3.3. Kinematics of the Disk Populations

For this discussion, we adopt the coordinate system of velocities \((U, V, W)\), which correspond to motion with respect to the Local Standard of Rest in the directions toward the Galactic Center \((X)\), in the direction of the motion of rotation \((Y)\) and the direction towards the North Galactic Pole \((Z)\), respectively. The IAU adopted standard velocity of the Local Standard of Rest, the motion expected for an object at the Galactic radius of the Sun orbiting the Galactic Center in true circular motion (i.e., in the absence of peculiar motions), is \(\Theta_{LSR} = 220 \text{ km s}^{-1}\), but recent work suggests that this value may be closer to \(184 \pm 8 \text{ km s}^{-1}\) (Olling & Merrifield 1998). The difference between the \(V\) velocity of a star and \(\Theta_{LSR}\) gives the rotational velocity of the star. Another name for \(V\) is the asymmetric drift velocity, which is 0 for the LSR.

Perhaps in no other property is the problem of the division (?) of the thin and thick disk components less certain than in the kinematics, with widely reported differences in the derived kinematics for the IPII among different groups. For example, the rotational velocity of the IPII has been reported to be anywhere from a single value in the range \(20 - 100 \text{ km s}^{-1}\) to a gradient between these values as a function of \(Z\)-distance (see summary in Majewski 1993). At least three uncertainties confound the problem: (1) uncertainty in the relative density laws (Figure 13) between Galactic components, (2) uncertainty regarding a gradient in the kinematics, as opposed to, say, a single valued rotational velocity, for the IPII (gradients in kinematics are expected for the thin disk due to secular heating of stars as they scatter off of giant molecular clouds complexes and other perturbations), and (3) the question of whether the thin disk and IPII are distinct components, or rather, parts of a single, contiguous population as described by Norris (1987). An illustration of the problem is given by the distribution of kinematical properties of all stars in complete samples at the Galactic poles. Figure 24 summarizes the mean asymmetric drift velocities and their dispersion as a function of mean \(Z\) of the sample. While there is general agreement in the global trend of kinematics between the surveys, the derived kinematical properties for specific Galactic components range broadly among these same samples. For example, Soubiran (1993) finds a rotational velocity of \(179 \pm 16 \text{ km s}^{-1}\) for her IPII, Spaenhauer (1989) finds a rotational velocity of \(140 - 160 \text{ km s}^{-1}\), while Majewski (1992) claims a gradient in IPII velocities from around \(200 \text{ km s}^{-1}\) near the Sun to about \(100 \text{ km s}^{-1}\) some 5.5 kpc above the plane. These differences are clearly related to disagreements in how the stars above the plane are distributed into populations (refer, again, to Figure 13): The IPII parameters \((h_{\text{IPII}}, \rho_{\text{IPII}})\) adopted for each of the three surveys just mentioned are \((0.7 \text{ kpc}, 6\%)\), \((1.3 \text{ kpc}, 2\%)\) and \((1.4 \text{ kpc}, 3.8\%)\), respectively. Majewski (1993, see Figure 6 in that paper) demonstrates that similar effects are seen in the derivation of rotational velocities from IPII tracer (i.e., selected star type) surveys when the mean heights of the samples are considered. Ultimately, of course, the decomposition of the disk into components (if indeed this is appropriate as opposed to a single, continuous population) must satisfy self-consistently and simultaneously the kinematical, chemical and starcount constraints, and may require multivariate analyses such as those described by Nemec & Nemec (1991, 1993), Soubiran (1993), and others.

Nevertheless, without such complete modeling some broad statements can still be made about the kinematical properties of the thin disk and IPII. First, given the relative spatial dimensions of the thin and thick disks and the results of Figure 24, it should be evident that the kinematics of the IPII/thick disk are more extreme (larger velocity dispersions, slower rotational velocity) than those of the thin disk. Then, by bearing in mind the relative ages of the IPII (which is apparently uniformly old) and thin disk
Figure 24. The run of asymmetric drift velocity and dispersion as a function of height (in pc) above the Galactic plane for complete surveys of proper motions at the Galactic poles. Error bars have been left out for clarity, but generally range from a few km s$^{-1}$ to $\sim 20$ km s$^{-1}$ vertically. General agreement is seen between the surveys for the global kinematics of all stars, even when the individual surveys derive widely different values for the IPII kinematics after distributing the stars (differently!) into populations. From Majewski (1994a).

Figure 25. The run of the velocity components $U$, $V$ and $W$ (left panels), and their dispersions (right panels) with age for a set of disk stars having ages determined from chromospheric HK measures by Rocha-Pinto et al. (1998). The spread in velocity dispersion with age is obvious in the thin disk, though note the apparent saturation of the increase beyond several Gyr. From Rocha-Pinto et al. (1998).

(which apparently spans a range of ages from 0 up to that of the IPII) populations, we may surmise a timeline for the formation of stars now at different velocities: At the extremes, the older stars of the IPII have hotter kinematics while the younger stars of the disk have colder kinematics. From dynamical considerations, this is what one might expect – hotter kinematics are required to support a more extended spatial structure.

Figure 25 shows that this basic timeline is, in fact, a more general age-velocity relation (AVR), well established for some time for the thin disk. The growth in the velocity
dispersion of thin disk stars as a function of age is obvious, as is the trend to a slower rotational velocity (increasingly negative $V$ velocity) with age. Note that most of the increase in velocity dispersion occurs early on (see below). It is reasonable to assume that the age-velocity trends shown proceed through to the even older stars of the IPII, since it is known that the kinematics as well as the vertical distribution of the IPII stars are more extreme than those shown in Figure 25. From the approximate AVR we observe (Figure 25), and the disk star formation history observed (Figure 18), we may assemble an observed dynamical population box for the disk populations, as shown in Figure 26. Note that without knowledge of the birth kinematics of the populations, or how disk stars may have evolved to the observed distributions, we cannot construct a reliable dynamical population box for the stars as they formed.

3.3.4. Formation Scenarios for the Disk

We now have collected several important observational facts about the disk population(s). First, the IPII appears to be an old population, formed at about the same time as the halo, and before much of the thin disk. Second, the almost non-existence of an AMR and large scatter in abundances for the disk populations at any age hints at very poor mixing of the pre-stellar gas. Third, there is an observed AVR across the disk populations.

How did the observed AVR for disk stars arise? Why are the older disk stars dynamically hotter, and therefore more spatially extended, than the younger thin disk stars? Basically there are two possibilities: (1) the older stars were born that way, or (2) the older stars were dynamically heated and diffused from colder, thin disk-like orbits.

