

NEAR FIELD COSMOLOGY – THE ORIGIN OF THE GALAXY AND THE LOCAL GROUP

JOSS BLAND-HAWTHORN

Sydney Institute for Astronomy, University of Sydney, NSW 2006, Australia; jbh@physics.usyd.edu.au

KEN FREEMAN

Mount Stromlo Observatory, Australia National University, Weston Creek, ACT 2611, Australia; kcf@mso.anu.edu.au

Draft version April 2, 2013

ABSTRACT

The Galaxy has built up through a process of accretion and merging over billions of years which continues to this day. Astronomers are now embarking on a new era of massive stellar surveys over the coming decade. These campaigns will derive three-dimensional space motions and heavy element abundances for millions of stars throughout the Galaxy and its neighbours. The new observations will reveal signatures of the formation and early evolution of the Local Group; this is what we mean by ‘near field cosmology.’ We set this new course of study within the context of fossil signatures from galaxy surveys and the high redshift universe. We discuss the complex relationship between baryons and dark matter over cosmic time, and introduce a synthetic framework that will allow both numerical simulations and the impending data deluge to be compared. We also include relevant source materials for the young near-field cosmologist and some historical perspectives.

Key Words: Cosmology, Local Group, Stellar Populations, Stellar Kinematics

1. PROLOGUE

In an earlier review, we outlined the concept of near field cosmology to emphasize the fact that there are ancient signatures all around us today providing evidence of the formation processes that led to the Galaxy and the Local Group (Freeman & Bland-Hawthorn 2002). We see ancient stars around us in the old thin disk, the thick disk, the stellar halo, the inner bulge, and in satellite dwarf galaxies. About half of all stars in the Galaxy today formed before a redshift $z \sim 1$. The Hubble Space Telescope (HST) probes these early formation epochs directly during a time when gas accretion, star formation and feedback processes were greatly enhanced. But how do we connect the past with the present? There are no easy answers. Here we revisit our earlier review, extend our discussion to observable fossils from the Big Bang to the present day, with an updated discussion of the complete baryon inventory.

The desire to connect present-day galaxies and the distant Universe has a long history, long before it was possible to compare both epochs directly. We now speak of galactic and stellar archaeology. Eggen, Lynden-Bell & Sandage (1962) were the first to show that it is possible to study galactic archaeology using stellar abundances and stellar dynamics; this is probably the most influential paper on the subject of galaxy formation. In 1966, Fred Hoyle wrote “It is not too much to say that the understanding of why there are these different kinds of galaxy, of how galaxies originate, constitutes the biggest problem in present day astronomy. The properties of individual stars that make up the galaxies form the classical study of astrophysics, while the phenomena of galaxy formation touches on cosmology. In fact, the study of galaxies forms a bridge between conventional astronomy and astrophysics on the one hand, and cosmology on the other.” (Hoyle 1966).

In the Local Group, we can see the imprint of processes that date back to at least $z \sim 5$ and maybe earlier still. Reliable ages do not exist for individual stars beyond the Sun but at least some of the most metal-poor stars discovered so far may date to early generations immediately following the first stars. We are coming into a new era of galactic investigation, in which one can study the fossil remnants of the early days of the Galaxy in a broader and more focussed way, not only in the halo but throughout the major luminous components of the Galaxy. The goal of these studies is to reconstruct as much as possible of the early galactic history. We will review what has been achieved so far, and point to some of the ways forward.

But first we review important fossils from the high-redshift universe ($z \gtrsim 10$) and provide a summary of some key findings from galaxy redshift surveys ($0 \lesssim z \lesssim 10$). These provide a context for our later discussions on the formation and evolution of the Local Group.

2. FAR FIELD COSMOLOGY

It is now firmly established that the Universe underwent a period of rapid exponential expansion (acceleration) about a picosecond after the Big Bang. This was the birth of space-time, the forces of nature, and the Higgs field¹ that imparted mass to energetic particles and dark matter. The Universe today is isotropic, homogeneous and flat on the largest scales. A mathematical description of the expanding Universe is given by the Friedmann-Robertson-Walker (FRW) metric of general relativity although a useful description is provided by Newtonian analysis (Wright 2003). Using either approach, for an appropriate choice of density and pressure, it is possible to arrive at Hubble’s law, $v = H_0 d$, where v is the recession velocity² and d is the proper distance. This simple formula describes the linear expansion of the Universe (Lemaître 1927; Hubble & Humason 1931) for which modern estimates of Hubble’s constant fall within $H_0 = 70 \pm 4 \text{ km s}^{-1}$ (e.g. Freedman et al 2001).

It therefore came as a surprise to (most of³) the astronomical community that distant supernovae appear to be accelerating from us (Reiss et al 1998; Perlmutter et al 1998), providing clear evidence for a second period of inflation for the Universe as a whole. The acceleration is powered by a mysterious source of ‘dark energy’ that has dominated the energetics of the Universe only relatively recently ($z \lesssim 0.5$; Peacock 2000). Dark energy is the name we give to Einstein’s cosmological constant Λ or something that acts like it. As we shall see, this secondary inflationary phase is already affecting the way in which structure forms throughout the Universe. The evidence for the existence of the dark sector is strong, but its properties are only loosely understood.

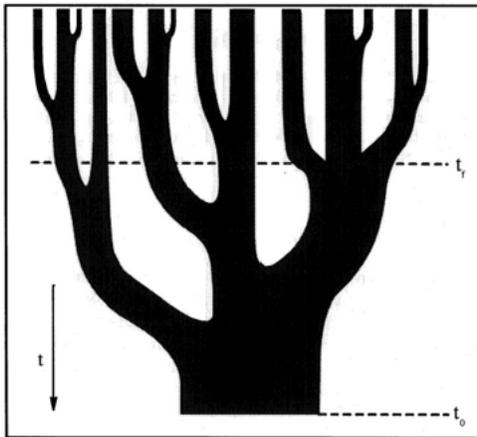


FIG. 1.— A merger tree depicting the hierarchical merging of dark matter structures over cosmic time t (Lacey & Cole 1993); the time t_0 corresponds to a dark matter halo today. We can identify a time in the past t_f when the halo first formed.

The first inflationary phase diluted all previous structure except for quantum fluctuations which cannot be smoothed out. These primordial fluctuations grew in time to become the seeds of all structure formation (Liddle & Lyth 2000). As the Universe expanded, dark matter across the hierarchy began to collapse to form larger structures (Fig. 1), and the cosmic radiation began to cool (Davis et al 1985). By a redshift of $z \sim 3000$, the Universe cooled enough to allow electrons and protons to form atoms. This was the ‘recombination era’ when a fog of atomic hydrogen and helium permeated the Universe for the first time.

In the 1940s, Gamow, Alpher and Herman began to consider the consequences of a hot expanding Universe and realised that the Universe must be filled with cool radiation today, a relic of these early times. This ancient signal was first detected in the 2.3K rotational temperature of cyanogen (McKellar 1940; cf. Roth et al 1993) although was not recognised as such until Penzias and Wilson (1965) detected the microwave signal at a temperature of 2.7K using a horn antenna. This relic radiation is known as the cosmic microwave background (CMB) and provides us with our most powerful probe of the high-redshift universe.⁴

It is instructive to ask what is the source of the CMB photons? During the ‘recombination era,’ any excess energy from the recombination of electrons and protons radiates away as photons. But these photons made a negligible contribution to the CMB. A viable source of photons must have occurred at an earlier epoch. In the very early Universe, during the so-called ‘baryogenesis era,’ electrons and positrons annihilated to produce 511 keV photons. Interestingly, the neutrinos were not heated by this process such that their temperature today is predicted to be 1.9K rather than 2.7K characteristic of the CMB.

The baryogenesis era also accounts for one of the most extraordinary fossils from the inflationary epoch – matter-antimatter asymmetry. The universe began with equal amounts of matter and antimatter, and yet today matter dominates over antimatter by up to ten orders of magnitude, the same ratio by which photons dominate over baryons

¹ It is presently believed that the Higgs field is distinct from the inflaton field that drove inflation (Guth 1997).

² We note with interest that, exactly 100 years ago, the first galaxy radial velocity was measured by Slipher (1913). Two years later, the first ‘rotation curve’ was measured for NGC 4594 (Slipher 1914), long before it was understood what these observations were telling us. We refer the reader to an excellent review by Peacock (2013) on the remarkable achievements of V. Slipher that have been largely overlooked.

³ Few can claim to have foreseen $\Lambda > 0$ from their published work (e.g. Efstathiou et al 1990; White et al 1993; Yoshii & Peterson 1995).

⁴ Gravity wave (Abbott et al 1993) and neutrino (An et al 2012) experiments can also provide constraints on the post-inflationary epoch but these endeavours are in their infancy.

today. The reason for the spectacular asymmetry is not known, but it may involve a charge conjugate-parity (CP) violation in quark/anti-quark interactions, a complex process that will soon be explored directly with neutrino/anti-neutrino experiments (An et al 2012).

2.1. *The cosmic microwave background*

The CMB is a redshifted relic of our hot beginnings taking us back to a time within 99.99% of the age of the Universe (Fig. 2). See Lineweaver (1999) for an exceptionally intuitive discussion of the CMB. The CMB is of immense importance in our understanding of the early universe because it allows us to see the imprint of dark matter fluctuations within 370,000 years of the Big Bang, i.e. a redshift of $z \approx 1100$. As we travel back in time, the Universe was smaller, hotter and denser. Anyone who has tried to see through mist with car headlights will understand what is meant by the ‘surface of last scattering’ – the light does not penetrate very far into the mist. And so it is with the CMB which arises from the ‘surface of last scattering’ that we can penetrate with microwave photons. Our view back to this time is optically thin and so we are probing large-scale structures when the Universe was a billion times smaller by volume.

The CMB is highly thermalized and therefore a near-perfect isotropic black body at a temperature of 2.725 ± 0.002 K. The spectrum is described by the Planck function as was demonstrated in spectacular fashion by the Cosmic Background Explorer (COBE) satellite. When the Principal Investigator John Mather first showed the spectrum at the American Astronomical Society in January 1990, he received a standing ovation. This remains the most precise Planck spectrum across all fields of applied physics. Consistent with a flat universe, wherever we look across the sky, the mean temperature is the same to within roughly 0.1%. But stamped onto this uniform signature is the effect of our motion through the Universe: CMB radiation is slightly warmer in the direction of our motion, and slightly cooler in the opposite direction. From this signal, we measure our motion to be 552 ± 6 km s⁻¹ in the direction of Hydra.

When the CMB dipole signal is removed, small temperature fluctuations are observed across the sky with amplitudes less than 0.01%. These are the imprints of early structure formation. Before recombination, the free electrons had scattered and trapped the thermal radiation in a photon-baryon fluid; the abrupt elimination of these free electrons allowed the thermal radiation to move nearly without scattering, and so the tiny variations in temperature trace variations in the local potential at $z \approx 1100$. When we form a power spectrum of the CMB variations, a set of ‘acoustic peaks’ can be seen on angular scales below 2°, the projected horizon scale. These are due to sound waves within the photon-baryon fluid as it reacts to the presence of dark matter fluctuations. Lineweaver (1999) provides a marvellous series of illustrations to describe how the acoustic peaks came about. Structure on angular scales greater than 2° dates back to the first picosecond after the Big Bang (inflation), and therefore constitute our most ancient fossils from the early universe at the present time.

It is often overlooked that the dark matter fluctuations are essential to our very existence (unlike dark energy)! Before $z \sim 1100$, the photon-baryon fluid was subject to strong diffusion (Silk damping) due to viscosity and thermal conductivity. This allowed photons to diffuse from hot regions into cold regions wiping out any fluctuations on scales less than about 10 Mpc as observed today. But fortunately for us, the dark matter hierarchy continued to evolve and thereby preserved a semblance of structure. At a later time, this mass spectrum of fluctuations was to drive the formation of the first stars and first galaxies. Precision measurements of the distribution of CMB radiation, most recently by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite (Komatsu et al 2011) are in beautiful agreement with the relativistic theory of how the present concentrations of mass in the Universe grew out of the distribution of dark matter at the time of radiation breaking free from the photon-baryon fluid. The baryon acoustic oscillations (BAO) detected in the CMB have now been detected in galaxy redshift surveys at low redshift (Percival et al 2001; Eisenstein et al 2005; Cole et al 2005; Blake et al 2011).

2.2. *The first stars*

How did the first stars form? Since Big Bang nucleosynthesis – which takes us back to the first minute of cosmic time – only produced elements up to boron, the existence of metals (C, N, O...) is proof enough that something dramatic took place after the recombination era. The first stars were unique to their time – very massive, free of elements heavier than boron (metal free), and very short-lived (Yoshida et al 2004; Gao & Theuns 2007). Such a star has never been observed directly but we live in hope of identifying its chemical signatures in low-mass stars that formed in the second or subsequent generations.

A useful summary of how the first stars formed is given in a highly readable, short text by Loeb (2010); a more extensive review is given by Bromm & Yoshida (2011) and Karlsson, Bromm & Bland-Hawthorn (2013). By a redshift of $z \sim 200$, gas began to accrete through gravitational instability towards local dark matter concentrations (Barkana & Loeb 2007). Whether a forming gas cloud is able to collapse is a competition between gravity and internal pressure. The increase in gas density towards the centre of the cloud generates a pressure (sound) wave that propagates outwards. The pressure wave attempts to counteract the pull of gravity but if the wave does not have sufficient time to cross the cloud, the cloud will collapse. This situation is usefully described by the Jeans mass given by

$$M_J = \frac{4}{3} \pi \langle \rho \rangle r_J^3 \quad (1)$$

where $\langle \rho \rangle$ is the mean density within a radius r_J . If the gas is collapsing onto a dark matter fluctuation, which is assumed in most contemporary models, the mass density includes the combined contribution of the gas and dark

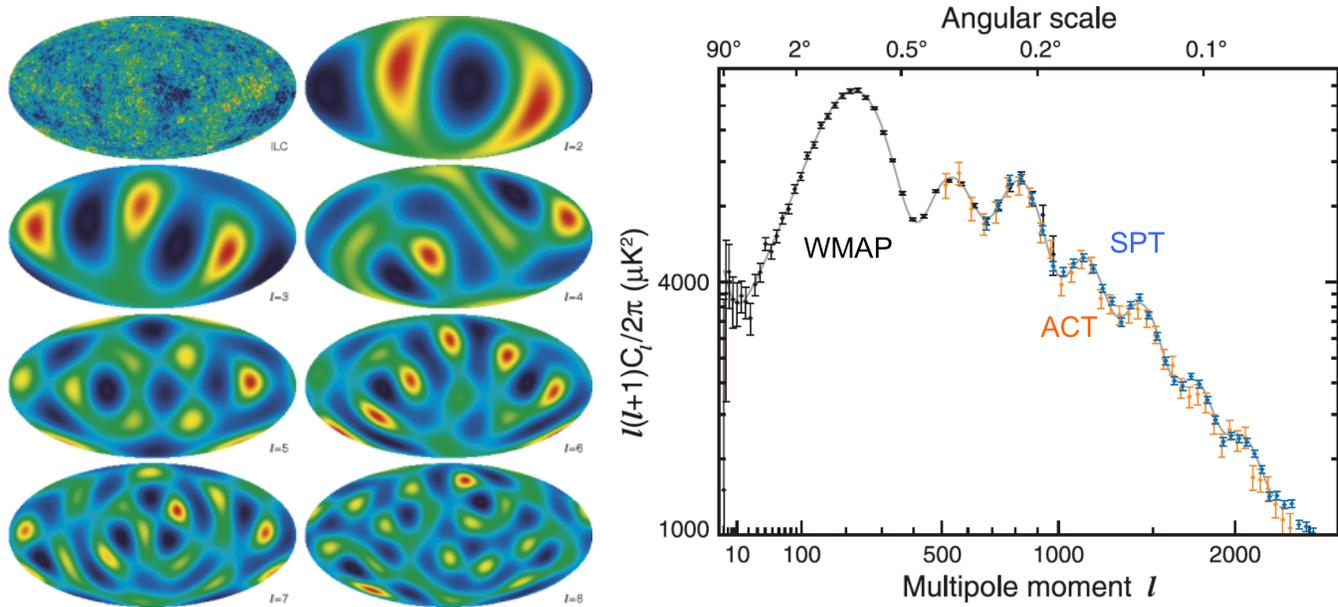


FIG. 2.— (Left) Cosmic microwave background (CMB) observed by the WMAP experiment separated into its fundamental and higher harmonics; the dipole contribution due to the Sun’s motion has been subtracted. (Right) 9 year analysis of the angular power spectrum where the fundamental harmonic is clearly visible. The declining tail – due to ‘Silk damping’ in the early universe (§2.1) – is compiled from observations by the South Pole Telescope (SPT) and the Atacama Cosmology Telescope (ACT). [We acknowledge the WMAP team for this figure.]

matter. If the gas temperature is roughly the temperature of the CMB, $M_J \sim 10^5 M_\odot$ independent of redshift (Peebles 1993; Haiman, Thoul & Loeb 1996).

