30

GERARD F. GILMORE*

Stellar populations and dynamics in the Milky Way galaxy

Our Galaxy offers a unique opportunity to deduce the important physics involved in galaxy formation from observations of those old stars that were formed at the time of the formation of the Milky Way, and whose present properties contain some fossil record of the Galaxy's history. Only in the Milky Way galaxy and its immediate neighbour satellite galaxies can one obtain the true three-dimensional stellar spatial density distributions, stellar kinematics and stellar chemical abundances. Knowledge of how stars move and how they are distributed in space measures the Galactic gravitational potential, including its dark matter content, while knowledge of stellar kinematics, ages and chemistry constrains the star formation and gas accretion history.

The scientific goal is to build on the powerful concept of stellar populations, which basically distinguishes old, metal-poor stars (Population II) from younger, more metal-rich stars (Population I), to include the rich complexity of the real Universe.

No single section of the electromagnetic spectrum provides the 'best' view of the Galaxy. Rather, all views are complementary. However, some views are certainly more representative than others. The most fundamental must be a view of the entire contents of the Galaxy. Such a view would require access to a universal property of matter that is independent of the state of that matter. This is provided by gravity, since all matter, by definition, has mass. Mass generates the gravitational potential, which in turns defines the size and the shape of the Galaxy. While the most reliable and comprehensive, such a view is also the hardest to derive. Kinematics and distance data are, however, the closest approach to such a view that is possible. An ideal astrophysicist would have astrometric eyes!

Available information strongly suggests that the Galactic extreme Population II subdwarf system formed early, though the duration of its aggregation into the proto-Galaxy remains unclear. This subdwarf system now forms a flattened, pressure-supported distribution, with axial ratio ~2:1. The thick disk formed close in time to the subdwarf system, with at least the metal-poor tail of the thick disk being comparable in age to the globular cluster system. The thick disk is probably chemically and kinematically discrete from the Galactic old disk – by which we mean those stars of the thin disk with age greater than a few billion years – implying a discontinuous Galactic history. The inner Galaxy is mostly old, almost certainly barred, but as yet remains to be studied in detail, especially in the innermost regions. Importantly, recent dynamical analyses lead to the conclusion that there is no statistically significant amount of non-luminous mass in the solar neighbourhood, and hence no evidence for dissipative dark matter.

We consider in turn the general questions of galaxy formation which the science of stellar population studies aims to address, followed by discussion of star count and infrared studies of Galactic structure, and astrometric and kinematic studies of the Galactic mass distribution. The primary space missions which have contributed to these subjects are the Hubble Space Telescope (HST), whose excellent spatial resolution allows study in crowded regions, whose dark sky background allows the study of very faint objects and, most importantly here, allows reliable image distinction between stars and resolved objects (galaxies); the sequence of infrared observatories, which have allowed observations through the optically thick dust obscuring the inner Galaxy, Infrared

* University of Cambridge, United Kingdom

Gerard F. Gilmore, The Century of Space Science, 699–719

Astronomical Satellite (IRAS), Diffuse Infrared Background Experiment (DIRBE) on the Cosmic Background Explorer (COBE) and the Infrared Space Observatory (ISO); and the High Precision Parallax Collecting Satellite (Hipparcos), arguably ESA's most substantial contribution to astrophysics in the twentieth century.

THE FORMATION OF DISK GALAXIES

Current understanding of the formation and early evolution of disk galaxies allows a description of the important physical processes at various levels of complexity and generality, and illustrates what we have learned, what we are attempting still to learn, and the relative importances of surveys, targeted research projects and newer space-based studies. At one extreme, one simply considers the global evolution of a gas cloud, and assumes that mean values of relevant parameters suffice for an adequate description of generic properties. Alternatively, one gives up general applicability and instead adopts specific numerical values for those parameters that quantify the important physics, attempting a detailed confrontation of model predictions with observed stellar populations. The relation of any model prediction to detailed observations at a single radius in a specific galaxy clearly needs to be considered with some care. Mindful of this caveat, we outline here the most important timescales and physical processes that are likely to play a role in the determination of the observable properties of galaxies like the Milky Way.