In fact, one or the other of these two answers form the bases for most creation scenarios for the IPII thick disk. A detailed accounting of the basic models and their variants is described in Majewski (1993). Here we concentrate mainly on the two families of IPII formation corresponding to the two methods of creating the observed AVR. I call these two families the (1) “top down” models, wherein the AVR is more or less created in place as gas from the forming Galaxy settles into an ever thinner, dynamically colder, plane; and (2) the “bottom-up” models wherein an initially thin, relatively cold population is dynamically heated in some way. As in many situations in astronomy, the true history of the disk may well involve some combination of the extremes.
Among dynamical diffusion processes, there is a firm theoretical foundation for secular dynamical heating. We have already seen in Section 1.2 that open and globular clusters that travel through or within the disk see time varying gravitational fields that act to accelerate stars within, and cause eventual destruction of the cluster through evaporation. These same fields act on individual stars once they have escaped from clusters. The predominant secular heating process is the Spitzer-Schwarzschild mechanism (1953), whereby the accumulated effect of repeated collisions of stars with large, $10^{5-6} M_\odot$ molecular cloud complexes is a steady increase in the velocity dispersion of a coeval disk population with time, according to

$$\sigma[v(t)] \sim \sigma[v(0)][1 + (t/\tau)]^{1/3},$$

where $\sigma[v(0)]$ is the initial velocity dispersion of the stellar population, and $\tau$ is the characteristic timescale for energy exchange in the collision process. The value of $\tau$ is about several 0.1 Gyr. This approximate functional form for the rate of stellar dispersion increase is evident in the data in Figure 25.

The leveling out of the velocity increase for older stars reflects a saturation effect with the scattering process: As the velocity dispersion of a population increases, the average time between encounters increases (as stars spend more time away from the thin molecular cloud layer), and the rate of velocity dispersion increase slows. While not universally observed among all studies to date (perhaps relating to problems of separating thin disk and IPII/thick disk stars from one another), some studies (as shown above) show a saturation effect at ages well within the domain of thin disk stars alone. In either case, given that the present observed density of giant molecular clouds is only sufficient to increase the vertical velocity dispersion $\sigma_W$ to at most $15 - 25$ km s$^{-1}$ (Villumsen 1985, Binney & Lacey 1988), it is difficult to understand how the Spitzer-Schwarzschild process can account for the velocities of the oldest thin disk stars, let alone the IPII stars, which appear to have $\sigma_W \sim 45$ km s$^{-1}$. Moreover, it is hard to understand how disk globular clusters, and the high Z-distance open clusters, could have been pumped up to such large distances above the plane with a secular heating process.

Of other proposed secular heating mechanisms, (1) the passage of spiral arm waves (Barnabés & Woltjer 1967, Carlberg & Sellwood 1985) appears to be coupled mainly to increasing velocity dispersions within the plane, and unable to produce significant dispersion perpendicular to it (Carlberg 1987), and (2) a vast sea of fast moving, massive ($10^6 M_\odot$) objects (supermassive black holes?) penetrating the disk and creating short-lived perturbations, while able to heat some stars to large enough velocity dispersions, is apparently unable to form a 10% mass (see Figure 12) IPII component (Ipser & Semenzato 1985, Lacey & Ostriker 1985). If they are the major constituent of the dark matter, some $10^5 - 10^6$ of these $10^6 M_\odot$ objects would be needed, and one might expect that they would have been seen in the various microlensing experiments. Indeed, the latter experiments suggest that the main contributor to the microlensing phenomenon are only of mass $0.5 M_\odot$.

Secular heating almost certainly acts on disk stars, but is insufficient to accommodate the large velocity dispersions seen in the old disk and IPII stars. Alternatively, the disk may have experienced one or more episodes of violent heating, as expected during minor mergers of small satellite galaxies. Due to the processes of dynamical friction, the orbits of satellite galaxies in the presence of dark matter halos should decay and lead to accretion by parent galaxies. The Milky Way is in the process of just such an accretion event, as the recently discovered Sagittarius galaxy clearly is in the midst of tidal disruption (see below). Dynamical simulations show that such merger events can
lead to the formation of thickened, heated disks that resemble the IPII (Figure 27). Because thin disks are “destroyed” in the merging process, most of the thin disk we see today must have formed subsequently to any merger event. Therefore, the thickened stars must be older than the oldest stars in the thin disk, and this may well be the case in our Milky Way. If this scenario is the explanation for the origin of the IPII, then, by the apparent age of that component, we have a relatively tight constraint that the merger event must have occurred early on in the life of the Milky Way, and, moreover, that no event of that magnitude has occurred since. It is interesting to note that since thick disks do not appear to exist in all disk galaxies (van der Kruit & Searle 1981, Morrison et al. 1994), the formation of a thick disk may not be a requisite phase of disk galaxy evolution. Thus, a stochastic process for the formation of the thick disk, like minor merger events, seems consistent with observations of external galaxies.

“Top-down” Formation

The idea that the disk kinematics we observe now may have been formed in situ has roots in the model of ELS. A key component of their model is the idea of “spin-up” – that as a proto-galactic cloud collapses it must preserve angular momentum and therefore it increases its rotational velocity as its radius shrinks. Another key component of the ELS model is that the proto-disk gas came from the leftover gas from the formation of the halo. While the timing of the formation of the IPII, which apparently began forming while the
halo itself was still forming, may now preclude the halo-then-disk phasing of ELS†, the basic concept of spin-up, as it is a direct result of the conservation of angular momentum, must occur at some point in the formation of the Galaxy. Sandage (1990) proposes that the ELS spin-up, rather than applying to the halo and disk, may apply solely to the disk components, with the IPII representing the point at which the collapsing proto-disk gas first encountered partial pressure support (a slowing period in the collapse). Thus, the IPII might represent a transitional phase in the formation of the disk, perhaps the point at which the gas becomes predominantly supported against collapse by rotational support (any previously formed halo would be supported by the kinetic pressure provided by large anisotropic velocity dispersions of stars in predominantly radial orbits).

Given the apparent narrow age range of the IPII (as given, for example, by the disk globular clusters), its formation must have been rather rapid and early on in the Galaxy’s life. Of course, this question is a function of how distinct the IPII is from the thin disk. If they are smoothly joined, then the requirement would seem to be only that the first phase of disk formation be dramatic, to produce quickly an extended population of stars and clusters, but not necessarily a population discrete in age. Both possibilities – a smooth, gradual formation of the entire disk (Larson 1976), or a disk formed with discrete components formed in separate bursts of star formation (Marsakov et al. 1990, Burkert et al. 1992, Katz 1992) – have been produced in chemodynamical models of disk formation without satellite merging. Pauses in the formation of stars during the collapse of the gas can come from one of several suppression mechanisms as a result of a previous star formation burst:

- **Disruptive mechanisms** in the form of tidal shocks, supernovae, and destructive cloud collisions could make it difficult for substantial star forming clouds to collapse (Gilmore 1984).
- **Temporary depletion of gas** of sufficient density to form stars might occur in a particularly vigorous starburst, and require the accumulation of more gas from larger radii to reinvigorate star formation. Initiation of star formation may require surpassing a threshold in mass density, and this threshold might be reached through replenishment from infall, possibly extragalactic or from a dying first generation of stars (Larson 1974, Gratton et al. 1996, Chiappini et al. 1997).
- **Intense heating** of the gas not used up in the previous starburst could come from the injection of energy, momentum, and metal-enriched material in the form of strong stellar winds (Marsakov & Suchkov 1977).