The first stars – so-called Pop III – are likely to have formed roughly 200 Myr ($z \sim 20$) after the Big Bang (Bromm et al 1999; Abel et al 2002) when the primordial gas was first able to cool and collapse onto small dark matter ‘mini haloes’ with masses of about $10^6 M_\odot$. Their large masses ($> 30 M_\odot$) arose from the limited cooling channels available to primordial gas, more specifically, cooling through radiation from simple molecules. It is now thought that Pop III star formation consisted of two distinct modes: one where the primordial gas collapses into a dark matter minihalo via H_2 cooling (Pop III.1), and one where the metal-free gas becomes significantly ionized prior to the onset of gravitational runaway collapse via HD cooling (Pop III.2).

Pop III.1 mode. The formation of the first stars is much simpler than the corresponding star formation process in the local Universe where the environment of giant molecular clouds is extremely complex (Bromm & Yoshida 2011). The Universe had no heavy elements, and therefore no dust to complicate the physics of cooling and opacity. The early post-recombination Universe was devoid of ionizing radiation and strong drivers of turbulence, such that the collapse into the minihalos assisted by H_2 cooling proceeded in a rather quiescent fashion.

The formation of the first stars inside minihalos provides us with a well-posed problem which can be tackled through rigorous computational studies. Since the late 1990s, several groups have simulated the formation of the first stars with sophisticated numerical algorithms with either smoothed-particle hydrodynamics (SPH) or adaptive-mesh refinement (AMR) techniques (Abel et al 2000, 2002; Bromm et al 1999, 2002; Yoshida et al 2006; O’Shea & Norman 2007, 2008). These calculations have converged on key parameters describing the formation of the first stars.

The most important result, where there is general agreement, is that Pop III.1 stars were predominantly massive. To first order, this can be understood as the consequence of a large Jeans mass in gas that cools only via H_2 . Prior to undergoing runaway collapse, the primordial gas settles into what is sometimes termed a quasi-hydrostatic ‘loitering’ state. This state is characterized by temperature and number density $T_{\text{ch}} \simeq 200 \text{ K}$ and $n_{\text{ch}} \simeq 10^4 \text{ cm}^{-3}$. Recent numerical simulations indicate that Pop III.1 stars formed with characteristic masses of $M_\star \gtrsim 30 M_\odot$.

It is possible that the first stars formed with a range of masses. The initial mass function (IMF) gives the number of stars formed per unit mass, N_\star , where present-day star formation is often described with a power-law with index β : $dN_\star/dm_\star \propto m_\star^{-\beta}$, or a sequence of such power-laws (Kroupa 2001). For Pop III, we do not know the complete functional form of the underlying IMF with any certainty. A rough estimate of the outcome of the star formation process can be made by focusing on the average mass: $\langle m_\star \rangle \propto \int m_\star dN_\star$. We refer to this as the ‘characteristic mass’ as it describes the typical result of star formation, with the understanding that some stars form with lower, and some with higher masses. It is not yet feasible to realistically simulate the assembly of an entire Pop III star, starting from realistic cosmological initial conditions. Recently, *ab initio* calculations have traced the evolution up to the point where a small protostellar core has formed at the center of a minihalo (Yoshida et al 2008).

This initial hydrostatic core has a mass $m_{\text{ch}} \sim 10^{-2} M_\odot$ very similar to present-day Pop I protostellar seeds. The subsequent growth of the protostar through accretion, however, is believed to proceed in a markedly different way. In

the early Universe, protostellar accretion rates are believed to have been more enhanced, due to the higher temperatures in the star forming clouds, which in turn is a consequence of the limited ability of the primordial gas to cool below the ~ 100 K accessible to H_2 cooling. The higher accretion rates, together with the absence of dust grains, and the reduced radiation pressure that could in principle shut off the accretion, conspire to yield heavier final stars in the early Universe. Estimates for the masses thus built up are somewhat uncertain, but a rough range is $m_* \sim 30 - 100 M_\odot$ although a higher mass cut-off is possible.

Recent simulations have shown that Pop III.1 stars may have formed as members of a binary or small multiple system (Turk et al 2009; Stacy et al 2010). Here the evolution is followed beyond the initial collapse of the first high density peak where one sees the emergence of a compact disk around the first protostar. This disk is gravitationally unstable and fragments into a dominant binary, possibly with a few more lower-mass companions. These simulations have not yet been able to reach the asymptotic end state, where all available mass is either accreted or permanently expelled. The reason is that radiative feedback processes in partially optically thick material cannot be neglected once a protostar has grown to $10 M_\odot$, when full radiation-hydro simulations are required (Hosokawa et al 2011; Stacy et al 2012). Currently, the binary fraction of Pop III stars is a major unknown.

Pop III.2 mode. While H_2 cooling drives the formation of Pop III.1, the HD molecule can also play an important role in cooling primordial gas in situations where the gas experiences strong ionization. This is largely due to the high abundance of electrons which serve as a catalyst for molecule formation (e.g. HD) in the early Universe (Shapiro & Kang 1987). The temperature can then drop well below 200 K. In turn, this efficient cooling may lead to the formation of primordial stars with characteristic masses of the order of $m_* \sim 10 M_\odot$, so-called Pop III.2 stars.

Efficient HD cooling can be triggered within the relic H II regions that surround Pop III.1 stars at the end of their brief lives owing to the high electron fraction that persists in the gas as it cools and recombines (Nomoto et al 2005; Yoshida et al 2007). The efficient formation of HD can also take place when the primordial gas is collisionally ionized, such as behind the shocks driven by the first SNe or in the virialization of massive DM halos. There is a critical HD fraction necessary to allow the primordial gas to cool to the temperature floor set by the CMB at high redshift: $X_{\text{HD}} = n_{\text{HD}}/n \sim 10^{-8}$ where n_{HD} is the number density of HD molecules and n that of all particles. Except for the gas collapsing into neutral minihalos, the fraction of HD typically increases quickly enough to play an important role in the cooling of the gas, facilitating the formation of Pop III.2 stars. Other environments conducive for Pop III.2 formation are the so-called ‘atomic cooling halos’ which have been suggested as candidates for the first galaxies (Bromm et al 2009).

In summary, there is an emerging consensus on the main processes involved in forming the first stellar generation because the hydrodynamical processes appear to be relatively well defined. But there are many questions. In particular, what is the role of turbulence in shaping the primordial mass function? If sufficient turbulence were present, a broad range of fragment masses could result, similar to the power-law extension towards high masses observed in the present-day IMF (Clark et al 2011). In comparison, the details of the hydrodynamic processes leading to subsequent stellar generations are poorly understood (Clark et al 2011; Grief et al 2011; O’Shea & Norman 2007; Norman 2011).

One of the most important themes of near field cosmology (Sec. 5) is the use of chemical signatures in the oldest stars today to interpret the sequence of events in the Galaxy’s formation. Unfortunately, the chemical yields of the first stars, and therefore their chemical signatures, are not well constrained at the present time (Karlssohn et al 2012). The signatures are complicated by a number of factors, in particular, the unknown combination of metals that falls back onto any black hole that forms in the subsequent supernova/hypernova explosion (Podsiadlowski et al 2002; Umeda & Nomoto 2003).

2.3. The first black holes

Stars in the mass range $9 \lesssim m_*/M_\odot \lesssim 40$ are believed to explode as core collapse SNe. At the end of hydrostatic burning, stars above $\sim 10 M_\odot$ contain an electron degenerate iron core. As the Fe core continues to grow by silicon shell burning, the electron pressure can no longer counterbalance the increasing gravitational pull and the core collapses. An explosive instability develops when the infalling outer layers bounce off the collapsed core (proto-neutron star) and are ejected into the ISM⁵. The details of the explosion mechanism is still a topic of debate (Burrows et al 2006; Janka et al 2007). Stars around $9 - 10 M_\odot$ instead form a degenerate O+Ne+Mg core which may collapse due to rapid electron captures on ^{20}Ne and ^{24}Mg , prior to the ignition of Ne (Nomoto 1987). If so, an electron-capture, or ONeMg SN is formed. Alternatively, the star loses its outer layers and forms a white dwarf (Garcia-Berro & Iben 1994).

At the high mass end of the core collapse SN regime ($25 - 40 M_\odot$), the potential energy of the stellar envelope is comparable to the kinetic energy of the explosion and the innermost layers of the star fall back onto the central neutron star which eventually collapses to a black hole (Fryer 1999). A weak or faint SN is formed if the black hole is non-rotating. For rotating black holes, however, a much stronger explosion may instead be expected in the form of a hypernova (HN; Fryer & Heger 2000). Stars with masses above $40 M_\odot$ collapse directly to a black hole.

For Pop. III stars with masses $\gtrsim 100 M_\odot$, the pair instability kicks in after central carbon burning, driven by the creation of electron-positron pairs. Below $\sim 140 M_\odot$, this instability causes the star to pulsate violently. While the outer layers are lost in SN-like explosions, the star settles down to form a massive Fe core. Eventually the star collapses quietly to form a black hole. On the other hand, if the star is in the mass range $140 \lesssim m/M_\odot \lesssim 260$ it will face complete disruption as a result of the pair instability, and no remnant is formed (Heger & Woosley 2002). In still

⁵ An elegant visual aid of a related process is to drop a basketball onto the ground with a tennis ball rested on it. When the basketball hits the ground, it bounces a little while the tennis ball shoots up to a great height.

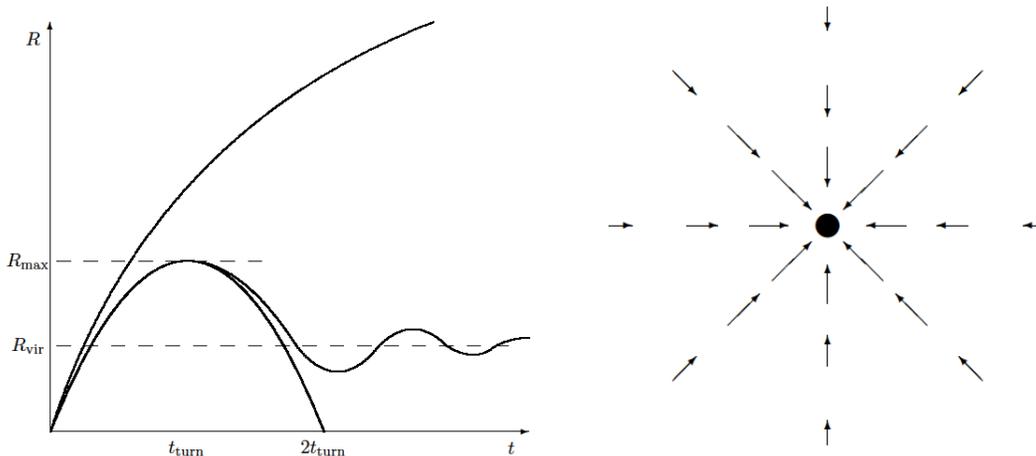


FIG. 3.— (Left) Evolution of a ‘spherical top hat’ perturbation (after Kaiser 2002). The top curve describes the expansion of material external to the perturbation. The bottom curve describes the formation of a black hole in the unlikely event that the mass fluctuation collapses perfectly radially. The middle curve describes the formation of a virialized halo: here the collapsing material essentially bounces before achieving some form of equilibrium. (Right) Divergence-free flow pattern due to the presence of a point mass. Mass accumulates at the origin at a constant rate because the infall velocity increases as r^{-2} (shown by the vectors) while the surface decreases by the same factor.

more massive stars, the energy released from explosive oxygen and silicon burning, caused by the pair instability, is instead used to photodisintegrate the nuclei in the central core. The explosion is halted and a black hole is formed (Fryer et al 2001).

Thus some of the first stars may well have seeded the first black holes in their cores. The existence of high-redshift quasars provides direct fossil evidence of these first seeds. To see why, consider a black hole that converts rest-mass energy with an efficiency ϵ into radiation. Its luminosity is given by

$$L_A = \epsilon \dot{m} c^2 \quad (2)$$

$$= 7 \times 10^{11} \left(\frac{\epsilon}{0.05} \right) \left(\frac{\dot{m}}{M_\odot \text{ yr}^{-1}} \right) L_\odot \quad (3)$$

for which \dot{m} is the mass accretion rate. In reality, the accretion disk luminosity can limit the accretion rate through radiation pressure. The so-called Eddington limit is given by

$$L_E = \frac{4\pi G m_{\text{BH}} m_p c}{\sigma_T} \quad (4)$$

$$= 2 \times 10^{13} \left(\frac{m_{\text{BH}}}{10^9 M_\odot} \right) L_\odot \quad (5)$$

where m_{BH} is the black-hole mass, m_p is the proton mass and σ_T is the Thomson cross-section for electron scattering. This equation allows us to derive a maximum accretion rate \dot{m} , and therefore a maximum luminosity, for a given black-hole mass. (The 1990 Saas Fee lectures provide an excellent introduction to the physics of black holes.)

The critical Eddington accretion rate is proportional to the mass of the accreting black hole ($\dot{m} \propto m_{\text{BH}}$) which provides us with an exponential timescale for a black hole growing at its maximum rate. The so-called Salpeter timescale T_S is given by

$$T_S = \frac{\epsilon \sigma_T c}{4\pi G m_p} \approx 4.5 \times 10^7 \text{ yrs} \quad (6)$$

Consider the case of the newly discovered $z \approx 7.085$ quasar ULAS J1120+0641 (Mortlock et al 2011). A crude estimate of its mass ($m_{\text{BH}} \sim 1 - 3 \times 10^9 M_\odot$) can be obtained from its Mg II line width and luminosity assuming that it is shining close to its Eddington luminosity. We observe this object at a time when the Universe was only 770 Myr old⁶. If our current picture is correct, that the first stars could *not* have formed before ~ 100 Myr, then only ~ 15 Salpeter e-folding times had elapsed before this time. Thus an initial black hole seed of $\sim 30 M_\odot$, say, could have grown at the maximum rate to a mass of about $m_{\text{BH}} \sim 10^8 M_\odot$ in the time available. This is low compared to the estimated mass and therefore, given the uncertainties, only barely adequate to explain the quasar. Sources like ULAS J1120+0641 clearly occupy special environments (e.g. high pressure regions in multiple mergers) where sustainable long-term accretion was possible. These may have been the same environments that produced the oldest globular clusters ($\gtrsim 13$ Gyr) observed in the Galactic halo today. Similar QSO sources at even higher redshift would be a major challenge for our ‘standard model’ for the first stars, or tell us something entirely new about how black holes are formed (in merging groups?) or evolve in the early universe.

⁶ See Ned Wright’s cosmology calculator at <http://www.astro.ucla.edu/~wright/ACC.html> (Wright 2006).