General models of dissipational disk galaxy formation

The existence of cold, thin Galactic disks has strong implications for galaxy formation. To see this, consider a standard picture whereby galaxies form from growing primordial density perturbations, which expand with the background Universe until their self-gravity becomes dominant and they collapse upon themselves. Were there to be no loss of energy in the collapse, and neglecting angular momentum, the transformation of potential energy into thermal (kinetic) energy would lead to an equilibrium system with final radius equal to half its size at maximum expansion, supported by random motions of the constituent particles. Thus an equilibrium, purely gaseous proto-galaxy should have temperature

$$T \equiv T_{\rm virial} \sim \frac{GMm_{\rm p}}{kR} \tag{1a}$$

and a stellar proto-galaxy should equivalently have velocity dispersion

$$\sigma^2 \sim T_{\rm virial}\frac{k}{m_{\rm p}} \tag{1b}$$

where k and $m_{\rm p}$ are the Boltzmann constant and the mass of the proton, respectively. Numerically, $T_{\rm virial} \sim 10^6\, R_{50}^{-1}\, M_{12}\,{\rm K}$ for gravitational (half-mass) radius, R, in units of 50 kpc, and mass M in units of $10^{12}\, M_\odot$. Since the disks of spiral galaxies are cold, with $T \ll T_{\rm virial}$, energy must have been lost. Since this lost energy was in random motions of individual particles, the only possible loss mechanism is through an inelastic collision, leading to the internal excitation of the particles, and subsequent energy loss through radiative de-excitation. Clearly, particles with small cross-section per unit mass for collisions, such as stars, will not dissipate their random kinetic energy efficiently, so that dissipation must occur prior to star formation, while the galaxy is still gaseous. The virial temperature of a typical galaxy-sized potential well is $T_{\rm galaxy} \sim 10^6$ K, with a corresponding one-dimensional velocity dispersion of ~100 km s^{-1}.

The physical conditions in the Universe at the epoch of galaxy formation (redshift ~a few), as deduced from observations of quasar absorption lines (the Gunn–Peterson test for neutral hydrogen), are such that hydrogen is ionized, and correspond to temperatures of the proto-galactic gas of ~10^4 K, with a sound speed of only ~10 km s^{-1}. The collapse of this gas in Galactic potential wells will thus induce supersonic motions, and lead to both thermalization of energy through radiative shocks and subsequent loss of energy by cooling. It is this conversion of potential energy, first to random kinetic energy as described by the virial theorem, and then to radiation via atomic processes, the net result of which is an increase in the binding energy of the system, that is termed dissipation.

The rate at which excited atoms can cool obviously places a fundamental limit on the amount and rate of dissipational energy loss, and hence on the maximum rate at which a gas cloud can radiate its pressure support and collapse. A convenient measure of this timescale is the cooling time of a gas cloud, which is the time for radiative processes to remove the internal energy of the cloud. Defining the cooling rate per unit volume to be $n^2\Lambda(T)$, where n is the particle number density, and where the functional form of Λ is determined by the relative importances of free–free, bound–free and bound–bound transitions and is thus an implicit function of the chemical abundance, gives

$$t_{\rm cool} = \frac{3nkT}{n^2\Lambda(T)} \propto \frac{T}{n\Lambda} \tag{2}$$

It is usually of most interest to compare this timescale with the global gravitational free-fall collapse time of a system, which is the time it would take to collapse upon itself if there were no pressure support. This timescale depends only on the mean density of the system, and is given by

$$t_{\rm ff} \sim 2 \times 10^7 n^{-1/2}\ {\rm yr} \tag{3}$$

The term 'rapid' is often used to describe evolution that occurs on about a free-fall time.

The important mass scale of any condensations is set by gravitational instability, quantified by the Jeans mass. The Jeans mass is that minimum mass at which gravity overwhelms pressure so that density perturbations of mass $M \gtrsim M_J \sim 10^8\, T_4^{3/2}\, n^{-1/2} M_\odot$ are unstable and collapse upon themselves, where the numerical factor is for temperature T in units of 10^4 K and number density n in units of particles per cubic centimetre.

The discussion above is based on an extremely idealized model of a proto-galaxy, in that only the *global* cooling and collapse timescales of a *uniform* gas cloud are considered. No analytic descriptions of more plausible models exist.