Thus, because of the possibility of suppression mechanisms in a dissipational collapse, discrete thick disk and thin disk components would seem to be an insufficient condition to discriminate between collapse and merger models for IPII formation. Note, even without a gap in star formation activity, it is possible to produce a rather chemodynamically distinct IPII and thin disk if the end of the IPII phase is coincident with a rapid increase in the rate of dissipation and collapse of the gas. One proposed possibility is that when the gas self-enriches to a certain metallicity ([Fe/H]~ −1), the dominant dissipation mechanism of the gas switches from the less efficient process of free-free transitions of electrons in encounters with positive ions (bremsstrahlung) to the much more efficient cooling by line radiation (Wyse & Gilmore 1988, Burkert et al. 1992). The increased

† Moreover, from data not presented here, it is fairly clear that the disk and halo are rather distinct from one another in a chemodynamical sense. That is, the halo and IPII are not smoothly joined, but rather, they are disjoint in their chemodynamical distributions. Combined with the near simultaneity of their age, it is difficult to see how their formations could have been part of a smooth, contiguous process. It is more likely that the leftover halo gas proceeded to make up the bulge of the Galaxy (Wyse & Gilmore 1993).
cooling rate causes rapid dissipation of energy and the cloud is able to collapse more quickly. The quick transition to higher densities probably induces an increased rate of star formation.

In summary, two formation scenarios for the formation of the old, thin disk and/or IPII seem to be viable at the present time: (1) some kind of merger event at early times, a process that we would expect to create (from the “bottom-up”) a distinct, likely disjoint IPII component, with likely little kinematic gradient perpendicular to the disk, or (2) some form of “top-down” formation with ELS spin-up during a dissipational collapse, which may or may not form an IPII that is discrete and disjoint from the thin disk, but which might show a global kinematic gradient.

In my opinion, despite, and because of, several recent claims each way, the jury is still out on the discreteness of the thin and thick disk populations. Without a clear idea of whether and how the disk should be divided into thin/thick disk populations, it is too early to discriminate between models, and it is certainly premature to attempt to construct a true birth dynamical population box for the disk populations. The best we can do at present is formulate what the observed distribution of thin and IPII thick disk dynamics are like (Figure 26).

3.4. The Halo

3.4.1. Age Characteristics of the Halo

A great deal of what we know about the age of the halo comes from the halo globular cluster system. Indeed, the halo globulars provide critical constraints on the age of the universe, and, thereby, the size of the Hubble Constant. A few constraints on the age of the halo come from halo stars, but generally this work is less well developed compared to the long history associated with age-dating globular clusters:

- The white dwarf luminosity function for halo stars obviously would provide a very useful constraint on age, or at least a convenient relative chronometer to the disk. However, as shown in Figure 17 above, data on field halo white dwarfs is extremely scant, so the techniques described for age-dating the disk with this technique cannot yet be brought to bear on the halo field. But, because halo white dwarfs are considered the “least unlikely” explanation for the MACHO microlensing events (Méra et al. 1998), a number of new searches for halo white dwarfs are now underway. In the meantime, deep HST imaging on halo globular clusters now makes possible the production of CMDs reaching faint enough to see the cluster white dwarf sequences (Richer et al. 1997; see King and Castellani contributions to this volume). Thus, we are on the verge of bringing a new technique to verify the cluster age scale as determined from isochrone fitting, or, inversely, we can use the latter as a check on white dwarf cooling theory.

- The main sequence turn-off color for the most metal-poor subdwarfs is near $B - V = 0.36$, which is similar to the main sequence turn off of the metal-poor clusters M15 and M92 (Sandage & Kowal 1986, VandenBerg 1991). This confirms that the age for the bulk of the youngest halo field subdwarfs is similar to that of the oldest globular clusters.

- Strömgren photometry of metal-poor, halo stars by Marquez & Schuster (1994) does show that there is an age scatter of several Gyr in the halo field. Moreover, they find that the stars with apogalactica outside of 10 kpc have a mean age some 2 Gyr younger and a smaller age dispersion than the stars interior to this radius. This result is in direct contradiction to expectations from the original ELS picture, which describes a more or less ordered progression of Milky Way formation from the outside in. The Marquez & Schuster results are consistent with the latest findings concerning the order of formation of the halo globular cluster system (Zinn 1993a, Sarajedini et al. 1997).
- The *halo globular clusters* have provided us with some of the most important breakthroughs in understanding the formation of the halo in the past several decades. In principle the mean age of the halo globular clusters sets the time of halo formation, while the existence of any cluster age range defines the duration of the formation epoch for the halo. The former is more difficult to determine since absolute ages depend critically on accurate determinations of cluster abundances (not only [Fe/H], but [O/H], etc.), as well as distance moduli, which must be known to 0.10 magnitude accuracy to achieve a 10% precision in the age. There has been up to a factor of two range in the age scale among various groups in recent years, although the most recent results based on HIPPARCOS data (see review by Reid 1999) tend toward the lower age range (i.e., clusters more distant, on average, than previously thought). The details of absolute age determinations are beyond the scope of the present discussion, but see discussion by Castellani (also, e.g., Bergbusch & VandenBerg 1997, Gratton et al. 1997, Chaboyer et al. 1992, 1998). Here we focus on *relative* cluster ages, where some of the most fruitful results are found from the point of view of Galactic structure and formation studies.