2.4. The first dark haloes

Most of the structure in the Universe came together in the redshift range $z \sim 3000$ to about $z \sim 0.5$ when the Universe was flat and matter dominated. Excellent discussions are provided by Peacock (1999), Kaiser (2002) and Binney & Tremaine (2008). From the FRW metric, the universal scale factor a obeys $a(t) \propto t^{2/3}$ and the mean density is given by $\rho_o(t) = (6\pi G t^2)^{-1}$. Consider a spherically symmetric density fluctuation δ_i within the expanding universe at a time t_i which, when averaged over a volume with radius r_i , is much less than unity. The total mass in this sphere is then

$$M = \frac{4\pi}{3}(1 + \delta_i)\rho_s(t_i)r_i^3. \quad (7)$$

If there is no flow of material across the surface defined by r_i , and no pressure within the collapsing region, then a small lump of matter experiences an acceleration given by

$$\ddot{r} = -\frac{GM}{r^2(t)}. \quad (8)$$

The parametric (cycloidal) solution to this equation is

$$r(\eta) = r_o(1 - \cos \eta) \quad (9)$$

$$t(\eta) = \sqrt{r_o^3/GM}(\eta - \sin \eta) \quad (10)$$

where we have assumed $r = 0$ at the Big Bang. (These same equations can be used to describe the expanding universe where $\dot{a} = (da/d\eta)/(dt/d\eta)$ and Hubble's scale factor is given by $H_o = \dot{a}/a$.) The 'turnaround radius' is defined as $r_{\text{turn}} = 2r_o$ where the expansion halts and the collapse begins ($\eta = \pi$).

If the average density inside the sphere is $\rho_s(t)$, then the average density compared to the universal mean is

$$\delta(t) = \frac{\rho_s(t)}{\rho_o(t)} \quad (11)$$

$$= \frac{9(\eta - \sin \eta)^2}{2(1 - \cos \eta)^3} \quad (12)$$

which at turnaround is equal to $\delta_{\text{turn}} = \frac{9\pi^2}{16} = 5.55$. If we now expand both equations 10 and 12 to eliminate r_i from equation 7, we arrive at equations for the turnaround radius and turnaround time:

$$r_{\text{turn}} \approx \frac{(GMt_i^2)^{1/3}}{\delta_i} \quad (13)$$

$$t_{\text{turn}} \approx 1.1 \frac{t_i}{\delta_i^{3/2}} \quad (14)$$

This last equation is particularly interesting in the context of Λ CDM because it says that the larger the initial fluctuation δ_i , the sooner the collapse process gets going.

In practice, the 'spherical top hat' model for collapse within Λ CDM is overly idealised compared to what is seen in simulations. (See Sheth & Tormen 2001 for a treatment of ellipsoidal collapse.) The collapse cannot be purely radial otherwise the Universe would be the exclusive domain of black holes. In reality, the collapsing dark matter forms a halo which becomes virialized (stabilised) within some radius (see Fig. 3), the virial radius r_{vir} . After virialization, all of the energy is conserved and the potential energy of the system is $W = -\frac{3}{5}GM^2/r_{\text{turn}}$ such that the half-mass radius of the relaxed system is $r_h \approx 0.375r_{\text{turn}}$, or about a third of the turnaround radius. The virialized halo density ρ_h can be compared to the mean universal density after the halo has virialized, $\rho_o(2t_{\text{turn}})$, such that

$$\frac{\rho_h}{\rho_o(2t_{\text{turn}})} = \frac{9\pi^2}{8} \left(\frac{r_{\text{turn}}}{r_h} \right)^3 \quad (15)$$

$$\approx 200. \quad (16)$$

Thus we speak of r_{200} , i.e. the radius within which the mean density is 200 times the universal density at a given epoch, as the virial radius of a halo, and the mass within is taken to be the mass of the dark matter halo. In words, the factor of ~ 200 is made up of three parts: (i) a factor of 5.55 is needed to ensure that matter freezes out of the Hubble flow; (ii) a factor of 8 to describe the increase in density after the twofold collapse in radius; and (iii) a factor of 4 to reflect the drop in the mean background density during the time of the collapse, i.e. $t_{\text{turn}} < t < 2t_{\text{turn}}$. The precise value of the virial radius depends on the underlying cosmology, but $r_{\text{vir}} = r_{200}$ is a useful measure for most relevant cosmologies.

It is possible to go beyond this description of the virialized halo and derive some properties of the internal structure (Gunn & Gott 1972). An isolated point mass M_o generates a divergence-free (pressure free) flow since the infall velocity increases as $\delta v \approx GM_o/(H_o r^2)$ which compensates for the decreasing area of the advecting surface as one approaches the point mass. The amount of mass across the surface is given by $\delta M \approx 4\pi r^2 \rho_o \delta v / H_o$ or equivalently $\delta M \sim M_o$. In

an Einstein-de Sitter universe, the perturbation $\delta\rho/\rho = \delta M/M$ grows linearly with the scale factor a , such that the turnover radius is $r_{\text{turn}} \sim (\delta M/\rho_0)^{1/3} \propto a^{4/3}$. If we assume that the specific binding energy $\delta\phi \propto M(r)^{-1/3}$ of the infalling shell is conserved at some final settling radius r_{fin} , then it follows that $M(r_{\text{fin}}) \propto r_{\text{fin}}^{3/4}$ and $\rho(r_{\text{fin}}) \propto r_{\text{fin}}^{-9/4}$. This profile is *almost* equivalent to an isothermal halo, i.e. the expected form giving rise to flat rotation curves seen in galaxies. If we consider the timing (phase) of infalling material, it is possible to learn about distinct core-halo structures in galaxies even before considering the role of baryons (e.g. Binney & Tremaine 2008, §9.2.4), but we leave this discussion to a later review.

2.5. Reionization and the first galaxies

In Sec. 2.2, we attempted a consensus across a wide literature on the likely properties of the first stars. We know they must have existed but at present there are many unknowns. Did the first stars form in isolation or in groups? Were relatively few massive stars responsible for reionization or was it triggered by the collective effect of massive star clusters?

One of the most important questions is what constituted the earliest stable baryonic *systems* and whether relics of these survive to the present day. The first stellar generations changed the universe in many ways, in particular, the chemical properties and the equation of state of the intergalactic medium. Most of the diffuse atomic hydrogen and even gas collapsed within small halos became fully ionised. This is known as the ‘epoch of reionization.’ Is this exclusively the domain of the most massive stars, or can substantial intermediate and low mass stars form (Tsuribe & Omukai 2006, 2008; Clark, Glover & Klessen 2008)? In other words, did stellar populations observable today exist before reionization (Tumlinson 2010; Okrochov & Tumlinson 2010)?

Peebles (1968) considered the possibility that the first structures were fully baryonic, more specifically globular clusters that formed before the first galaxies came together (see also Ferraro et al 2009). These early baryonic systems may have been more massive than the typical cluster masses observed today in order to account for their enhanced metallicities (Conroy & Spergel 2011; Conroy 2012). How globular clusters fit into the early history of galaxy formation, and what they have to tell us about the first stellar generations (e.g. Caloi & D’Antona 2011), remains a mystery. But the contemporary view is that the earliest baryonic systems must have contained some dark matter to allow baryon cooling to proceed in order to form any stars at all, even the first generation of stars.

During reionization, the cosmic gas was heated above 10^4K which raised the critical Jeans mass to roughly $m_{\text{crit}} \sim 10^8 M_{\odot}$ (with roughly 10% in baryons). In order for the local gas density to survive heating through photoionization (which suppresses star formation), the confining dark-halo mass must have exceeded m_{crit} (Rees 1986; Ikeuchi 1986; Efstathiou 1992; Barkana & Loeb 1999). These arguments based on ionization balance and heating are supported by 3D numerical simulations (Quinn et al 1996; Navarro et al 1997; Gnedin 2000; Okamoto et al 2008). Naively, this means that galaxies (baryons+dark matter) less massive than the m_{crit} threshold could not have formed.

We stress that these earlier simulations do *not* completely exclude ongoing star formation for masses below m_{crit} during the reionization epoch (e.g. Dijkstra et al 2004). The calculations do not consider the time it takes for the gas to evaporate from the low mass halos. For a cosmic ionizing UV intensity $J_0 = 10^{-21} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \text{ sr}^{-1}$, with a moderate power-law form of $f_{\nu} \propto \nu^{-2}$, we obtain an ionisation flux of $\sim 2 \times 10^5 \text{ photons s}^{-1} \text{ cm}^{-2}$. The potential rate of evaporation of hydrogen atoms off the surface of a confined gas cloud, of radius 50 pc ($\sim 3 \times 10^{41} \text{ cm}^2$), assuming one ionising photon evaporates one hydrogen atom, is then given by

$$\dot{m}_{\text{evap}} \sim 1.5 \times 10^{-3} (n_r + 1)^{-1} M_{\odot} \text{ yr}^{-1}, \quad (17)$$

where the n_r term allows for the number of recombinations in the escaping wind. Thus, the timescale for evaporation for the $10^4 - 10^5 M_{\odot}$ of gas in the models shown here ($\tau_{\text{evap}} \sim f_b M_{\text{min}}/\dot{m}_{\text{evap}}$) can exceed 10^8 yr , which is long enough for star formation to proceed in the interior while the outer regions are evaporating. Others have also argued that star formation was diminished but not fully suppressed, such that galaxies with masses of order $10^8 M_{\odot}$ form stars during and immediately after reionization (Gnedin & Kravtsov 2006; Salvadori & Ferrara 2009). An alternative route to suppressing star formation in low mass systems has been to consider the effect of starburst-driven winds (Mac Low et al 1999).

So what were the first stellar systems? There are few if any reliable observational constraints thus far. Primordial objects are important to identify at any redshift because these may retain chemical signatures of the first and second generations of stars (Karlsson et al 2012). Whether long-lived, low mass stars emerged depends in large part on whether dust formed from the ejected material of the first stars (Clark et al 2008). If dust did not form in appreciable amounts, the CMB inhibits gas cooling, in which case low mass stars may not have emerged until $z \lesssim 10$ when mean metallicities reached $\sim 10^{-4} Z_{\odot}$ (Schneider & Omukai 2010).

3. LESSONS FROM GALAXY REDSHIFT SURVEYS

Over the past twenty years, multi-object spectrographs around the world have revolutionised our understanding of galaxies out to redshifts $z \sim 1$ and beyond (see Fig. 4). These surveys have been made possible by high quality u, g, r, i, z photometry from various sources, in particular, the HST and the Sloan Digital Sky Survey (www.sdss.org). A complete list of all surveys to date is maintained by I.K. Baldry (www.astro.ljmu.ac.uk/~ikb/research). The highest redshift sources confirmed spectroscopically are a gamma ray burst at $z \approx 8.2$ (Tanvir et al 2009), a QSO (Mortlock et al 2011) and two Lyman break galaxies at $z \sim 7$ (Vanzella et al 2011), and a host of higher redshift ‘candidates’ with less accurate photometric redshifts (e.g. Robertson et al 2010).

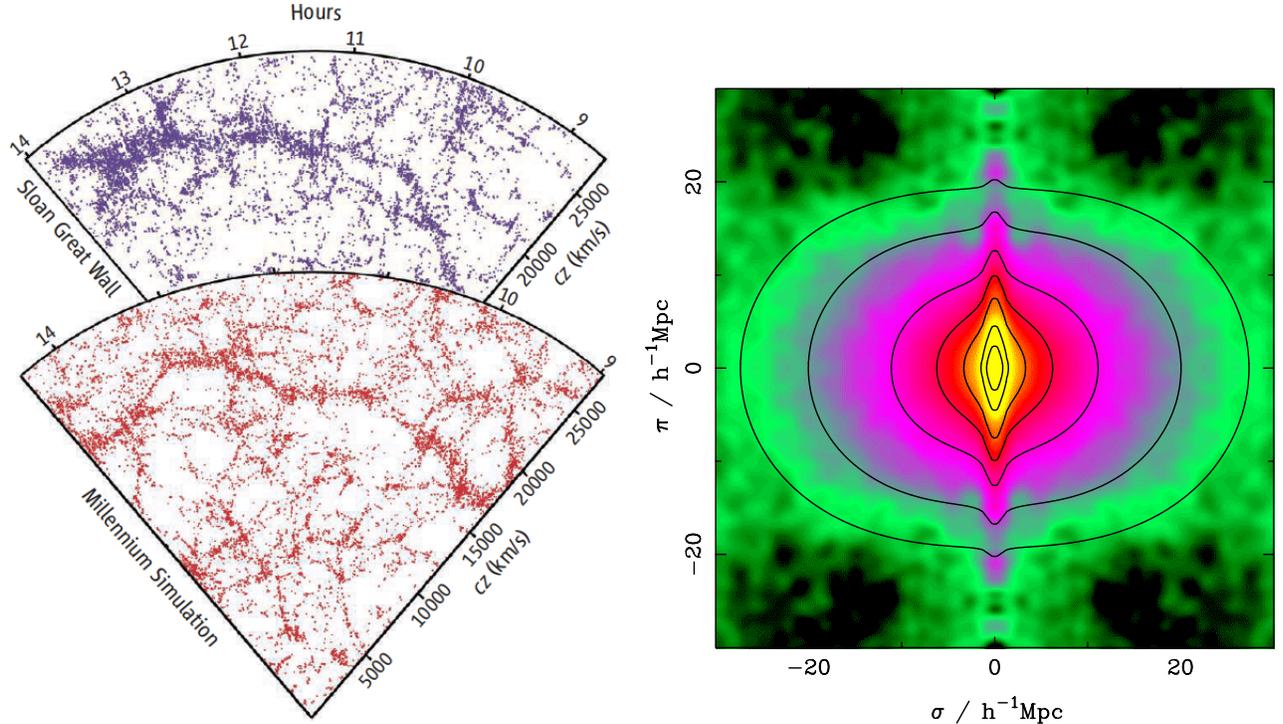


FIG. 4.— (Top left) The “Great Wall of Galaxies” that emerged from the SDSS and 2dFGRS galaxy redshift surveys. (Bottom left) 2D slice from the Millennium Λ CDM simulation which shows that such structures, while relatively rare, are not unexpected (Springel et al 2005). The Great Wall contains in excess of 10,000 galaxies and stretches over 75° (400 Mpc at $z \approx 0.07$). (Right) The redshift-space correlation function $\xi(\sigma_{12}, \pi_{12})$ for the 2dFGRS galaxy redshift survey plotted as a function of the transverse (σ_{12}) and radial (π_{12}) pair separation (see Peacock et al 2001). This figure shows clear redshift distortions with ‘fingers of god’ on small scales and a pronounced flattening on large scales due to infall onto filaments and sheets (see text).

A comprehensive review of galaxy surveys requires an entire book (e.g. Mo et al 2010); we can only give a flavour of the excitement and progress that has been possible in an era of HST and 8-10m telescopes to give some context to the importance of the Local Group. Useful recent reviews include Glazebrook (2012), Silk & Mamon (2012) and Driver et al (2013). We encourage the reader to consider the ramifications of Fig. 5: understanding galaxies is an extremely challenging proposition requiring as much information as we can gather at all cosmic epochs.

3.1. Evolution and environment

Most forms of galaxy activity (star formation, AGN, winds, mergers, etc.) appear to have peaked in the redshift range of $z = 1 - 5$ with a steady decline thereafter (Lilly et al 1996; Hopkins & Beacom 2006). Within the Λ CDM paradigm of an evolving hierarchy, galaxies must have been smaller in the past, and this appears to be supported by extremely deep imaging from the HST (Bouwens & Illingworth 2007). Galaxies were certainly more complex in character, with higher levels of asymmetry, clumpiness and internal velocity dispersion measured in the gas (Genzel et al 2006; Law et al 2007; Förster-Schreiber et al 2009). From a detailed study of the Hubble Ultra Deep Field, Elmegreen et al (2005) identify four main classes of galaxy morphology: ‘chains,’ ‘clump clusters,’ ‘doubles’ and ‘tadpoles.’ Out of this mess emerged modern-day disk galaxies which account for roughly half of the total stellar mass today (Driver et al 2007; Bower et al 2006). While well formed disks are a relatively modern phenomenon, there are occasional discoveries of ‘normal’-looking disk galaxies at early times, in particular, a grand-design spiral recently discovered at $z \approx 2.18$ (Law et al 2012).