The above discussion can say nothing about when or how local Jeans-mass condensations actually form stars; the inherent assumption is that cooling is necessary and sufficient for efficient star formation, though the critical distinction between global and local timescales is rarely made explicit. However, it is clear that the existence of gaseous disks requires the star-formation efficiency to be low during the early stages of disk formation. A realistic discussion of galaxy formation must consider the hydrodynamics of the gas in a proto-galaxy. The general conclusion from available studies is that, while it is possible to build models which are somewhat like observations, it is necessary to specify the most sensitive parameters (viscosity and, in effect, the star formation rate) in an *ad hoc* way. Hence the need to be guided by observations.

The angular momenta of galaxies that formed in environments of different density might also be expected to differ. During the build-up of structure, initially strongly bound particles lose both energy and angular momentum, while the weakly bound particles gain energy (become more weakly bound) and also gain angular momentum. There is overall alignment of the angular momentum vector of different shells in binding energy. Disks of spiral galaxies would then form without significant angular momentum transport, provided the baryons remain gaseous until the virialization of the dark halo, and shock-heating, as described earlier, would homogenize the gas. The predictions of these models could be tested in detail if we knew the angular momentum distribution of the outer spheroid of our Galaxy; all we know at present is that the kinematically selected subdwarfs in the solar neighbourhood have a lower specific angular momentum than the disk stars, by roughly a factor of five, and that the metal-poor globular cluster system is consistent with zero net rotation to galactocentric distances of ~30 kpc.

The angular momentum distribution of the material destined to form the disk controls the range of galactocentric radii over which infall occurs at a given epoch, and the duration of the infall at a given location. Thus models of disk chemical and dynamical evolution that appeal to continual infall must also satisfy angular momentum constraints.

Specific models of Milky Way galaxy formation

The most widely referenced model of the formation of our Galaxy is that of Eggen, Lynden-Bell and Sandage (1962; henceforth ELS), which was developed primarily to understand their observations of the kinematics and chemical abundances of stars near the Sun. Certainly the most important effect of the ELS analysis was to emphasize that quantitative conclusions about the epoch of galaxy formation could be derived from observations of stellar abundances and kinematics near the Sun today – a task in which we are still engaged.

The ELS model requires the stellar spheroid to have formed during a period of rapid collapse of the entire proto-galaxy, after which the remaining gas quickly dissipated into a metal-enriched cold disk, in which star formation has continued until the present. The ELS model was designed to provide conditions under which the oldest stars populated radially anisotropic orbits, while stars that formed later had increasingly circular orbits, in accordance with data which implied that the most metal-poor stars, assumed to be the oldest stars, were on more eccentric, lower angular momentum orbits than the more metal-rich stars. This model is based on two crucial assumptions: first, that a pressure-supported, primarily gaseous galaxy (where $T = T_{\mathrm{virial}}$) is stable against star formation, in which case the *global* cooling time is the shortest timescale of interest, and thermal instabilities must be suppressed; and second, that stellar orbits cannot be modified to become more radial after the formation of the star.

If the first assumption were valid, the observed high-velocity stars must have formed from gas clouds that were not in equilibrium in a pressure-supported system. If the second assumption above were valid, these clouds formed stars while on radial orbits at large distances from the Galactic centre. Thus, in this picture, these clouds must have turned around from the background universal expansion, and be collapsing towards the centre of the potential well. Hence, the oldest stars of the Galactic spheroid must have formed as the proto-galaxy coalesced. To determine the rate of the collapse, ELS analysed the evolution of the radial anisotropy of a stellar (or gas cloud) orbit as the Galactic potential changed, and showed it to be approximately conserved during a slow collapse but to become more radially anisotropic in a fast collapse. They argued against a slow collapse on the grounds that such a collapse requires tangentially biased velocities (remember that pressure support has been excluded by assumption), and this tangential bias will be unaffected by the resulting slow changes of the gravitational potential. The observed radial anisotropy of the stellar orbits then implies an initially radially biased velocity ellipsoid, while the calculations of ELS show that such a velocity ellipsoid will have become more radially anisotropic during collapse. Hence, ELS deduced that the gas clouds were in free-fall radial orbits, and that the consequent collapse must have been rapid, with 'rapid' in this sense meaning that the timescale for collapse is comparable to an orbital or a dynamical timescale, which is of the order of 10^8 years. It should be noted that Isobe (1974) came to the opposite conclusion from his analysis of the ELS data, and favoured a slow collapse, while Yoshii and Saio (1982) augmented the ELS sample and also concluded that the halo collapsed over many dynamical times.