The availability of high precision color-magnitude diagrams and accurately determined, spectroscopic metallicities and radial velocities has made a substantial impact in the study of differential ages in globular clusters, to where relative ages can in most cases be determined to 1 Gyr or better (modulo, of course, the age scale based on the adopted distances and isochrones) between appropriately chosen pairs of clusters. Most revealing are studies of “second parameter pairs” – pairs of clusters with the same abundance but very different looking horizontal branches: one cluster with a predominantly red HB and the other one with a blue HB. The HB morphology can be quantified as \((B - R)/(B + V + R)\) – where \(B\) is the number of horizontal branch stars blueward of the RR Lyrae gap, \(R\) is the number of HB stars redward of the gap, and \(V\) is the number of RR Lyrae stars. The “primary parameter” driving the nature of the horizontal branch is mean abundance, [Fe/H]. However, the distribution of cluster HB morphologies against [Fe/H] reveals a spread of HB morphologies for the abundance range \(-2 < \text{[Fe/H]} < -1\) (Figure 28). Some examples of different HB morphologies in clusters are shown in King’s contribution to this volume. The main dependence of HB Type on [Fe/H] is obvious in Figure 28. The *second parameter* that causes the spread in HB Type at any given
[Fe/H], and which may actually be a combination of effects (i.e., “second parameter” = “second parameter” + “third parameter” + “fourth parameter” + ...), has been variously ascribed to, or partly to, a number of causes, including (see Fusi Pecci et al. 1996 for a summary of how some of these effects operate on the position of the HB):

- age
- cluster density
- rotation
- α-capture element abundances
- oxygen abundances affecting mass loss rates
- helium abundances and “deep mixing” of helium in highly convective stars
- helium core mass at helium flash
- differences in the number of planets swallowed up in the red giant phase.

While many details need to be worked out, and there is certainly no consensus on this, evidence seems to be pointing to age as at least one (though possibly not the sole) second parameter dictating HB morphology, after [Fe/H] (Sarajedini et al. 1997). An age range in the halo globulars, suspected since the seminal study by Searle & Zinn (1978, “SZ” hereafter) discussed the strong second parameter effect in the outer halo, is now generally accepted: Even groups that strongly disagree on the source of the second parameter and the amount of age spread in the halo globular clusters admit to finding at least some globulars that are clearly younger (by 1 or more Gyr) than the bulk of the halo system, and the usual culprits (the likes of Arp 2, IC 4449, Rup 106) have unusually red HBs for their metallicities (though some others – Pal 12 and Ter 7 – are so metal rich that their red HBs are expected by the primary parameter). The list of identified “younger” halo globular clusters encompass almost the full range of halo cluster metallicities. Surveying these results, an age spread of 1 Gyr seems certain, and this is generally found by groups in the compressed cluster age scale camp (e.g., Reid 1998). At the other extreme, and generally claimed by groups in the large cluster age scale camp, a halo cluster age spread has been given as large as 5 Gyr (Chaboyer et al. 1992, Sarajedini et al. 1997).

Most intriguing are relative age rankings of objects in different cluster populations. The comparison of the CMDs of several halo clusters (Pal 12, Terzan 7) to that of the archetypal disk globular 47 Tuc reveals these particular halo clusters to be about 30% younger (Buonnano et al. 1998, Rosenberg et al. 1998); clearly, some clusters in the halo were still forming long after the disk began to form, as first suggested by SZ. According to the MAI analysis in Figure 19, there are even open clusters older than the halo globulars clusters Terzan 7, Arp 2, Pal 12 and IC4449.

Detailed analysis of the abundance and kinematical properties of groups of clusters selected via their distribution in the [Fe/H]-HB Type plane has led to important breakthroughs in our paradigm for the structure and origin of the globular cluster system and its subsystems since the simpler picture presented in Figure 16a. It was noted by SZ that the second parameter effect was most prominent in the outer halo, while at the same time there is no radial abundance gradient in the \( R_{GC} > 8 \) kpc globulars; on the assumption that age was the second parameter, these observations prompted SZ to conclude that the outer part of the halo formed slowly. But, on the basis of the large abundance spread among the clusters in the outer halo, coupled with the fact that there is little abundance spread within globulars, the outer halo clusters must have formed within larger structures (they used the term “fragments”) that were independently evolving, self-enriching, and mixed internally. Later these transient fragments could be destroyed by supernovae or tidal effects, and the debris continued to fall into dynamical equilibrium with the Milky Way well after any main collapse of the Galaxy had begun to form the interior regions.

A slow, SZ-like halo formation (spanning 3 – 5 Gyr according to some studies) requires
repositories of gas in the halo ("fragments") and with a range in enrichment levels well after any global collapse would have collected mass to the central regions. Natural candidates for such repositories would be either (1) smaller primordial density fluctuations left behind by a main collapse (presumably at large Galactocentric radius) and on a "slow path" to initial dwarf galaxy star formation, or (2) larger, satellite galaxies capable of sustaining multiple star formation epochs and enrichment. As sites of prolonged star, and presumably cluster, formation (see the example of the Carina dwarf, in Figure 5), satellite galaxies could contribute "young halo" clusters (as well as, of course, older clusters) to the Galaxy after being tidally disrupted. As if on cue, Nature and Fortune provided us with the recently discovered satellite galaxy of the Milky Way in Sagittarius (Ibata et al. 1995), which is now observed to be in the process of tidal disruption. In the process of this tidal disruption, Sgr is contributing second parameter clusters (Arp 2, Terzan 7) to the Galaxy, as well as two other clusters (M54, Ter 8) with "normal" HB types.

Other evidence suggests an association of second parameter clusters to dwarf satellite galaxies of the Milky Way. Two second parameter globular clusters, Pal 12 and Rup 106 have been shown (Lin & Richer 1992) to have radial velocities consistent with being tidal débris from the Magellanic Clouds. More detailed statistical analysis of the locations and possible orbits of all of the second parameter clusters show them to have a high degree of association with the positions and orbits of the dwarf satellite galaxies of the Milky Way (Majewski 1994b, Palma & Majewski 1999), while the globular clusters still attached to dwarf satellites of the Milky Way are almost exclusively of the Zinn "young halo" type with second parameter, redder HBs (Zinn 1993b). The weight of evidence favors a picture in which the second parameter clusters are derived from the tidal stripping of dwarf satellite cluster families by the Milky Way.

How does this affect our previous "division" of the globular cluster system into "disk" and "halo" systems? Zinn (1993a) demonstrates that there are actually two halo globular cluster systems that show distinctly separable chemodynamical properties (see also van den Bergh, 1993). The basis of the separation is the second parameter effect. Clusters with the bluest HB at any given [Fe/H] (those clusters along the ridge line in Figure 28), i.e., those that show no second parameter effect, define Zinn's "BHB" sample. If age is the second parameter, these objects are presumably the oldest clusters at every [Fe/H], and define the "old halo" in Zinn's (1993) analysis. All clusters falling away from this ridge line, the "RHB" sample, show a second parameter effect, and are termed the "young halo" clusters by Zinn (1993a). With this division into halo subpopulations, we find the properties shown in Table 2. As noted previously by Searle & Zinn, the outer halo has a dominant fraction of second parameter, "young clusters", so that the Galactic density distribution of RHB clusters is essentially spherical. In contrast, the BHB clusters are in a more oblate configuration, with generally smaller Z-heights. Note too that, as found by SZ in the outer halo, the RHB clusters at large $R_{GC}$ show no abundance gradient, while the BHB clusters, which are predominantly at smaller $R_{GC}$, do show an abundance gradient. Finally, the kinematics of the RHB clusters are vastly more extreme than those of the BHB clusters, with "hotter" orbits (Dinescu et al. 1999) and a mean rotation that is marginally (with respect to the uncertainty) retrograde. If true, the mean retrograde rotation rules out the possibility that the RHB could have

† In a more recent discussion, Zinn (1996) further divides this "non-second parameter" sample into two groups, the "BHB" clusters with $[\text{Fe/H}] \geq -1.8$ and the "metal-poor" ("MP") clusters with $[\text{Fe/H}] < -1.8$. The "second parameter" halo clusters are called the "RHB" sample in Zinn (1996)
formed in a grand collapse of the Galaxy, since the direction of angular momentum of a collapsing cloud cannot change directions. Rather, the properties of the RHB cluster system, taken as a whole, support the idea that these objects were accreted.