One of the most discernible aspects of evolution is the phenomenon of ‘downsizing’ over cosmic time (for an updated discussion, see Fontanot et al 2009). Before 2000, it was assumed that, in a hierarchy, small things form first, and larger objects appear through a process of merging of smaller systems (e.g. Fig. 1). In fact, with the benefit of hindsight, it was clear even in the 1960s that something was amiss. A large fraction of the most massive galaxies today are very old, red and dead (de Vaucouleurs 1961), whereas most low-mass systems show evidence of ongoing star formation. We now know that ellipticals have higher metallicities (Faber 1973) and enhanced α/Fe ratios (Ziegler et al 2005), consistent with the idea that massive red galaxies inhabit dense cluster regions where much of the activity took place at early times ($z \gtrsim 1$; Gonzalez et al 2011; Robertson et al 2010). Presumably all of the smaller building blocks with active star formation that likely filled the early universe have mostly been swept up in dense regions.⁷

We can look at this evolution in another way. Galaxy distributions in colour space (e.g. $g - r$) show a bimodal

⁷ This is reminiscent of the early Solar System when proto-fragments left over from planet formation bombarded the Earth and Moon (Late Heavy Bombardment) giving rise to the familiar ‘face in the moon’ around 4 Gyr ago.

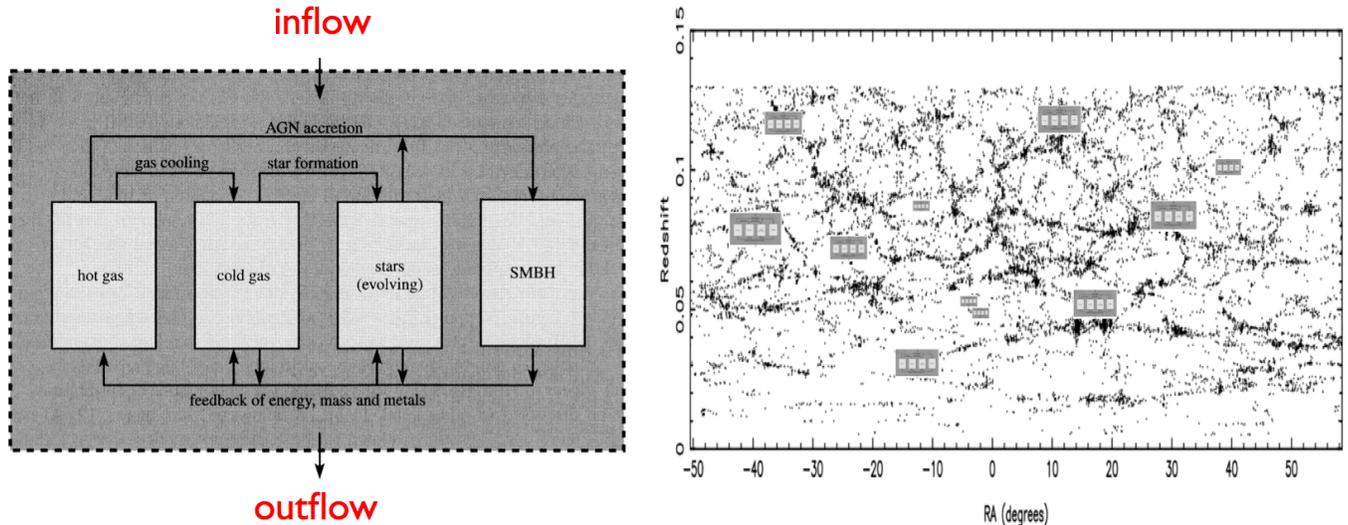


FIG. 5.— (Left) A schematic of the inner workings of a galaxy (adapted from Mo et al (2010)). (Right) An illustration of the manifest complexities involved in galaxy studies. The simple model is distributed across large-scale structure (RA, redshift) identified in Stripe 82 from the Sloan Galaxy Redshift survey. The size of the schematic is proportional to a galaxy’s halo mass. Model galaxies are distributed across a wide range of density contrast. Gas may be entering or leaving the galaxy, or a combination of both, or there may be no inflow or outflow (e.g. ram pressure stripping). Our goal is to understand the inner workings of galaxies across all environments and through cosmic time.

distribution where the details are strongly dependent on environment (Strateva et al 2001; Blanton et al 2003; Baldry et al 2004). The ‘blue sequence’ is dominated by disks with ongoing star formation; the ‘red sequence’ is more of a mixed bag of dusty spirals and spheroids. The role of environment in galaxy evolution has been reviewed comprehensively by Blanton & Moustakas (2009). A complicating factor is that galaxy surveys provide accurate positions on the sky but only rough distances can be deduced from a redshift. Thus a formal definition of environment is hard to come by. Many schemes have been explored but these amount to little more than statistical measures of crowding (Muldrew et al 2012 – Table 1).

In addition to environment, a galaxy’s evolution is affected by its mass at a given epoch, although mass and environment are coupled in the sense that more massive galaxies appear to be highly clustered at all epochs (Mo et al 2010). Galaxy masses are difficult to determine at the present time although ‘photometric masses’ using red or infrared bands are a useful proxy (Baldry et al 2008). Whether evolutionary effects are dominated by a halo’s mass (or its central black hole) or its association with clustered galaxies embedded in a hot cluster medium are major questions that remain unresolved.

Thus we see evolution in time and evolution across large-scale structure. Interestingly, we also see evolutionary differences between groups of comparable mass. For example, the M81 group today has a similar inventory to the Local Group by mass, but appears to have more neutral gas, with extended tidal arms of H I and an HVC populations dominated by massive gas clouds. It seems likely that we are seeing comparable mass groups at different stages in their evolution (Nichols & Bland-Hawthorn 2009, 2013).

3.2. Accretion and feedback

Just how gas and dark matter move out of voids and into filaments, sheets and clusters are topics of key interest in modern cosmology. The Two Degree Field Galaxy Redshift Survey (2dFGRS; $z \approx 0.05$) provides clear evidence that 80% of galaxies are collapsing into sheets and filaments (Peacock et al 2001), the so-called Kaiser effect (Kaiser 1987). From these data, it is possible to determine a 2-point correlation function $\zeta(\sigma_{12}, \pi_{12})$ from the transverse (σ_{12}) and radial (π_{12}) pair separation. The ζ distribution measures the excess probability over random of finding pairs; for an isotropic random distribution, ζ has circular (iso)likelihood contours. But the radial vector π_{12} is determined from the redshift and therefore suffers redshift-space distortion. Small scales are dominated by the internal motions of clusters which leads to the familiar ‘finger of god’ effect. At larger separations, the likelihood contours are clearly flattened in the transverse direction consistent with galaxies having fallen away from the general Hubble flow (Peacock et al 2001). Inflow onto individual groups and clusters is also seen in limited redshift surveys (Ekholm & Teerikorpi 1994; Ceccarelli et al 2005) and possibly even in samples at $z \sim 2.4$ (Rakic et al 2011).

Sheet structures may already be evident at large redshift. High resolution spectroscopy of paired QSOs provides two closely spaced sight lines to explore all IGM material from the present day to $z \sim 4$. Rauch et al (2005) find evidence of gas sheets at different orientations presumably associated with galaxies in large-scale structure (see also Penton et al 2004), although large-scale expansion of the sheets cannot be ruled out. The distribution of Ly α galaxies at these epochs appear to show evidence of sheets and filaments (Matsuda et al 2005).

It is extremely difficult to probe gas beyond the nearby universe. H I 21cm emission is only observed to $z \sim 0.1$ directly although $z \sim 0.4$ has been achieved through stacking observations. Reaching H I 21cm beyond $z \sim 1$ is a major

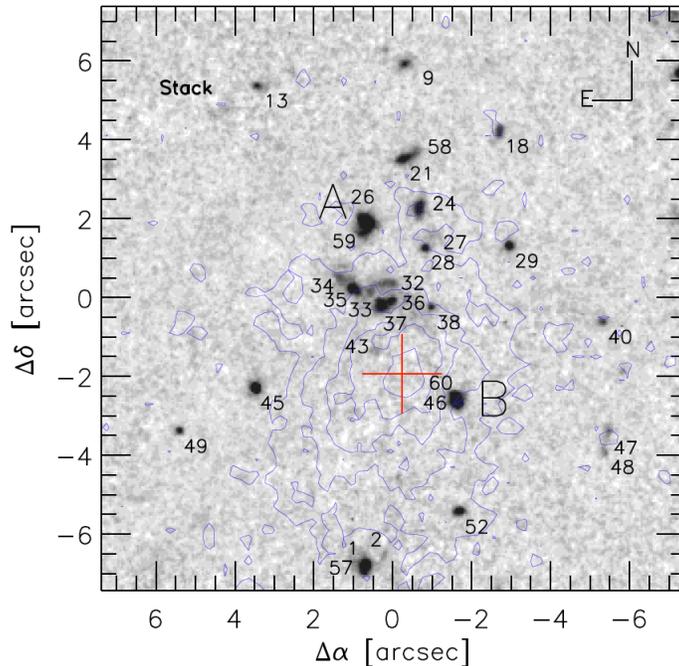


FIG. 6.— Giant Ly α nebula (blue contours) in LABd05 ($z = 2.66$) where the greyscale image comes from HST broadband imaging. The centroid of the Ly α emission observed with Subaru is shown as a red cross. The extent of the gas ($\gtrsim 100$ kpc) is consistent with the early formation of a compact massive group (Prescott et al 2012).

science goal of the Square Kilometre Array in the next decade. In the nearby universe, all HI clouds are associated with galaxies without exception (Sancisi et al 2008). Early claims of accreting protoclouds on megaparsec scales (e.g. Blitz et al 1999) have never been firmly established at any epoch. On smaller radial scales, evidence has been found for metal-rich gas accretion onto star-forming galaxies (Rubin et al 2012). *Clear unambiguous evidence of gas accretion on galaxy scales is rare at any epoch.*

Ly α imaging has been used to detect diffuse gas in halos of high redshift sources, most notably with the Grantecan Osiris tuneable filter at $z \sim 4.4$ (Kuiper et al 2011). Another spectacular example is shown in Fig. 6. An impressive demonstration of the generality of Ly α halos at this epoch has come from stacking ~ 100 images of Lyman break galaxies (Steidel et al 2011) where diffuse gas is seen out to 80 proper kiloparsecs. Galaxies must accrete gas in order to form stars, but whether this accretion is largely through infalling dark halos or a smooth flow associated with dark matter remains unclear. The specific processes involved with how gas gets into galaxies is discussed in Section 4.

Numerical simulations appear to show that there must be some form of ‘feedback’ within a forming galaxy to produce realistic disks and to match the details of the bimodal colour distribution. Starburst winds are often invoked to remove the low angular momentum build-up of material seen in synthetic simulations (Sharma et al 2012; Brook et al 2011). These are known to occur at all redshifts (Heckman et al 1989; Veilleux et al 2005), but whether the numerical outflows have any relationship to real winds is unclear (Springel & Hernquist 2003).

A stronger argument for feedback stems from the fact that the mass distribution of dark matter haloes predicted in Λ CDM simulations looks very different from the observed stellar mass distribution of galaxies (Fig. 7). Semi-analytic models (SAMs) that attempt to paint galaxies onto Λ CDM models through increasingly complex prescriptions fail to account for this behaviour without feedback. It seems increasingly likely that the complexity of baryonic physics has been underestimated and not adequately resolved nor understood in existing numerical simulations. In particular, star formation seems to be too complex to be simply gravity-induced, which has been the driving assumption in many numerical models.

In a recent review, Silk & Mamon (2012) appeal to processes that are strongly dependent on the dark halo mass. Supernova-driven feedback in starbursts may provide an acceptable fit to the turnover in the low mass end of the galaxy luminosity problem, but there are numerous problems. Low-mass galaxies have cores rather than cusps that arise in Λ CDM simulations so there are clearly other processes at work. It is possible that outflows can flatten cusps to cores (Governato et al 2012) if most of the activity occurred in the early universe, but these models fail to account for the rapid dwarf evolution at low redshift (Weinmann et al 2012), even accounting for the suppression of gas accretion onto dwarfs during reionization.

At the high mass limit, AGN feedback appears to account for the sharp cut-off in the bright end of the galaxy luminosity function (Croton et al 2006; Panter et al 2007). In a recent model, the cold flows onto the high mass haloes that drive the star formation are eventually interrupted by AGN-driven superwinds (Dubois et al 2012). But if the wind or jet momentum is sufficient to stop gas inflow, it may even *induce* in situ star formation (Silk & Mamon 2012).

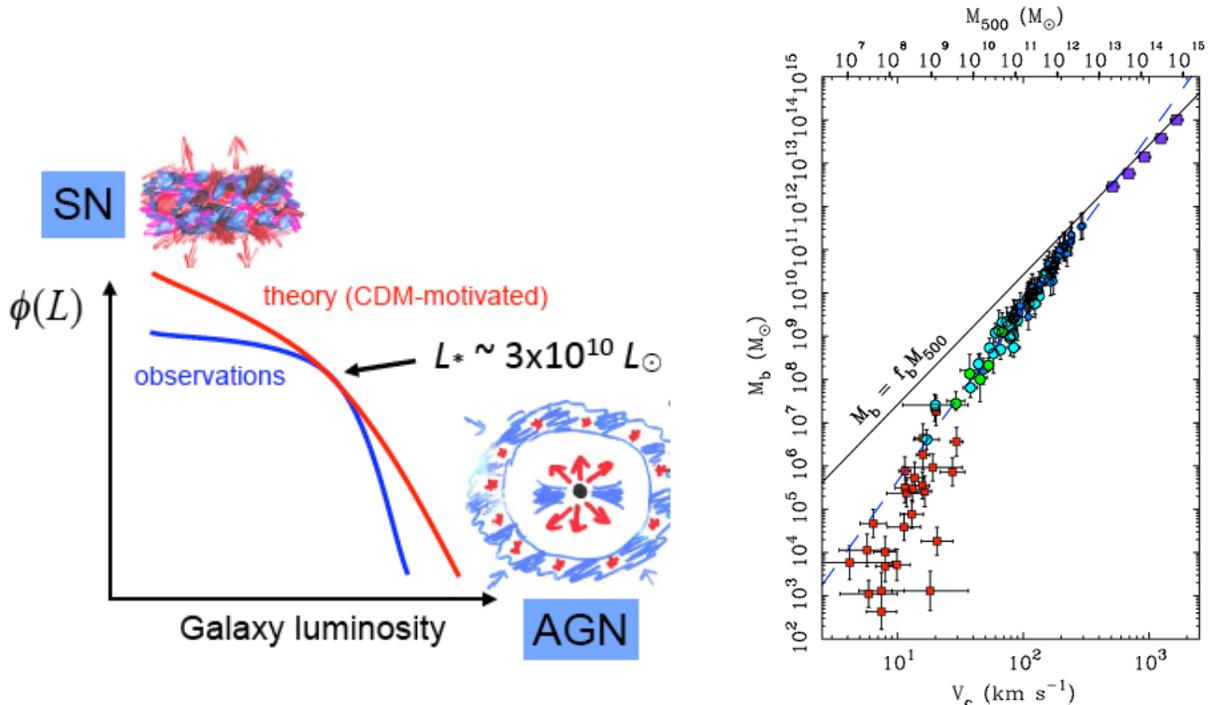


FIG. 7.— (Left) Schematic drawing of the Λ CDM “luminosity function” vs. the stellar luminosity function $\phi(L)$ illustrating the possible role of feedback in modifying the observed distribution. The discrepancy at the low mass end may arise from starburst activity, and at the high mass extreme from AGN activity (from Silk & Mamon 2012). (Right) Baryon mass M_b in clusters, groups (purple), galaxies (blue, green), Local Group dwarfs and ultra-faint dwarfs (red). The halo mass scale is shown at the top. The solid line is the derived baryon mass from assuming a universal baryon fraction $f_b^o \approx 0.17$ (McGaugh et al 2010).