Clearly if either of ELS's assumptions were violated, there need be no correlation between the *time* of a star's formation – which they infer from a star's metallicity – and its *present* orbital properties. In a non-rotating pressure-supported system, all stars formed would be on highly radial orbits, as a star has too small a surface area to be pressure supported by the gas. As mentioned above, assumptions about star formation in pressure-supported systems must be treated as *ad hoc* until we understand better the physics of star formation, so conclusions based on such assumptions are at best uncertain. If their second assumption were violated, then the stars that are now the high-velocity stars near the Sun could have originated from more circular orbits inside that of the Sun, and have present orbital properties that depend only on dynamical processes subsequent to their formation. The realization that a forming galaxy undergoes changes in its gravitational potential which are of order the potential itself (violent relaxation) means that stellar orbits can be modified considerably.

Later *N*-body models (e.g. May and van Albada 1984) for systems in which dissipation does not play a major role show that the final state of the collapsed system depends both on the degree of homogeneity and on the temperature of the initial state: clumps cause angular momentum and energy transport. Violent relaxation never goes to completion, so that final and initial orbital binding energies and angular momenta are correlated, the interior regions becoming more centrally concentrated and the outer regions being puffed up. The typical final steady-state velocity distribution is highly anisotropic outside (roughly) the half-light radius, and more isotropic within that radius. If violent relaxation were completely efficient, all systems would reach the same final state with isotropic velocity distribution. In the Galaxy, the spheroidal half-light radius is ~3 kpc, well inside the Sun's orbit. Thus the expected velocity distribution of old stars near the Sun after virialization of the spheroid is anisotropic, as observed by ELS, even though the dynamical evolution of the system is not as they envisage, and a correlation between kinematics and age is no longer an inevitable conclusion. One might, for example, imagine a situation where later (rapid) collapse of either the disk or the dark halo, or the merger of a few large substructures, could lead to the rapid dynamical evolution of a central spheroidal component which had previously formed on a longer timescale. Models of this type have yet to be studied in detail.

It is the continuing attempt to quantify the metallicity–kinematics distribution function of stars, with sufficient accuracy and large numbers to address these questions, that motivates projects such as Hipparcos, and the use of HST, COBE, IRAS and ISO to quantify the true distribution of stars in space. We now consider that specific continuing challenge.

THE SPATIAL STRUCTURE OF THE MILKY WAY GALAXY

Counting stars is one of the few truly classical scientific techniques used to study high-latitude (and therefore low-obscuration) Galactic structure. Early work in this subject is well reviewed by Paul (1993). The extensive data set and understanding available at that time is reviewed by Blaauw and Schmidt (1965). Relatively little further progress was achieved until the new deep, high-quality data of Ivan King and collaborators at Berkeley became available in the late 1970s. The application of computer modelling to these data by van den Bergh (1979) led to a considerable resurgence of interest, continuing to the present. An accessible overview of the recent advances in the subject is provided by Croswell (1995), while the continuing level of research activity generates several conferences per year. An important specific example, which includes both historical and research articles, is the proceedings of IAU Symposium 164, 'Stellar Populations', held on the 50th anniversary of Walter Baade's publication of the original concept (van der Kruit and Gilmore 1994).

The fundamentals of star-count analyses

The number of stars, N, countable in a given solid angle to a given magnitude limit, m, is given by a simple linear integral equation often known as 'the fundamental equation of stellar statistics':

$$N(m) = \int \Psi(M_{\mathrm{v}}, \boldsymbol{x})\, D(M_{\mathrm{v}}, \boldsymbol{x}) \mathrm{d}^3 x \tag{4}$$

where $\Psi(M_{\mathrm{v}}, \boldsymbol{x})$ is the distribution function over absolute magnitude and position, $D(M_{\mathrm{v}}, \boldsymbol{x})$ is the stellar space density distribution, and d^3x is a volume element. This (Fredholm) equation is rarely invertible, being ill-conditioned. In general, the luminosity function is too broad to allow any

http://www.springer.com/978-0-7923-7196-0

The Century of Space Science
Bleeker, J.A.; Geiss, J.; Huber, M. (Eds.)
2001, XLIX, 1846 p., Hardcover
ISBN: 978-0-7923-7196-0