In contrast, the properties of the BHB, “old halo” cluster system are consistent with a general collapse with ELS spin-up. Zinn (1993a) suggests that the BHB globulars may simply represent a more extended, metal-poor, kinematically hotter extension of the disk globular system (Figure 29). Indeed, several rather metal poor, BHB clusters (NGC 6254, NGC 6626 and NGC 6752) are found to have rather IPII-like orbits (Rees & Cudworth 1991, Dinescu et al. 1999). Taken together, the disk globular clusters and the old halo, BHB globulars seem to be consistent with formation in a single dissipational, collapsing, metal-enriching entity. Such a progression is suggested by the distribution of Z-max distances derived from orbits determined for clusters with derived proper motions (Figure 16b). A new paradigm for the division of the Galactic globular cluster system, based on Zinn’s (1993a) analysis, is shown in Figure 16c.

3.4.2. Dual Halo Scenarios

The need for two separate, metal-poor, halo globular cluster populations in our description of the Galaxy is apparently mirrored in the halo field star population. “Dual halo” models for the distribution of metal-poor stars in the Galaxy have become increasingly popular since Hartwick’s (1987) study of the spatial distribution of metal-poor RR Lyrae stars (see examples in Table 1). In these models, the space distribution of “halo” stars incorporates both a spherical component as well as a component that is significantly flattened (with exponential distributions perpendicular to the disk of approximately 2 kpc scaleheight). Such a dual component distribution has been used to describe the spatial properties of metal-poor, field RR Lyraes (Hartwick 1987), blue horizontal branch stars (Kinman et al. 1994), and ordinary main sequence stars, whether selected by low metallicity (Allen et al. 1991, Sommer-Larsen & Zhen 1990) or by large proper motion (Carney et al. 1996).

Curiously, while studies of metal-poor field populations have moved towards “dual halo” descriptions, studies of the Galactic IPII thick disk have pushed the metallicity limit of this population to as low as [Fe/H] = −1.6 and, in some cases, as low as [Fe/H] = −2.05 or less (see discussion in Section 3.3.2). If one accepts that such metal-poor IPII stars exist, then the ability to distinguish these stars from similar metallicity stars in any “flattened halo” becomes extremely difficult: The scaleheight of some studies of the IPII (Figure 12) is similar to that proposed for the “flattened halo” — but with a higher density normalization so that the IPII dominates the flattened halo. Moreover, at large distances from the Galactic plane the kinematics of the IPII become much more extreme in some studies (see Majewski 1993) and less separable from those of halo stars. These

<table>
<thead>
<tr>
<th>V_{rot} (km s^{-1})</th>
<th>σ_{los} (km s^{-1})</th>
<th>distribution</th>
<th>metallicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>70 ± 22</td>
<td>89 ± 9</td>
<td>oblate</td>
<td>∼ 0.5 dex from R &lt; 6 to 40 kpc</td>
</tr>
<tr>
<td>−64 ± 74</td>
<td>149 ± 24</td>
<td>spherical</td>
<td>no gradient</td>
</tr>
<tr>
<td>44 ± 25</td>
<td>113 ± 12</td>
<td>flattens with small R</td>
<td>no gradient R &gt; 8 kpc or</td>
</tr>
</tbody>
</table>

**Table 2.**
Figure 29. The evidence suggesting that the disk globular cluster system and the BHB, “old halo” globular clusters may represent a single continuous population formed in a dissipational, ELS-like collapse. Note the smooth progressions of spatial, chemical and kinematical properties of BHB, “old halo” clusters and disk globular clusters. From Zinn (personal communication).

Sources of confusion may be part of the reason for the discrepancies in the determined relative numbers of stars in the flattened and spherical halo components: For example, the Kinman et al. (1994) HB stars yield a relative apportionment of flat to spheroidal halo components locally as 4:1, but Sommer-Larsen & Zhen (1990) quote a balance of 2:3 for their metal poor stars. Indeed, the difficulty in distinguishing between stars in the IPII thick disk and those in a “flattened halo” calls into question the need to distinguish between them (Majewski 1995), at least until real evidence exists that such a division is warranted. If, as has been suggested above, the combination of the BHB/“old halo” and disk globular clusters may represent one dissipational entity, might not a similar picture work satisfactorily for the field stars (particularly if the field stars are partly derived from the break-up of globular clusters)?

In such a model, the field star analogy to the “disk + old halo” cluster system might be the Population I thin disk + IPII + “flattened halo”, where the latter represents the
most extreme tail of a more or less continuously distributed, Norris-like (1987) “extended disk”. Thus, in this view, as has been proposed for the globular clusters, we have two origin scenarios for the metal weak field star populations: one dissipational and one by accretion of satellites. Such a picture wedds the Eggen, Lynden-Bell & Sandage (1962) collapse model — applied to the Population I thin disk + IPII + flattened halo — with the accretion scenario of Searle & Zinn (1978) for the more spherical halo; this marriage has been discussed variously by Sandage (1990), Majewski (1993), Norris (1994) and others. Such a unified picture, of course, must be held with caution in the face of some evidence that the IPII thick disk may have formed via violent dynamical heating. However, even in this case the ELS + SZ picture may need only to be modified: Perhaps the “flattened halo” + thin disk were formed originally as a dissipational entity, and the IPII was created soon after by the merger of an early satellite into the disk. As can be seen, studies of the thin disk, IPII and inner halo populations are ripe for more work before a comprehensive origin scenario can be proffered with more certainty.