3.3. Baryon inventory and metal enrichment

The universal baryon fraction of the total matter content of the universe $f_b^o \approx 0.17$ is well determined from the angular power spectrum of the CMB (Komatsu et al 2011). There is a well established trend in the baryon fraction observed in dark matter haloes as a function of halo mass (Fig. 7). The most massive systems, i.e. clusters with total masses of order $10^{15} M_\odot$, have baryon fractions that approach the universal value (Dai et al 2010). For L_* galaxies like the Milky Way, $f_b \lesssim 5\%$; in the limit of ultrafaint dwarf galaxies, f_b approaches zero. The explanations for this phenomenon are many and varied: either specific accretion rates decline with halo mass, or the lower mass haloes are interacting with their environment, or starburst-driven winds remove baryons more easily, or the baryons are less visible in lower mass systems.

Shull et al (2012) have recently provided a useful baryon census of the local universe. In Fig. 8, the baryon census is given as 25% in a warm-hot intergalactic medium, 28% in the Ly α forest, 4% in the intracluster medium, 9% in galaxies and a further 5% in the circumgalactic medium (i.e. on scales of the halo virial radius); a further 29% is listed as ‘missing in action.’ Circumgalactic media (CGM) are in essence multiphase gaseous haloes that extend to the halo virial radius. Spectacular evidence for the widespread occurrence of the CGM has come from OVI absorption observed through the haloes of L_* galaxies at intermediate redshift (Tumlinson et al 2011). Interestingly, CGM galaxies are almost exclusively systems with active star formation; this gas appears to be circulating through the halo rather than escaping in a large-scale wind.

The baryon inventory is highly uncertain at the present time. The SSD budget (i) overlooks the intracluster medium in more numerous, low mass clusters ($\lesssim 10^{14} M_\odot$); and (ii) the existence of a diffuse shock-heated IGM with a temperature above 10^7 K. The budget suffers from problems of double counting phases, assumes fixed chemical abundances, uses ionisation corrections from numerical simulations which may not account for all sources of heating, and so forth. Thus it remains unclear whether there are missing or invisible baryons in the universal baryon inventory (cf. Danforth & Shull 2008; Shull et al 2012).

The dramatic evolution in the star formation rate density ($\dot{\rho}_*$) (Hopkins & Beacom 2006) is expected to leave a signature in the baryon mass locked up in stars vs. gas with cosmic time. More of the mass is expected to be locked up in stars as the universe ages. This is largely borne out by HST imaging of the stellar content of galaxies when compared to the cold gas content in Lyman limit systems (Wolfe et al 2006). The corresponding evolution in the universal metal content is also observed (Ryan Weber et al 2009; Rafelski et al 2012). The CIV absorbers show dramatic evolution at $z \sim 6$ suggesting a universal metallicity threshold of $Z_{\text{IGM}} \gtrsim 10^{-4} Z_\odot$ by this redshift (Ryan Weber et al 2009). These authors interpret this signature as fossil evidence of unseen stellar evolution at $z \sim 9$.

It is not known at what stage in cosmic time the last vestiges of pristine gas finally succumbed to stellar enrichment,

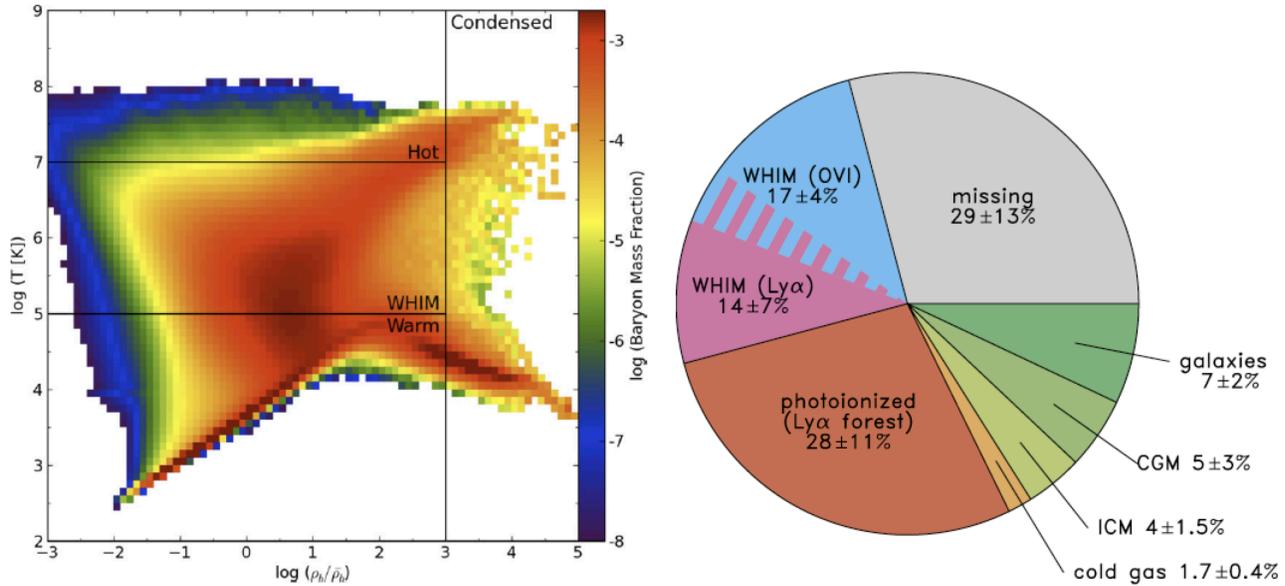


FIG. 8.— (Left) Physical state of the universal baryons (gas density vs. temperature) where colour refers to the relative baryon fraction. (Right) Baryon inventory: warm-hot ionized medium (WHIM), intracluster gas (ICM) and circumgalactic medium (CGM). The overlap of the WHIM phases illustrates the overlap between these phases when detected (Shull et al 2012). About a third of the baryons are unaccounted for. These figures are only illustrative; a lively debate continues about the relative proportions of the different phases.

but it is not inconceivable that some Pop III stars formed a billion years after the reionization epoch. This raises the tantalizing prospect that we can identify such regions in direct observations of the intermediate and high-redshift Universe. Extremely low values for the IGM metallicity have now been observed at $z \sim 3$ (Fumagalli et al 2011). Thus we may be seeing back to the material that formed some of the most metal poor stars associated with galactic halos like our own.

3.4. Chemical evolution in galaxies

We refer the reader to the excellent lecture notes provided by F. Matteucci as part of the 2007 Saas Fee series.

3.5. Milky Way and Local Group analogues in the real universe

We mention one more aspect of galaxy surveys that has important implications for near field cosmology. The GAMA galaxy survey (350,000 sources at $z \lesssim 0.1$; Driver et al 2011) is sufficiently densely sampled that we can ask questions such as ‘How common are Local Group assemblies in the Universe?’ This is a specific goal of the GAMA survey, i.e. to find close analogues to the Milky Way (i.e. two Magellanic companions, comparable mass) and the Local Group (two Milky Way [MW] mass galaxies with Magellanic companions). Robotham et al (2012) find that 12% of galaxies have a companion within a factor of two of the LMC’s mass, although this drops to 3% if we include an SMC mass companion as well. Two full analogues to the Milky Way were found where a closely grouped MW-LMC-SMC type triplet are all actively star forming. This amounts to less than 1% of galaxies in the GAMA survey. Thus while the Milky Way may be rather typical of late-type L_* galaxies, it seems to be relatively special when the satellite population is also considered even when star formation rates are ignored. The search for Local Group analogues is presently ongoing; one striking MW-M31 analogue has already been found (Fig. 9; Robotham et al 2012).

3.6. Milky Way and Local Group analogues in simulated universes

In parallel with the observations, an equally important goal is to identify realistic analogues in cosmological N-body simulations. Large-scale simulations involving up to 10^{10} particles include Millennium I (Springel et al 2005), MareNostrum (Gottloeber et al 2006), Millennium II (Boylan-Kolchin et al 2009), Bolshoi (Klypin et al 2011) and MultiDark (Prada et al 2012). Dark matter simulations have also been carried out with exquisite detail for individual haloes (Springel et al 2008; Stadel et al 2009).

Some of the most important simulations to date that are specific to near field cosmology are the Constrained Local Universe Simulations (CLUES; www.clues-project.org) that achieve realisations which are heavily constrained by local data (Hoffman & Ribak 1991; Kravtsov et al 2002; Klypin et al 2003; Yepes et al 2009; Gottloeber et al 2010). Forero-Romero et al (2011) have recently identified what appears to be a Local Group analogue. They confirm what is known from the stellar record that the Milky Way and M31 are products of a quiet mass assembly history where the last major merger occurred more than 10 Gyr ago (Wyse 2001). Much like galaxy redshift surveys, these simulations contain a wealth of information that is yet to be studied in detail.

4. GAS ACCRETION ONTO GALAXIES

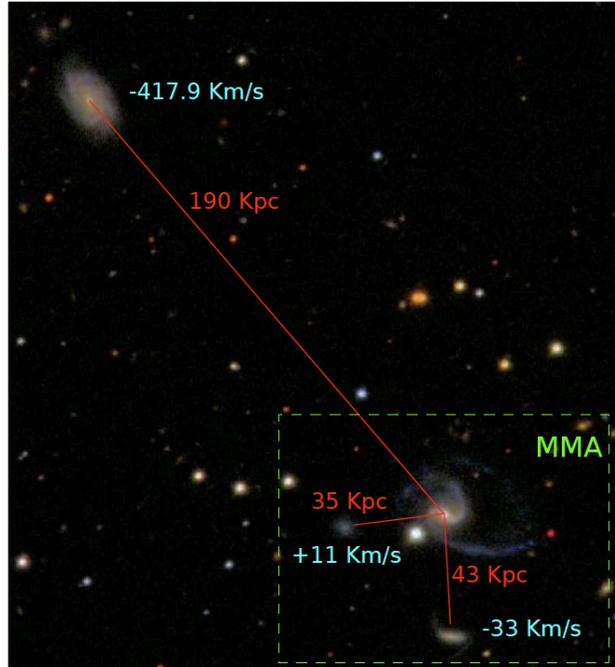


FIG. 9.— An SDSS image of the best analogue of the Local Group to emerge from the GAMA galaxy survey (Robotham et al 2012). The group marked MMA is remarkably similar to the Milky Way - Magellanic System where the three main galaxies are all forming stars at the present time.

4.1. Introduction

A major focus of these lectures has been to consider how gas gets into galaxies, with specific attention to the Local Group. We explore this below in terms of current fads – cooling flows, cold flows, warm flows – but we start by discussing early ideas on galactic accretion as these provide a more solid basis for understanding important physical processes.

As we have seen, it is well established that the observed baryons over the electromagnetic spectrum account for only a fraction of the universal baryon fraction f_b^o , particularly on scales of galaxies. For example, the Galaxy has numerous baryonic phases that have been studied in great detail over many years. A detailed inventory of the Galaxy reveals only a quarter of the universal fraction (Flynn et al 2006), and the same holds true for M31 (Tamm et al 2012). Moreover, the build-up of stars in the Galaxy requires an accretion rate of $1 - 3 M_\odot \text{yr}^{-1}$ (Williams & McKee 1997; Binney et al 2000), substantially larger than what can be seen from direct observation.

So how does gas get into galaxies? This is a major problem in contemporary astrophysics which remains largely unsolved. Does the gas arrive cold, warm or hot, or in the form of stars? Are the infalling baryons brought in by dark matter or as a steady rain of material? Part of the problem is that baryons have many phases (plasma, atoms, molecules, dust, metals, etc.) which call for a multiwavelength observational approach and hugely complex computer codes to track all of the ingredients. We may be a decade or more away from a physical picture that would convince most physicists of its validity.

4.2. Earliest epoch of gas accretion

Accretion prior to reionization. As we have seen, at the time of recombination, the distribution of matter was remarkably uniform, although tiny fluctuations at the level of 1 part in 100,000 can be measured from the CMB. But once the baryons and photons had become decoupled, the baryons were able to accrete onto evolving density fluctuations in the dark matter for the first time. In Fig. 10, the gas temperature fluctuations remain tied to the photon temperature fluctuations for a while after recombination. This is because the uncombined free electrons still experience Compton scattering with the CMB. But within the horizon scale ($k \gtrsim 0.01 \text{ Mpc}^{-1}$ at $z \sim 1000$), where gravity dominates, gas is clearly accreting onto dark matter halos, with a fluctuation spectrum converging on the Λ CDM spectrum by $z \sim 200$. By this time, the thermal coupling of the gas with the CMB is essentially over; the universal expansion is driving the adiabatic cooling such that the gas temperature fluctuations converge to $\sim 2/3$ of the amplitude of the gas density fluctuations.

It is interesting to ponder whether one can even observe these first accretion (fossil) events, long before the first stars formed – the so-called “Dark Ages” (Scott & Rees 1990; Madau et al 1997). This question is of particular interest to scientists considering the impact of the Square Kilometer Array (SKA) in detecting gas phases at these early epochs (Barkana & Loeb 2007; Furlanetto & Loeb 2004). H_I 21cm in its ground state has hyperfine splitting due to parallel spins (triplet state) of the proton and electron pair having a slightly higher energy than the anti-parallel (singlet) state.

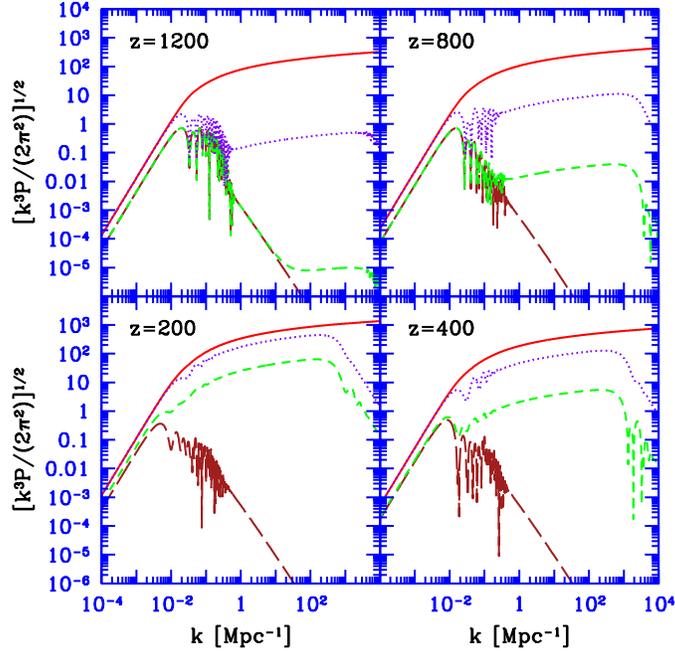


FIG. 10.— Power spectrum plots of density and temperature fluctuations in the early universe (CMBFAST; Barkana & Loeb 2007) as a function of comoving wavenumber $k = 2\pi/\lambda$ where λ is a physical scale in Mpc. The fluctuation spectra are shown for four redshifts: 1200, 800, 400, 200. The four curves correspond to fluctuations in Λ CDM density (solid), baryon density (dotted), baryon temperature (short-dashed) and photon temperature (long-dashed). Note how gas is well and truly accreting onto DM halos by $z = 200$.

The pre-reionization era was marked by subtle variations in H ionisation fraction, gas density and spin temperature. The energy density of the redshifted 21cm emission is expected to be about 1% of the CMB.