### 3.4.3. Other Evidence for Accretion in the Halo

In contrast to the impasse just described for the more flattened, inner components of the Milky Way, there is a growing cumulation of evidence for Galactic accretion events, particularly in the outer halo. Indications of accretion of globular clusters from satellite galaxies and into the outer halo of the Milky Way have already been discussed. The same tidal stripping processes that would lead to the accretion of clusters by the Milky Way should also lead to the accretion of stars that are the tidal débris of disintegrating parent objects, presumably dwarf galaxies like Sagittarius. Theoretical modeling of satellite encounters with massive, Milky Way type galaxies (McGlynn 1990, Moore & Davis 1994, Velázquez & White 1995, Johnston et al. 1996a,b) show that stellar systems with relatively small (≤ 10 km s$^{-1}$) velocity dispersions (globular clusters or dwarf spheroidal satellite galaxies) may be tidally pulled into long and long-lived streams of stars strung out along the orbit of the disintegrating parent body (as modeled by Toomre & Toomre 1972). The tidal arms of Sagittarius galaxy have now been mapped to a stretch of more than 40° of the sky (Siegel et al. 1997, Mateo et al. 1998). The destruction of Sgr may provide us with a useful paradigm for both the structure and origin of the halo, for, unless we are at some privileged epoch, events like the destruction and dispersal of Sgr must either be reasonably common or extremely long-lived.

To unwitting observers that happen to be studying a particular direction of the sky along which one of these tidal streams lies, the coherent débris trails would be observed as moving groups of halo stars. Indeed, it was as such a moving group that the Sagittarius galaxy first became known (Ibata et al. 1995). Other evidence for “moving groups” of halo stars has occasionally been alluded to by in situ surveys of the Galactic halo. Typically (and often as a “by the way” discussion) authors have pointed out possible halo moving groups manifested as small numbers of stars in a particular survey field having the same distance and radial velocity (see review by Majewski et al. 1996a). However, the possibility, postulated by Oort (1965), that tidal streamers may be a pervasive, even dominant, component of the field halo star population is suggested by the clumpy halo velocity distribution in the magnitude-limited survey of Majewski et al. (1996b). More recent searches (Harding et al. 1998, Majewski et al. 1998) specifically directed at finding this phase space substructure in the halo have shown that halo moving groups

† Because of stronger phase mixing in the less spherical potential of the inner halo, spatial coherence from accretion is expected to have a shorter lifetime closer to the Galactic center, although coherence in velocity should remain, even increase, with time (see Helmi & White 1999).
are relatively common. It may well be that the outer Milky Way contains a network of crossing tidal débris streams like that shown schematically in Figure 30.

That the halo field star population contains the remnants of disrupted satellite galaxies has long been suspected from the existence of A type stars with high velocity (Rodgers et al. 1981, Lance 1988). Main sequence stars of this spectral type must be relatively young, but their velocities and distances above the Galactic plane are difficult to reconcile with their age if the stars were formed in the Population I disk. Other examples of kinematically hot, yet apparently young stars can be seen in the left panels of Figure 25 (see also Soderblom 1990). These stars could be explained if their formation sites were within objects already having halo-like kinematics, i.e., within one or several recently disrupted satellite galaxies. Alternatively, these stars may have been created as a result of the impact of high velocity gas with the disk (Lance 1988). Contributing to the peculiar age distribution of dynamically “halo-like” stars is a population of “blue metal poor” stars that are apparently only $3 - 10$ Gyr old (Preston et al. 1994); these stars have also been attributed to the break-up of a Galactic satellite within the last 10 Gyr.

The existence of “young” stars in the halo points to dwarf satellite galaxies, which are known to contain intermediate age (e.g., Carina; Figures 4 and 5) and even young populations (the Magellanic Clouds, Carina), as the likely progenitors of tidal débris. However, as we have seen (Section 1.2), globular clusters should also contribute to the field star population (Surdin 1995). Indeed, Oort (1965) suggested that the “pure races of the halo population II” (presumably the extended, spherically distributed field stars counterpart to the Zinn RHB/“young halo” clusters described above) might contain significant structure in the form of intermingling “tube-like swarms” from the breakup of some hundreds or even thousands of globular clusters.

Harris (1991) points out possible problems with this scenario: Globular cluster orbits appear to be more isotropically distributed than those of halo stars and globular clusters are more metal-poor than halo stars at the same $R_{GC}$. However, Lee & Goodman (1995) point out that these might not be problems when one considers that the orbits of surviving clusters are expected to be less eccentric than those of destroyed clusters (see Section 1.2). In the end it is likely that globular clusters do make at least some contribution to
the halo field star population, as tidal tails have now been observed around some globular clusters (Grillmair 1998).

3.5. The Bulge

At a mass of $2 - 4 \times 10^{10} M_\odot$ (Blum 1995), the bulge is nearly half the mass of the disk and of order 30% the Galaxy’s luminous mass. Yet the bulge population is nearly entirely confined within a radius of 1 kpc (compared to the 25 kpc radial extent of the disk, or the > 100 kpc extent of the stellar halo. Other defining characteristics of the bulge include triaxiality (the presence of a bar), relatively rapid rotation, and high mean metallicity. Unfortunately, the Galactic center happens to be at the point of highest density for all stellar populations, a more troublesome situation for separating Galactic populations than that already discussed with regard to separating the thin disk, IPII and halo near the Sun. The properties of the bulge population overlap with those of the inner disk to a degree that clear membership of individual objects cannot be assigned. I now turn to a brief description of the bulge as a stellar population.

Structure: It is thought that the shape of the bulge mimics that of the near infrared surface brightness distribution seen by the DIRBE experiment on the COBE satellite – that is, a triaxial shape consistent with a bar, and even better fit by a boxy shape (Dwek et al. 1995). Strong evidence for a bar, with some tilt to the Galactic plane, is cited by Blitz & Spergel (1991) on the basis of 2.4 μ maps of the Galactic center. It is found that the distribution of Mira variables generally follows the near infrared light distribution, while the late M type stars drop off faster, with a power law of $n = -4.2$ (Whitelock 1993). On the other hand, the distribution of RR Lyrae stars in the bulge follow an $n = 3$ power law, which suggests that they may largely be representatives of the inner halo (Minniti 1996). This is supported by the finding that only in the innermost bulge do the RR Lyrae stars show any evidence for a bar (Alcock et al. 1998). In contrast, metal-rich population stars – red giant clump stars (Alcock et al. 1998), asymptotic giant branch stars (Weinberg 1992), and Miras (Whitelock 1993) – show a clear bar-like distribution.

Abundances: With a recent downward revision of the abundance scale for bulge giants (McWilliam & Rich 1994), a mean abundance of $[\text{Fe/H}] = -0.25$ is obtained in the inner kiloparsec of the bulge; however, stars as metal rich as $[\text{Fe/H}] = +0.5$ are still found. The abundance distribution function is broad (Rich 1988, Minniti et al. 1995, Harding 1996) and, unlike the disk, well fit by a closed box, chemical evolution model, so that there is apparently no G dwarf problem in the bulge (Rich 1990). It is significant if the mean metallicity of the bulge is less than the disk in the solar neighborhood, especially considering that the disk shows a trend of increasing abundance towards the center: Either the bulge formed before the disk or it has been replenished by infall of low metallicity gas. Leftover gas from the halo would naturally sink to the bulge, not the disk, based on the low angular momentum of the halo; thus it is possible that the halo preceded formation of the bulge (Carney et al. 1990, Wyse & Gilmore 1992), an idea consistent with many age determinations (see below).