Are the spatial fluctuations in the early gas distribution observable? The hyperfine levels tend to be in thermal equilibrium with the CMB. To be observable, the 21cm level populations must correspond to a spin temperature that is different from the CMB in order to be observed in emission or absorption. Two possible ways to do this involve continuum UV from early sources of radiation ($Ly\alpha$ pumping), or large-scale IGM shocks or small-scale shocks (spin exchange) as gas collapses onto halos. It remains unclear which, if any, of these mechanisms is likely to lead to a successful SKA detection (Furlanetto & Loeb 2004; Shapiro et al 2006).

In Sec. 2.2, we discussed how ($z \gtrsim 40$) major cooling onto a small fraction of dark matter minihalos led to the formation of the first stars. This ultimately gave rise to inhomogeneous reionization as the first stars bleached their way through the extended IGM. The suppression of low-mass galaxies during the reionization epoch (Sec. 2.5) is modulated by fluctuations in the timing of reionization (Babich & Loeb 2006). What are the likely effects of patchy reionization in the nearby Universe? It is conceivable that the faint end of the galaxy luminosity function varies across the hierarchy because of the varying impact of the UV field (Efstathiou 1992). But present galaxy surveys do not go deep enough to allow us to study the dwarf population far beyond the Local Group. At present, there is no clearly identified fossil signature of the reionization epoch in the Local Group galaxy population (Grebel & Gallagher 2004). *Accretion after reionization.* Doroshkevich et al (1967) first considered the impact of a UV background in quenching star formation. They realised that the photoheated gas becomes too hot to stay confined by the shallow potential well of the galaxy (see Sec. 2.5). The issue is particularly relevant to the Local Group because N-body simulations repeatedly overpredict the number of dwarfs that are visible today (e.g. Moore et al 1999; Klypin et al 1999). It is plausible that many low-mass dwarfs were suppressed for all time during the reionization epoch. Are there thousands of dark matter minihaloes moving through the Local Group devoid of baryons? There are no strong constraints at the present time.

Loeb (2010) provides three useful formulae for relating halo properties with cosmic epoch:

$$R_{\text{vir}} = 1.5 \left(\frac{M}{10^8 M_{\odot}} \right)^{1/3} \left(\frac{1+z}{10} \right)^{-1} \text{ kpc} \quad (18)$$

$$V_c = 17.0 \left(\frac{M}{10^8 M_{\odot}} \right)^{1/3} \left(\frac{1+z}{10} \right)^{1/2} \text{ km s}^{-1} \quad (19)$$

$$T_{\text{vir}} = 1.0 \times 10^4 \left(\frac{M}{10^8 M_{\odot}} \right)^{2/3} \left(\frac{1+z}{10} \right) \text{ K.} \quad (20)$$

R_{vir} is the virial radius (see Sec. 2.4), i.e. the natural scale of the evolving system within which infall material is no longer collapsing (or expanding). V_c is the circular velocity at the virial radius such that $V_c = (GM/R_{\text{vir}})^{1/2}$. T_{vir} is the virial temperature at R_{vir} defined as $T_{\text{vir}} = \mu m_p V_c^2 / 2k$ for which μ is the mean molecular weight (0.6), m_p is the proton mass and k is Boltzmann's constant.

The critical mass m_{crit} below which galaxies are unlikely to have formed is not fully resolved. Gnedin (2000) argued that m_{crit} corresponds to the scale over which baryonic perturbations are smoothed in linear theory. But this was found to be an order of magnitude too severe by Okamoto et al (2008) who showed that it was the gas temperature at the virial radius which determines whether a halo can accrete gas, i.e. the usual argument relating to Jean's mass where the contribution of the dark matter to the gravitational potential is included. Gas infall depends sensitively on the Jeans mass (Sec. 2.5). Thoul & Weinberg (1996) found a 50% reduction in the collapsed gas mass due to heating for a halo with circular velocity $V_c \sim 50 \text{ km s}^{-1}$ at $z \sim 2$ and complete suppression below $V_c \sim 30 \text{ km s}^{-1}$ (cf. Rees 1986). With the inclusion of self-shielding by the gas, Kitayama & Ikeuchi (2000) determined slightly lower velocity thresholds. Dijkstra et al (2004) found substantially lower thresholds were possible at $z \gtrsim 10$ even while the UV background is ramping up.

In later sections, we consider in more detail just how gas manages to find its way into the evolving dark matter haloes. We start with some early ideas that continue to be relevant today because they have a solid analytic basis.

4.3. Early ideas on galaxy accretion

The problem of accretion onto a spherical body was studied by Bondi et al (1947; hereafter BHL). Their focus was on stellar accretion but this seminal paper is an important starting point for understanding galactic accretion. In the BHL framework, there are three characteristic accretion radii. The first, R_0 , is due to the cross-sectional area of the body. A galaxy moves through the intergalactic medium, sweeping up anything in its path. The second, R_A , is the radius within which the particle velocity is less than the escape velocity and the particles are therefore accreted gravitationally onto the galaxy. This does not take into account collisions, except on the accretion axis. The final radius is the Bondi radius, R_B , which is based on a fluid dynamical treatment of the case where the galaxy is stationary relative to the gas and therefore simply accretes spherically symmetrically through gravitation. For the Galaxy, Hunt (1971) calculated the radii as $(R_0, R_A, R_B) = (12, 160, 54) \text{ kpc}$ and notes that these values are within an order of magnitude and that there is therefore no single dominant process. This is in contrast with the situation of stellar accretion, where the values are (0.005, 4, 700) AU.

For galactic accretion, the dominant term is dependent on the velocity of the galaxy through the gas. Hoyle & Lyttleton (1939) showed that the rate of accretion is given by

$$A = \frac{4\pi(GM)^2\rho_\infty}{V^3} \quad (21)$$

given certain assumptions such as the temperature not being too high. However, this analysis was performed by assuming the density of gas to be infinite along the axis the gas was accreting onto. Bondi & Hoyle (1944) replaced this with a more physically reasonable accretion axis, which led simply to the incorporation of a multiplicative factor α , a parameter which lies between 0.5 and 1. For the case they were interested in, a star moving perpendicularly into a flat cloud, Bondi & Hoyle calculated α to be approximately 0.625. Applied to the Galaxy, using the calculated value of R_A , Equation (4.1) gives an accretion rate of $10 M_\odot \text{ yr}^{-1}$ (Hunt 1971), a high value by modern standards.

The work discussed above ignored particle interactions except on the accretion axis. However, for accretion of the IGM onto a galaxy, the mean free path of particles is only a few parsecs, much smaller than the relevant accretion radii, which are typically in the region of tens or hundreds of kiloparsecs. Thus particle collisions can be important and therefore a fluid dynamical treatment is required. However, as Hunt (1971) noted, for the case of highly supersonic fluid flow, the kinetic energy term of the fluid equations dominates over the pressure term, meaning that the fluid dynamical treatment is unnecessary. The only place where pressure is important is on the accretion axis, where an accretion column similar to Bondi & Hoyle's is formed, so Eq. 21 is relevant here.

For most cases where the velocity is not highly supersonic, the full fluid dynamical treatment cannot be performed analytically. However, a case where it is possible is accretion onto a stationary body, with the flow being spherically symmetric. Bondi (1952) showed that the accretion rate for this case is given by

$$\frac{4\pi\zeta(GM)^2\rho_\infty}{c_\infty^3}, \quad (22)$$

where c_∞ is the sound speed in the gas at infinity and ζ is a positive number which is found by matching the pressure of the gas at the surface of the body with the surface pressure of the body. This gives an accretion rate of $1.5 M_\odot \text{ yr}^{-1}$ for the Galaxy, using the calculated R_B above, a value that is sufficient to sustain the current star formation rate in the Galaxy (Williams & McKee 1997).

The two equations above are similar, which led Bondi to suggest that the accretion rate for the intermediate case, which cannot be calculated analytically, be interpolated from the low and high velocity cases, i.e.

$$A = \frac{2\pi(GM)^2\rho_\infty}{(c_\infty^2 + V^2)^{3/2}} \quad (23)$$

This gives an accretion rate for the Galaxy of $2.5 M_{\odot} \text{ yr}^{-1}$, similar to what is required to explain the observed rate of star formation. Hunt believed that this was the best estimate for the accretion rate of the galaxy. Dodd (1953) confirmed that this formula was at least approximately correct for the case of isothermal flow past a gravitational point source at four different supersonic velocities.

So is there observational evidence for a basic accretion process of the kind envisaged by Bondi, Hoyle and Lyttleton? What are the expected signatures? Unquestionably, the process is complicated, and the signatures will be difficult to dig out. Quite apart from the motion of the Galaxy around the Local Group barycentre (Cox & Loeb 2008), the Local Group has a comparable velocity with respect to the intergalactic medium (Tully et al 2008, Table 3). There are no clear accretion signatures at the present time other than that inferred from the high-velocity cloud (HVC) population. The important point is that the current baryon inventory of the halo fails by a substantial factor to account for the required star formation rate in the Galaxy (Bland-Hawthorn 2009; Aumer & Binney 2009).

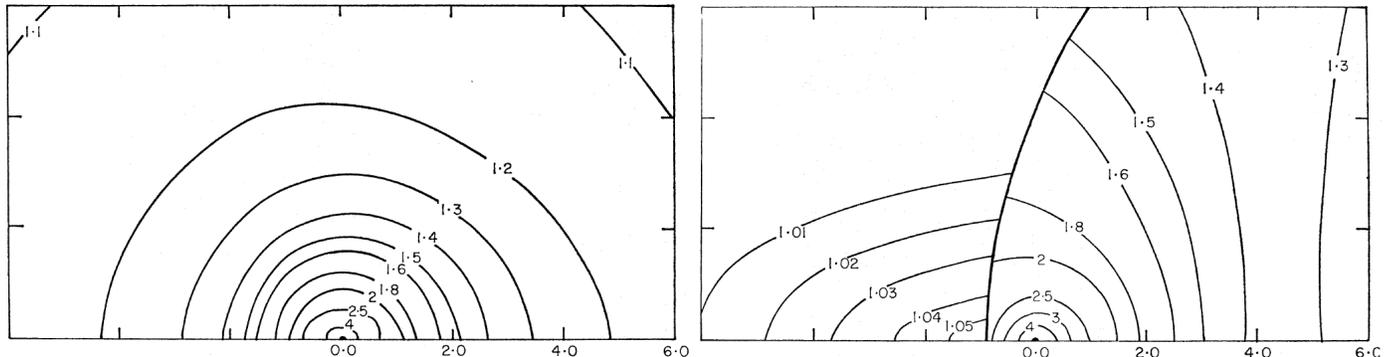


FIG. 11.— Accretion flows onto the Galaxy centred at the origin for two different Mach numbers: (left) $\mathcal{M} = 0.6$; (right) $\mathcal{M} = 2.4$ (from Hunt 1971).

4.4. Accretion shocks

If the Galaxy moves through ambient gas supersonically, a shock wave will appear, much like the heliopause ahead of the Sun. A shock wave is a very sudden change in the properties of the medium, usually involving a large jump in pressure, density and temperature. The condition required for supersonic accretion is that the virial temperature of the halo is greater than the temperature of the accreting gas (e.g. Binney 1977).

Models of accretion shocks show that the location of the shock depends on the relationship between cooling times and dynamical times. The warm-hot intergalactic medium is expected to be in the form of a plasma at 10^7 K, which means that cooling times are longer than dynamical times (Sutherland & Dopita 1993) and the shock is therefore expected to occur close to the virial radius (Benson 2010). A shock can either be a *bow shock*, where the shock is located in front of the body, or a *wake shock*, where the shock follows the body. A bow shock will appear if the medium can provide the sufficient pressure to maintain it, otherwise it will be a wake shock. The nature of the Galactic shock has been the subject of debate since Hunt's early work.

The Galaxy is travelling through the intragroup medium with a transonic velocity, i.e., its velocity is of the same order as the sound speed in the medium (Portnoy et al 1993). This means that the accretion rate cannot be found analytically (§ 4.1). Hunt (1971) tested Bondi's interpolation (Eq. 23) using numerical integration and investigated the nature of the shock in cases where the flow is supersonic. He wrote the fluid equations as conservation laws because this allows a shock wave to appear automatically in the solutions. The shock itself does not work well with numerical methods because it is a discontinuity, but the conservation law method uses continuous quantities and the shock appears as steep rises over a few intervals, rather than as a discontinuity. Hunt found that for the subsonic case he tested, $\mathcal{M} = 0.6$, where \mathcal{M} is the Mach number, the flow was largely isotropic, with a mean anisotropy of only 5 percent. The density is very similar to Bondi's case of spherically symmetric flow, with only 3 percent average deviation. His conclusion was that if the Galaxy is moving subsonically through the intergalactic medium, it would receive matter as if it were stationary.

Hunt also modelled the transonic cases $\mathcal{M} = 1.4$ and $\mathcal{M} = 2.4$ and finds that the flow exhibits a bow shock. This result is in contrast to Spiegel (1970) who argued for a tail shock because there was an insufficient pressure head to support a bow shock. Importantly, Hunt claims that a pressure head *is* formed due to the increase in density as material is accreted onto the Galaxy, which results in pressure forces acting away from the Galaxy. In other words, a constant rain of accreting material can assist in the formation of a bow shock. For $\mathcal{M} = 1.4$, Hunt finds the position of the shock at $0.87 R_B$ in front of the body and $3.9 R_B$ perpendicular to the direction of motion; for $\mathcal{M} = 2.4$, these numbers are 0.11 and $0.36 R_B$. The accretion rate is found to agree with the prediction from Bondi's formula and the $\mathcal{M} = 2.4$ result is close to the expected analytic result for large Mach speed. As seen in Fig. 11, except for the shocked region, the flow is nearly isotropic. The deviation from the stationary case was only at the 25% level.

Shima et al (1985) looked at the same problem as Hunt⁸, but used an integration scheme they considered to be better for capturing discontinuities. To test the accuracy of their scheme, they used it to calculate the flow of air past a rigid sphere and found that it conformed to experimental results. For the cases studied by Hunt, Shima et al also found a bow shock but with a twofold increase in accretion rate. They suggested that the accretion rate in Bondi's interpolation formula should be increased by the same factor, approaching the formula provided by Hoyle & Lyttleton for large Mach number.

Portnoy et al (1993) came to a different conclusion based on a model of accretion onto an *extended* gravitational source (see also Hunt 1975), rather than the point source of Hunt (1971). They found a bow shock only when they increased the mass of the Galaxy relative to its scale radius, consistent with the suggestion that they would have found a bow shock if they had used a point source. In general, they found a tail shock as suggested by Spiegel (1966). However, including replenishment of gas from stellar halo mass loss in their model allowed a bow shock to form, with the shock supported by this gas. A bow shock appeared if the ram pressure of the outside flow and the pressure of the replenished gas were comparable.

To date, there has been little observational evidence for something akin to a 'heliopause' ahead of either the Galaxy or the Local Group. These early ideas on accretion have strong echoes in numerical simulations of cluster galaxies generating wakes as they move through the intracluster medium, although here the focus tends to be on stripping galaxies of their gas (e.g. McCarthy et al 2008). Possibly related bow-shock structures have been observed on scales of galaxy clusters (e.g. Bagchi et al 2006): the outer boundary of the radio continuum emission is remarkably well confined within a circle defined approximately by the virial radius of the central cluster.

4.5. Cooling flows

Since the 1990s, the notion of galaxies sweeping up gas has been largely displaced by the paradigm of hot accretion. In this picture, gas and dark matter are accreted together onto central haloes, groups and clusters (Silk 1977; Rees & Ostriker 1977; White & Rees 1978). The gas becomes shock-heated to the virial temperature of the evolving dark matter system (White & Frenk 1993; Kaiser 1986). This picture finds support on megaparsec scales from the existence of hot coronae in groups and clusters.

Hot accretion has been a basic tenet of Λ CDM simulations for over twenty years. The belief was that the hot gas has time to cool in the central regions and accretes onto the central galaxy or group forming a 'cooling flow' (Fabian 1994). The mass cooling flux \dot{M} is assumed to be steady, spherical, and subsonic, and that any gradients in the potential are small compared to the square of the sound speed. In this case the pressure is constant, and mass conservation requires that $\dot{M} = 4\pi\rho v_i r^2$ where v_i is the inflow velocity. The cooling radius r_c interior to which the cooling time t_c is smaller than the age of the dark matter halo t_h is set by the condition $t_c \sim t_h$. The flow time from r_c is $t_f \sim r_c/v \sim 4\pi\rho_c r_c^3/\dot{M}$, where ρ_c is the gas density at r_c . We assume $t_f \sim t_c$, so that the gas has time to cool before reaching $r = 0$. This sets $v \sim r_c/t_h$ at r_c . If the cooling function $\Lambda(T_e)$ does not vary rapidly with temperature T_e , the density and temperature within r_c scale nearly as $\rho \propto r_c/r$, $T_e \propto r/r_c$ (Fabian & Nulsen 1977).

On the scales of galaxies, possible evidence for hot accretion of this form has been widely discussed (Maller & Bullock 2004; Kauffmann et al 2004) but never clearly demonstrated. In fact, beyond the Galaxy, the first reliable detections of disk galaxy coronae have only been possible in recent years for two of the most massive disk systems known (Anderson & Bregman 2011). When scaled to the mass of the Galaxy, the predicted halo accretion rate of $0.1 M_\odot \text{ yr}^{-1}$ is much too low to be a useful source of baryon accretion.

On group scales, unless the Local Group is very old, it is unlikely that a steady-state flow has been established (e.g. Tabor & Binney 1993), especially as infall of gas into the Local Group is likely to be intermittent and winding down over cosmic time (Maloney & Bland-Hawthorn 2001). Blitz et al (1999) made the spectacular suggestion that the HVC population is evidence of such an event on megaparsec scales, but these clouds are now known to lie within 50 kpc of the Galaxy (Bland-Hawthorn et al 1998; Putman et al 2003).

For massive clusters, Peterson et al (2003) carried out *XMM-Newton* spectroscopy of 14 clusters with the highest \dot{M} rates estimated from using the above analysis. A wide range of ions and ionisation states were detected and the results are unambiguous, a relatively rare occurrence in extragalactic astronomy. In all cases, *the cluster gas was found to be much too hot to support cooling flows*. For some (or all?) of these systems, there is evidence that central AGNs have kept the gas hot through intermittent jet interactions with the ICM. Thus there appears to be little or no support for cooling flows on galaxy to cluster scales.

4.6. Cold flows

The hot accretion picture leads to a delay in the star formation because the gas needs time to cool to trigger star formation. While being an early supporter of hot accretion, Binney (1977) also pointed out that some of the accreted gas is likely to be 'cool' (10^4K) because any dense clumps are unstable to cooling processes. High resolution simulations ultimately confirmed that much of the gas can cool and accrete through filamentary streams (Fardal et al 2001; Birnboim & Dekel 2003; Katz et al 2003; Keres et al 2005; Keres & Hernquist 2009). The most recent simulations (e.g. Fig. 12) find that cold flows dominate accretion onto all halo masses before $z \gtrsim 3$ and exclusively galaxies less massive than L_* since that time (Keres et al 2009; Brooks et al 2009) – this is not a view that is universally supported

⁸ Several authors comment on the difficulty of reproducing Hunt's classic papers in 1971 and 1975. The elegant analysis stymied by typographical errors that we have identified and corrected for; it is then possible to reproduce this important work.

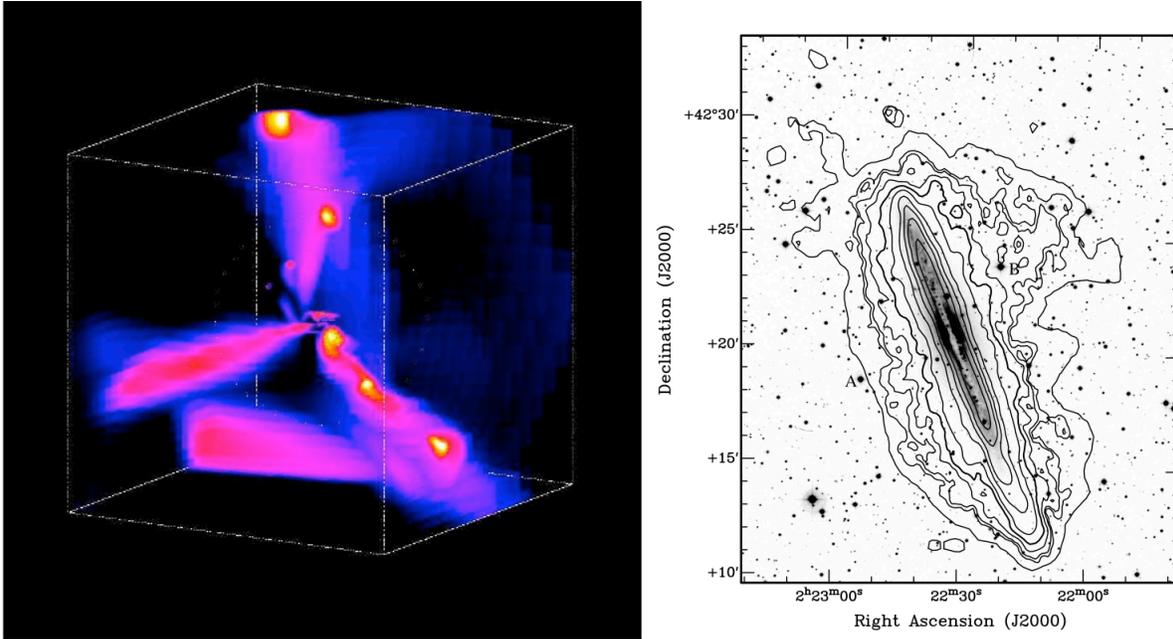


FIG. 12.— Cold gas flows onto a $10^{12} M_{\odot}$ dark matter halo observed in a 3D hydrodynamic simulation (Dekel et al 2009). The size of the box is about 300 kpc. Nothing like this has been observed directly although an intriguing HI filament (~ 25 kpc) has been observed in NGC 891, a galaxy that shares most of the Galaxy’s attributes (Oosterloo et al 2007).

(e.g. Benson & Bower 2011). The build-up of the hot halo initially helps the cold flow to slow down during accretion, but ultimately the halo can become quite dominant and shut off the accretion process.

The ‘cold flow’ paradigm has some interesting consequences, most notably in explaining angular momentum build-up in disks. In the early ‘hot accretion’ picture, the dark matter and gas in the extended halo acquire essentially all of their angular momentum through tidal torques with the surrounding mass distribution. This largely occurred at turn around ($z \sim 5$) when galaxies fell away from the Hubble flow. Within this paradigm, Fall & Efstathiou (1980) present what has become the standard paradigm for disk formation: the shocked gas settle to a disk which then smoothly accreted inwards while conserving angular momentum, originally proposed by Mestel. Thus, for example, the Galactic disk with its exponential scale length of ~ 5 kpc has conserved angular momentum from a collapsing gaseous halo of order ~ 300 kpc in size (see White 1984).

In the context of cold flows, Keres et al (2005) state that the infalling streams merge onto disks “like streams of cars entering an express way” converting some fraction of their infall velocity to rotational velocity. The ‘French school’ (Pichon et al 2011; Codis et al 2012; Kimm et al 2012) extend this idea to overturn earlier ideas on how angular momentum is advected through the virial sphere. In their simulations, they find that a typical L_{\star} galaxy has 2-3 cold flow streams although only one is particularly active at a given time. These are long-lived coherent flows that can build up a disk’s spin, scaling properties and orientation. The widths of these filaments do appear to depend on the resolution of the simulation (Rosdahl & Blaizot 2012).

At the present time, no clear evidence of ‘cold flows’ have been observed which is somewhat surprising given that they are believed to occur, albeit at a lower level, to the present day (Keres & Hernquist 2009). An interesting trend is to associate extended Ly α emission around galaxies at high redshift with cold flow accretion (e.g. Rauch et al 2011) although the situation is undoubtedly complicated (Rosdahl & Blaizot 2012). Other hints of cold accretion come from absorption site lines to high redshift quasars (Ribaudo et al 2011) but the evidence remains circumstantial.

Whether cold streams survive in future hydrodynamic simulations, with higher resolution and better adaptive mesh algorithms to capture important microprocesses, remains to be seen. But it is interesting to consider what disk properties are signatures of the formation process at high redshift; this is an issue we return to below in our discussions on near-field cosmology.

4.7. Warm flows

Λ CDM models consistently fail to produce realistic disks (Bullock et al 2001). Galaxy disks are thin and relatively fragile, and are easily disrupted or heated from infalling structure. Kormendy et al (2010) showed that two thirds of his galaxy sample are pure disks with only small central bulges⁹ and found no correspondence with the big-bulge thick disks arising naturally in the simulations. These are simply too violent to form the thin disks we observe today predicting very high merger rates of infalling dark matter clumps and too many clumps at all epochs (Springel et al 2005). The random motions between simulated galaxies are much higher than observed in peaceful suburbs like our own Local Group of galaxies, and so on.

⁹ Small central bulges are expected in pure disk evolution; see Sec. 6.1.

An even more extreme example of disks are the super thin disks first identified by Mathews & Gallagher (1999). The gathering consensus is that gas must enter a disk galaxy quietly (Freeman 2008). Newly contrived models that encourage quiescent gas accretion do seem to produce better looking galaxies (Brook et al 2011).

Marinacci et al (2011) have proposed coronal halo accretion driven by star formation in the disk driving gas shells into the halo. This followed from the realisation that the disk-halo interaction rate in the Galaxy ($\gtrsim 10 M_{\odot} \text{ yr}^{-1}$) is much larger than the required accretion rate onto the disk (Bland-Hawthorn 2008).

Another quiescent channel is a ‘warm flow’ that follows the breakdown of cold gas as it disrupts during its passage through the hot halo (Bland-Hawthorn 2009; Heitsch & Putman 2009). This process is clearly observed in the disruption of the Magellanic Stream in a shock cascade (Bland-Hawthorn et al 2007). These authors have shown that much of the Magellanic Stream is in the form of warm gas due to a shock cascade along the Stream; this warm gas is raining onto the Galaxy. The Magellanic Stream may well be the source of most high velocity clouds seen throughout the halo (Sec. 7.1.5). The warm gas accretion rate is at least as high as the HI accretion rate, and may even dominate. But the ionised+neutral accretion rates are still well below the levels required to sustain the present star formation rate.

4.8. Accretion via major and minor mergers

In our discussion of gas accretion onto galaxies, little has been said about whether gas and dark matter accrete together onto galaxies. If so, is the gas carried by the dark matter, or does it enter separately? Since we do not understand the main channels of gas accretion, this is a difficult question to answer.

Major mergers. The recent review by Sancisi et al (2008) of HI in the local universe finds that cold gas disconnected from galaxies is always associated with interactions between galaxies – in other words, there is no evidence for isolated (primordial?) HI gas at present. Major galaxy mergers and collisions involving both dark matter and baryons can be an important aspect of the accretion process (Hibbard & van Gorkom 1996).

The stellar record tells us that the Galaxy has not experienced a major accretion event (i.e. at least as massive as an L_{\star} galaxy) over the past 10 Gyr (Wyse 2001). Evidence of less massive gas-rich accretion events is more easy to come by. A particularly spectacular interaction phenomenon is the Magellanic HI Stream that trails from the LMC-SMC system (10:1 mass ratio) in orbit about the Galaxy. Since its discovery in the 1970s, there have been repeated attempts to explain the Stream in terms of tidal and/or viscous forces (q.v. Mastroiello et al 2005; Connors et al 2006). Indeed, the Stream has become a benchmark against which to judge the credibility of N-body+gas codes in explaining gas processes in galaxies. A fully consistent model of the Stream continues to elude even the most sophisticated codes. The total gas content of the Magellanic System falls well short of the required gas accretion rate ($\lesssim 0.2 M_{\odot} \text{ yr}^{-1}$; Putman et al 2012).

Minor mergers. Kauffman, Li & Heckman (2010) have investigated the possibility that dwarf galaxies are major sources of baryons. After compiling a large sample of isolated galaxies from the SDSS survey, they find strong correlations between the total mass of gas in satellites and the colours and specific star formation rates of central galaxies. The inferred accretion rate from the *observed* baryons is two orders of magnitude too small to account for the star formation rate in the central galaxy, which led them to postulate that dwarfs are harbouring large reserves of warm gas. An attractive aspect of the model is that warm gas is more easily concealed because its emissivity depends on the square of the gas density (Bland-Hawthorn et al 2007), unlike HI gas where the projected column has a linear dependence on density.

Warm gas within dwarf galaxies however is easily stripped by the Galactic halo through ram pressure stripping (Nichols & Bland-Hawthorn 2009, 2011). The stripped gas becomes part of the warm-hot halo and can conceivably seed clouds for later accretion (Joung et al 2012). An additional problem is the lack of strong evidence for massive hot haloes around dwarf galaxies. Gallagher et al (2003) used the Wisconsin H α Mapper (WHAM) to obtain very deep limits over a 1° beam on the presence of ionised gas in the Draco and Ursa Minor dwarf spheroidals. No warm gas was seen down to a mass limit that is of order 10% of the stellar mass fraction in both objects.

4.9. Accretion of high velocity clouds

4.9.1. Decelerating gas clouds in a hot corona

We now consider the infall of individual HI clouds towards the Galactic disk. Do they survive to reach the disk or break up to become part of the hot corona? This important question bears on how much warm/hot gas is presently concealed in the Galactic corona. The question is very difficult to settle from direct observation because emission from a relatively cool Galactic corona ($T_e \approx 2 \times 10^6 \text{ K}$) peaks in the extreme ultraviolet region.

The Galactic halo is home to many high velocity HI clouds (HVCs) and gas streams of unknown origin (Wakker 2001; Lockman et al 2008), although the Magellanic Stream is a likely source for a significant fraction. The survival and stability of these clouds is a problem that has long been recognized (e.g. Benjamin & Danly 1997). It is likely that many or all halo clouds have experienced some deceleration during their transit through the hot gas in the lower halo.