While Tyson & Rich (1993) find no evidence for a metallicity gradient in the inner bulge, a point they use to argue for bulge formation via a dissipationless collapse, a summary of previous determinations of mean metallicities in a range of bulge fields by Minniti et al. (1995) shows a clear abundance trend in the inner 3 kpc. However, an unanswered question is whether this gradient is intrinsic to the bulge population itself, or the result of changing contributions by halo, disk and bulge stars in the inner Galaxy. A true metallicity gradient in the bulge would support its formation in a dissipational collapse. Minniti et al. (1995) argue that a comparison of the number of stars at the
location of metal-rich giants in color magnitude diagrams from different bulge fields gives strong evidence for a metallicity gradient in the inner 1 kpc of the bulge.

It is interesting to note that the abundances of RR Lyrae in the bulge range across $-0.3 \geq [\text{Fe/H}] \geq -1.65$, with a mean around $[\text{Fe/H}] = -1$ (Walker & Terndrup 1991); this is significantly more metal poor than the mean abundance for other bulge stars, and is further evidence that most of the RR Lyrae may not be a true bulge population.

**Kinematics:** The dynamics of the gas in the inner bulge also suggests that it is strongly dominated by a bar (Binney et al. 1991). The kinematics of metal poor stars in the bulge generally are found to have hotter velocity dispersions than the metal rich stars (Harding 1996, Minniti 1996, Tiede & Terndrup 1997), although it is possible that this is not entirely a real population difference, but instead a result of projection effects along the line of sight (Tiede & Terndrup 1997). Representative bulge kinematics are given by the results of Harding (1996) for K giants in a field at $(l, b) = (-10^\circ, -10^\circ)$: For a metal rich sample ($[\text{Fe/H}] > -0.9$) he finds $V_{\text{rot}} = 113 \pm 11 \text{ km s}^{-1}$ and an azimuthal velocity dispersion of $59 \pm 12 \text{ km s}^{-1}$, while for metal poor stars ($[\text{Fe/H}] < -1.5$) he finds $V_{\text{rot}} = 29 \pm 21 \text{ km s}^{-1}$ and dispersion $123 \pm 18 \text{ km s}^{-1}$. This difference has been argued by Harding and others to reflect the presence of two populations. The latter stars have kinematics typical of the halo, while Harding’s metal rich giants have kinematics similar to other bulge tracers, e.g., Miras, OH/IR stars, and planetary nebulae. The hot, halo-like kinematics of the metal poor K giants are also found to be characteristic of the bulge RR Lyrae sample (Gratton 1987, Tyson 1991).

**Age:** The age distribution of the bulge population is still not well resolved. Direct measurement of the age of the stellar population by use of the color of the main sequence turnoff is complicated due to uncertainties in the extreme reddening, the spatial distribution along the line of sight, and the large abundance spread in the bulge. One attempt (Holtzman et al. 1993) suggests that there are a significant number of intermediate age ($< 10 \text{ Gyr}$) stars in the bulge. Houdashelt (1995) also finds a best fit isochrone to Baade’s Window stars for an 8.0 Gyr population with $[\text{Fe/H}] = -0.3$. The relative numbers of red clump to red giant stars in the bulge also argues for a $< 10 \text{ Gyr}$ population (Paczyński et al. 1994). However, Ortolani et al. (1995) find that two near solar abundance globular clusters in the bulge, NGC 6528 and NGC 6553, have ages within a few Gyr of the age of halo globular clusters. Moreover, Ortolani et al. , after aligning the red clump (horizontal branch) of these globular clusters and the bulge field stars, note a striking similarity in the luminosity functions that suggests little age difference between the bulge field and the these clusters. Ortolani et al. also find little spread in age for the bulge MSTO stars. This analysis indicates an age for the bulge field stars perhaps similar to that of metal rich disk clusters like 47 Tuc, though a still younger age for the bulge with respect to disk clusters (a difference of 4 Gyr) has been suggested by analysis of Baade’s Window MSTO stars by Fullton (1995).

Minniti (1995, 1996; see also Harris, in this volume) has argued that on the basis of their kinematics, abundance, and concentration – which reflect those of metal-rich bulge giant stars – the metal-rich globulars in the innermost 3 kpc of the Galactic center should rightfully be associated with the bulge, rather than a metal-rich extension of the disk, as suggested by the representations in Figures 16c and 29. However, as Zinn (1996) points out, the apparent match of systemic rotational velocity between K giants and the innermost clusters is strongly influenced by a few clusters in the sample at larger $|l|$ that may well be members of the true disk system. Removing these clusters from the sample, or restricting the sample to the most metal rich clusters, removes any signal of rotation. Unfortunately, these analyses of the inner cluster population are affected by the small number statistics of the known clusters. The issue of the proper population
assignment of the metal-rich, inner globulars, if settled, would be of great value because of the use of the clusters for dating populations. Age determinations for clusters under debate include several (NGC 5927 and NGC 6553) for which significantly younger ages are found compared to other disk globulars (Demarque & Lee 1992, Fullton 1995; but see Ortolani et al. 1995 and above).

A population of stars with rather younger ages is implied by the presence of Miras in the bulge (Glass et al. 1995) with periods longer than those found in globular clusters. The large number of > 300 day period Miras suggests an age for these stars more like the young disk Miras found in the disk near the Sun.

On the other hand, there are also RR Lyrae stars in the bulge, which might suggest that at least some stars in the bulge are old. Lee (1992) looked at the change in the abundance distribution of these RR Lyrae and concluded that the changes were the result of a decreasing age with $R_{GC}$ which suggests an inside-out formation, with the oldest stellar population in the bulge forming about one Gyr before the inner parts of the halo. However, if a majority of the RR Lyrae stars belong to the inner halo, rather than the bulge (as suggested by their density law, velocity dispersion, and metallicity), quite the opposite conclusion can be reached. By removing these stars from the mix, and noting strong trends of kinematics with abundance for bulge giants, Minniti (1996) argues that the bulge may have formed in an ELS-like, dissipational collapse, after formation of the halo. On the other hand, Rich (1996) warns that some of the abundance-kinematics trends might be a product of disk star contamination.

In summary, the age distribution for the bulge obviously spans a large range. Certainly intermediate age stars are present, but it is not clear whether the true bulge population has stars with an age as great as that of the halo, or whether the oldest bulge stars are several Gyr younger than the halo globulars, perhaps with an age similar to the disk clusters. Determining the age distribution of bulge stars is of course critical to fitting it into a sensible timeline with the formation of the other stellar populations in the Milky Way. The confusing situation regarding the mean age of the bulge is partly a result of the fact that it is a repository for much “waste product” (stars and gas) in the evolution of the Milky Way. For example, we have seen (Section 1.2) that clusters with small apogalactica orbits face the quickest demise from bulge shocking; therefore we might expect at least some fraction of the bulge to be cluster débris. In particular, the inner RR Lyrae population could be tracing the remains of a destroyed halo globular cluster population. Clearly it would be interesting to determine the kinematics of these stars to determine if they are on elongated orbits expected of primordial clusters most susceptible to destruction. Note that on such orbits, débris from these clusters would not feel the bar potential long enough to respond to it, and this is consistent with the observations of the RR Lyrae population (Alcock et al. 1998).

4. Putting It All Together: Chemodynamical Pictures of Milky Way Formation

In this section I attempt, via the “population boxes”, to illustrate the history of stellar populations in the Galaxy as the data presented here suggest it. Several cautionary notes are worth mentioning. First, the quick review here must necessarily gloss over details, and even whole areas of Galactic studies (like the wealth of information now available from detailed chemical abundance studies) have been left out. It is hoped that other teachers at the school will fill in some of the blanks.

Second, it would be ill-advised to believe that the ideas presented here enjoy a complete consensus within the community. While I have tried to be fair in presenting what I believe
to be the main themes and general results in each area, it is unavoidable that personal
taste and “leanings” towards certain models creep in. This is always the case when
the answers are still out of reach, and several viable models to explain the data exist.
Much work remains to iron out a number of problems in the ripe, and presently very
lively, research area of Galactic populations. This, then, is a call of encouragement to
the next generation of Galactic astronomers who, armed with bigger telescopes, much
larger surveys, and more realistic chemodynamical models, will no doubt soon (perhaps
already?!) look on the present discussion as hopelessly naïve.

Nevertheless, let me summarize what I believe to be the general themes of our un-
derstanding of Galactic stellar populations, as presented here. The result is presented
schematically as the composite Hodge chemical population box of Figure 31 and the
composite dynamical population box in Figure 32, which are essentially Figures 22 and
26 with the addition of halo populations. Here the boxes are illustrated in projection
from above in order to show details. The SFRs for the disk populations have been taken
directly from Figure 18, and, although the reliability of chromospheric age dating of
disk stars is not free from criticism, the data presented in Figure 18 at minimum give
a flavor for what the disk SFR may have been like. The SFRs shown for the halo and
IPII components are pure speculation. In both figures, the bulge has been left out due
to uncertainties in the age-metallicity-kinematics relations and the definition of the true
bulge population, but it is possible that the evolution of the bulge is qualitatively similar
to (and would therefore overlap with) that of the early disk as shown.

While much larger samples of stars with full chemical and kinematical data are still
needed, the present observational data seem to call for chemodynamical models incorpo-
rating multiple formation processes (collapse, accretion) just to explain what previously
would have been considered the “classical” halo populations of the Milky Way. I illus-
trate the “dual halo” model of the Galaxy with clearly separated populations in the
population boxes. In all populations, a tail to lower abundances, as suggested by ob-
served metallicity distribution functions, is indicated. In these highly schematic figures I
illustrate the flattened halo, the IPII and the oldest parts of the disk as partly contiguous,
though I point out again that it is still not at all clear whether the IPII and flattened
halo are parts of the same population, or whether they are clearly separable populations.
Perhaps the old, flattened halo is the extreme extension of the disk whereas the IPII is
a manifestation of an accretion event. In either case – the IPII as a transitional phase of
a general, dissipational, ELS spin-up collapse, or the IPII as the result of violent heating
of the thin disk – the resultant observed population boxes, shown here, might look the
same. Indeed, this is part of the problem of discriminating between these models.

The dynamical population box shown here is based on the observed properties of the
various stellar populations. If the IPII were formed from an accretion event, then the
stars in that population would have been born with higher log (|V_{rot}|/σ), perhaps closer
to 0.5, and then pumped up to the presently observed value. The figure also conveys the
SZ picture of the halo forming in “fragments” even while the inner parts of the Galaxy
were organized into a flattened, dissipational structure with ELS spin-up. Note that the
dynamics shown are for the “fragment” population as a whole; the internal dynamics of
the individual fragments (dwarf galaxies? globular clusters? both?) would have smaller
velocity dispersions.

Finally, the age scales of Figures 30 and 31 are meant to convey something of a middle
ground between the nearly factor of two range among the various quoted values for the
age of Galactic clusters. If the recent trend towards young globular cluster ages is born
out, then the scale of the abscissas should, to first order, be compressed.
Figure 31. Schematic, composite Hodge population box for stellar populations in the Milky Way. The “box” is observed from above as a contour plot in order to show detail. The contour levels correspond roughly to 0.5 levels in SFR/$<\text{SFR}>$ as shown in Figure 18 for the disk populations, with the lowest contour at 0.5. The bursts from Figure 18 are labeled.

It is a great pleasure to thank Ricky Patterson and my students, Jeff Crane, Johnny Johnson, Chris Palma, and Michael Siegel for their generous assistance with assembling this manuscript. Jeff Crane and Chris Palma went beyond the call of duty in helping to create some of the original figures. I also appreciate valuable conversations with and comments from colleagues, especially Eva Grebel, Kathryn Johnston, Bill Kunkel, Neill Reid, Bob Rood, Bill Saslaw, and Jesper Sommer-Larsen. Neill Reid and Ricky Patterson made numerous valuable suggestions for improving this manuscript.

REFERENCES

Figure 32. Schematic, composite dynamical population box for stellar populations in the Milky Way, with properties as observed today. As in the previous figure, the “box” is observed from above as a contour plot in order to show detail. The contour levels are the same as those shown in the previous figure. Note the use of $|V_{rot}|$ to accommodate the possibility of a retrograde “young halo”.

Berendzen, R. & Hart, R. 1973 JHA iv, 73.
Carney, B. W. 1993 In Galaxy Evolution: The Milky Way Perspective, ed. S. R. Majewski,
92, p. 307. ASP.
Hanson, R. B. 1989 BAAS 21, 1107.
Harris, W. E. 1991 ARAA 29, 543.
Heatherington, N. S. 1975 JHA vi, 115.
Hodge, P. 1989 ARAA 27, 139.
Kapteyn, J. C. 1906 Plan of Selected Areas, Hoitsema Brothers.
Lindblad, B. 1927 MNRAS 87, 553.
REID, I. N. 1999 *ARAA, in press*.


Struve, F. G. W. 1847 *Etudes d’astronomie stellaire*, St. Petersburg Imperial Acad. Sci.