To understand HVCs, we consider the drag equation for an HI cloud moving through a stationary medium,

$$\mu_c \dot{v}_c = \frac{1}{2} C_D \rho_h(z) v_c^2 - \mu_c g(z) \quad (24)$$

where μ_c is the surface density of the cloud. Equation 24 only holds as long as the cloud stays together. The drag

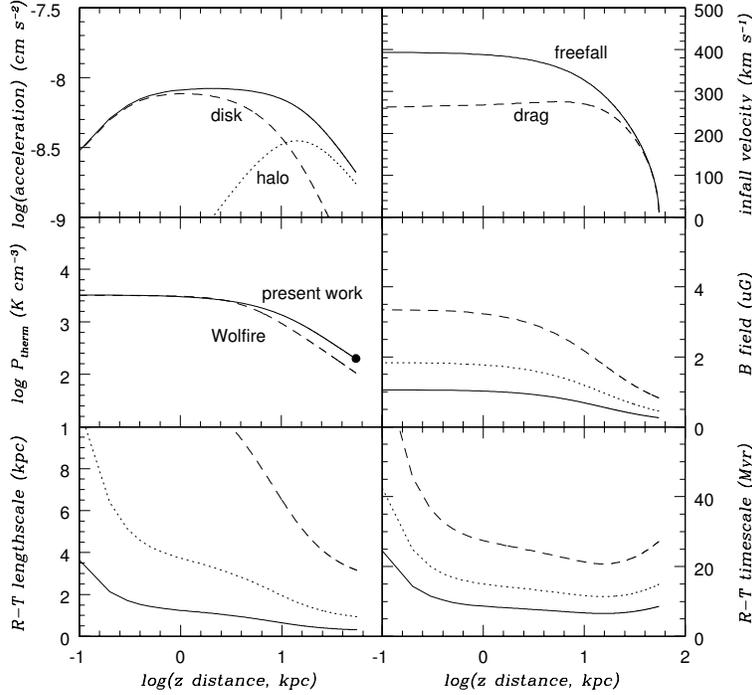


FIG. 13.— (a) gravitational acceleration due to the disk+halo – all plots are shown as a function of vertical height above the disk at the Solar Circle; (b) infall velocity for a point mass starting from the Magellanic Stream with halo drag ($C_D = 1$; dashed curve) and without (solid); (c) coronal gas pressure for our model (solid line) compared to the Wolfire model (dashed) – the dot indicates the halo pressure used in our hydrodynamical model; (d) magnetic field strength for $\beta = 1$ (dashed), $\beta = 0.3$ (dotted), $\beta = 0.1$ (solid); (e) minimum lengthscale for RT instability; (e) timescale for RT instability.

coefficient C_D is a measure of the efficiency of momentum transfer to the cloud. For the high Reynolds numbers typical of astrophysical media, incompressible objects have $C_D \approx 0.4$ (e.g. a rough sphere) which indicates that the turbulent wake behind the plunging object efficiently transfers momentum to the braking medium. The leading face of a compressible cloud may become flattened, such that the approaching medium is brought to rest in the reference frame of the cloud; in this instance, $C_D \gtrsim 1$ may be more appropriate (we adopt $C_D = 1$ here).

Let us consider the Galactic hot corona to be an isothermal gas in hydrostatic equilibrium with the gravitational potential, $\phi(R, z)$, where R is the Galactocentric radius and z is the vertical scale height. We adopt a total potential of the form $\phi = \phi_d + \phi_h$ for the disk and halo respectively; for our calculations at the Solar Circle, we ignore the Galactic bulge. The galaxy potential is defined by

$$\phi_d(R, z) = -c_d v_{circ}^2 / (R^2 + (a_d + \sqrt{z^2 + b_d^2})^2)^{0.5} \quad (25)$$

$$\phi_h(R, z) = c_h v_{circ}^2 \ln((\psi - 1)/(\psi + 1)) \quad (26)$$

and $\psi = (1 + (a_h^2 + R^2 + z^2)/r_h^2)^{0.5}$. The scaling constants are $(a_d, b_d, c_d) = (6.5, 0.26, 8.9)$ kpc and $(a_h, r_h) = (12, 210)$ kpc with $c_h = 0.33$ (e.g. Miyamoto & Nagai 1975). The circular velocity $v_{circ} \approx 220$ km s $^{-1}$ is now well established through wide-field stellar surveys (e.g. Smith et al 2007).

We determine the vertical acceleration at the Solar Circle using $g = -\partial\phi(R_o, z)/\partial z$ with $R_o = 8$ kpc. The hydrostatic halo pressure follows from

$$\frac{\partial\phi}{\partial z} = -\frac{1}{\rho_h} \frac{\partial P}{\partial z} \quad (27)$$

We adopt a solution of the form $P_h(z) = P_o \exp((\phi(R_o, z) - \phi(R_o, 0))/\sigma_h^2)$ where σ_h is the isothermal sound speed of the hot corona. To arrive at P_o , we adopt a coronal halo density of $n_{e,h} = 10^{-4}$ cm $^{-3}$ at the Stream distance (55 kpc) in order to explain the Magellanic Stream H α emission (Bland-Hawthorn et al 2007), although this is uncertain to a factor of a few. We choose $T_h = 2 \times 10^6$ K to ensure that OVI is not seen in the diffuse corona consistent with observation (Sembach et al 2003); this is consistent with a rigorously isothermal halo for the Galaxy. Our solution to the above equation is shown in Fig. 13(c) and it compares favorably with the pressure profile derive by others (e.g. Sternberg et al 2002).

A solution specific to our model is shown in Fig. 13(b) where the freefall velocity is now slowed by about 35%. In practice, the cloud's projected motion can be considerably less than its 3D space velocity (e.g. Lockman et al 2008). In all likelihood, infalling HVCs have experienced significant deceleration through ram pressure exerted by the corona.

But even before the cloud reaches terminal velocity, the cloud is expected to break up (Murray & Lin 2004; Heitsch & Putman 2009).

So how do clouds resist the destructive forces of RT and shock instabilities? One prospect is the stabilizing influence of magnetic fields when a cloud passes through a magnetized medium. The halo magnetic field is poorly constrained at the present time (Sun et al 2008). We describe the uniform magnetic field in terms of the pressure of the halo medium, or

$$\frac{B^2}{8\pi} = \beta P_h \quad (28)$$

such that $B \approx 1 \mu\text{G}$ at the distance of the Stream (55 kpc) if the field is in full equipartition with the corona. But there is evidence that the field is weaker than implied here ($\beta \approx 0.3$; Sun et al 2008), at least within 5 kpc of the Galactic plane. For the warm, denser low-latitude gas (Reynolds layer), we adopt the new parameter fits of Gaensler et al (2008) from a re-analysis of pulsar data. The lower β value finds support from recent magnetohydrodynamic simulations of the Reynolds layer (Hill et al 2008).

4.9.2. Stability limits and growth timescales

We consider the surface of a high velocity cloud as a boundary between two fluids. In practice, the Galactic ionizing radiation field imparts a multiphase structure to the cloud. At all galactic latitudes within the Stream distance, HVCs with column densities of order 10^{20} cm^{-2} or higher have partially ionized skins to a column depth of roughly 10^{19} cm^{-2} for sub-solar gas due to the Galactic ionizing field. Between the warm ionized skin and the cool inner regions is a warm neutral medium of twice the skin thickness; both outer layers have a mean particle temperature of $\lesssim 10^4 \text{ K}$.

The cloud is denser than the halo gas. Because of the gravitational field, RT instabilities can grow on the boundary. Furthermore, KH instabilities may also develop due to the relative motion of the cloud with respect to an external medium. Buoyant bubbles in galaxy clusters are stabilized against RT and KH instabilities by viscosity and surface tension due to magnetic fields in the boundary (De Young 2003). Here we examine whether HVC boundaries are similarly stabilized against disruption in the Galactic halo.

When there is no surface tension, no viscosity and no relative motion between the two media, the growth rate of the RT instability for a perturbation with wavenumber k is $\omega = \sqrt{gk}$, where g is the gravitational acceleration at the fluid boundary. The wavenumber is related to a perturbation length scale, $\ell = 2\pi/k$. The instability requires a few e-folding timescales to fully develop; the timescale is given by (De Young 2003)

$$t_{grow} = \omega^{-1} = \sqrt{\frac{\ell}{2\pi g}}. \quad (29)$$

In the presence of a magnetic field, the transverse component (B_{tr}) provides some surface tension which can help to suppress RT instabilities below a lengthscale of (Chandrasekhar 1961)

$$\ell_{min} = \frac{B_{tr}^2}{2\rho_c g}. \quad (30)$$

Here ρ_c is the mass density of the denser medium, i.e. the cloud, and B_{tr} is the average value of the transverse magnetic field at the boundary.

In order to illustrate when RT instabilities become important, we assume a flat rotation curve for the Galaxy such that (Kaffe et al 2012)

$$g \approx \frac{v_{circ}^2}{R} = 1.6 \times 10^{-8} \left(\frac{v_{circ}}{220 \text{ km s}^{-1}} \right)^2 \left(\frac{R}{R_o} \right)^{-1} \text{ cm s}^{-2}. \quad (31)$$

This is only a rough approximation to the form expected from the coupled equations in Eq. 26. We stress that the actual behaviour discussed below, and shown in Fig. 13(e) and (f), solves for the gravitational potential correctly.

Shortly after the discovery of HVCs, it was thought that they may be self-gravitating. But this would place them at much greater distances than the Magellanic Stream (e.g. Oort 1966) which is now known not to be the case (e.g. Putman et al 2003). Instead, we consider two cases: (i) HVCs in pressure equilibrium with the coronal gas; (ii) HVCs with parameters fixed by direct observation. In (i), because the temperature is not strongly dependent on radius, but the number density decreases rapidly with increasing radius, we expect the increased pressure to compress the clouds at lower latitudes.

We estimate the impact of RT instabilities using equations 27 and 29 for a cloud temperature of $T_c = 10^4 \text{ K}$ in pressure equilibrium with the hot halo, the electron density is given by

$$\begin{aligned} n_{e,c} &\approx n_{e,h} \frac{T_h}{T_c} \\ &= 0.02 \left(\frac{R}{55 \text{ kpc}} \right)^{-2} \left(\frac{T_h}{2 \times 10^6 \text{ K}} \right) \left(\frac{T_c}{10^4 \text{ K}} \right)^{-1} \text{ cm}^{-3}. \end{aligned} \quad (32)$$

We use equations 30, 32 and 28 to estimate ℓ_{min} as a function of Galactocentric radius. The minimum length scale for instability is

$$\begin{aligned} \frac{\ell_{min}}{R} &\sim \frac{8\pi\beta k_B T_c}{m_p v_{circ}^2} \\ &= 0.004 \left(\frac{T_c}{10^4 \text{K}} \right) \left(\frac{\beta}{0.1} \right) \left(\frac{v_{circ}}{220 \text{ km s}^{-1}} \right)^{-2} \end{aligned} \quad (33)$$

and its associated growth timescale using equation (29)

$$\begin{aligned} t_{grow} &\sim \sqrt{\frac{4\beta k_B T_c}{m_p v_{circ}^2}} \Omega^{-1} \\ &= 1.1 \text{ Myr} \left(\frac{T_c}{10^4 \text{K}} \right)^{\frac{1}{2}} \left(\frac{\beta}{0.1} \right)^{\frac{1}{2}} \left(\frac{v_{circ}}{220 \text{ km s}^{-1}} \right)^{-2} \left(\frac{R}{R_o} \right) \end{aligned} \quad (34)$$

where the angular rotation rate is given by $\Omega = v_{circ}/R$.

Because we have assumed that $B^2 \propto n_{e,h}$ (equipartition) and $n_{e,c} \propto n_{e,h}$ (pressure equilibrium), neither the minimum scale length or its growth timescale depend on the halo density or temperature. They do depend on the temperature of the clouds and the ionization state. If the clouds are hotter than 10^4K , then $n_{e,c}$ is overestimated under the assumption of pressure equilibrium. This would lead to larger minimum instability lengthscales and growth timescales. If the Galactic rotation curve drops faster than the flat profile implied by Eq. 27, we would have underestimated both the minimum instability scale length and its associated growth timescale at large radii.

Under the assumption of pressure equilibrium, the falling clouds become more compressed as they approach the disk which can hasten cooling. This effect may help to stabilize against break up, particularly if a cool shell develops (cf. Sternberg & Soker 2008).

In the absence of gravitational instability, the flow is stable against the KH instability if (Chandrasekhar 1961)

$$U^2 < \frac{B_{tr}^2 (\rho_c + \rho_h)}{2\pi \rho_c \rho_h} \quad (35)$$

where U is the relative velocity between the two fluids. When $\rho_c > \rho_h$, this requirement becomes

$$\begin{aligned} U &< \sqrt{\frac{8\beta k_B T_h}{m_p}} \\ &= 115 \left(\frac{\beta}{0.1} \right)^{\frac{1}{2}} \left(\frac{T_h}{2 \times 10^6 \text{K}} \right)^{\frac{1}{2}} \text{ km s}^{-1} \end{aligned} \quad (36)$$

and we have described the magnetic field in terms of the halo pressure using equation 28. This requirement is also independent of the halo density as we have related the magnetic field to the halo pressure, although it is dependent on the halo temperature. This requirement is nearly satisfied for HVCs if the magnetic field is near equipartition.

5. NEAR FIELD COSMOLOGY

5.1. Introduction

The formation and evolution of galaxies is one of the great outstanding problems of astrophysics. Within the broad context of hierarchical structure formation, we have only a crude picture of how galaxies like our own came into existence¹⁰. A detailed physical picture where individual stellar populations can be associated with (tagged to) elements of the protocloud is far beyond our current understanding.

As we have shown, important clues about the sequence of events involved in galaxy formation are already apparent from the high redshift universe (far-field cosmology) but now we turn our attention to fossil evidence provided by the Galaxy (near-field cosmology). Detailed studies of the Galaxy lie at the core of understanding the complex processes involved in baryon dissipation. This is a necessary first step towards achieving a successful theory of galaxy formation.

What do we mean by the reconstruction of early galactic history? We seek a detailed physical understanding of the sequence of events which led to the Milky Way. Ideally, we would want to tag (i.e. associate) components of the Galaxy to elements of the protocloud – the baryon reservoir which fueled the stars in the Galaxy.

From theory, our prevailing view of structure formation relies on a hierarchical process driven by the gravitational forces of the large-scale distribution of cold, dark matter (Λ CDM). The Λ CDM paradigm provides simple models of galaxy formation within a cosmological context (Peebles 1974; White & Rees 1978; Blumenthal et al 1984). N-body and semi-analytic simulations of the growth of structures in the early universe have been successful at reproducing some of the properties of galaxies. Current models include gas pressure, metal production, radiative cooling and heating, and prescriptions for star formation.

¹⁰ The idea of hierarchical formation predates Λ CDM (q.v. de Vaucouleurs 1970a)

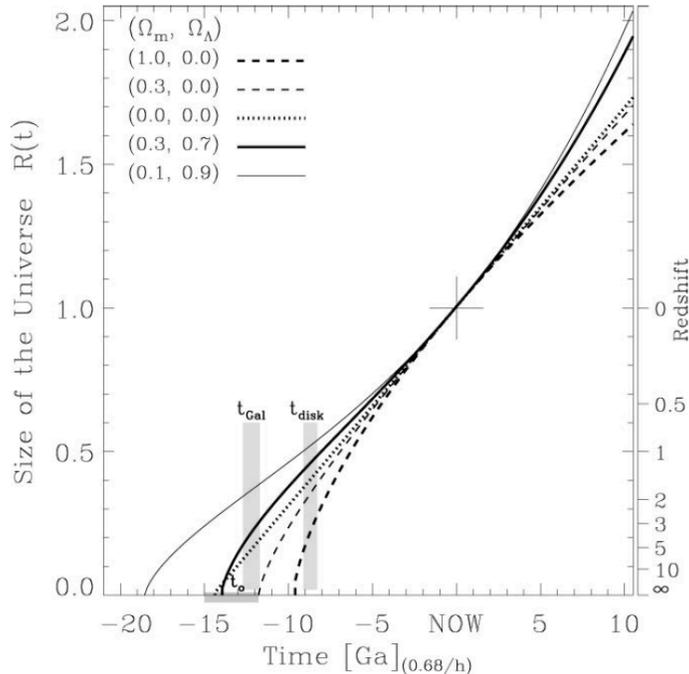


FIG. 14.— Look-back time as a function of redshift and the size of the Universe for five different world models. The approximate ages of the Galactic halo and disk are indicated by hatched regions. [We thank C. Lineweaver for modifying an earlier version of this figure for us.]

The number density, properties and spatial distribution of dark matter halos are well understood within Λ CDM (Sheth & Tormen 1999; Jenkins et al 2001). However, computer codes are far from producing realistic simulations of how baryons produce observable galaxies within a complex hierarchy of dark matter. This a necessary first step towards a viable theory or a working model of galaxy formation.

In this review, our approach is anchored to observations of the Galaxy, interpreted within the broad scope of the Λ CDM hierarchy. Many of the observables in the Galaxy relate to events which occurred long ago, at high redshift. Fig. 14 shows the relationship between look-back time and redshift in the context of the Λ CDM model: the redshift range ($z \lesssim 6$) of discrete sources in contemporary observational cosmology matches closely the known ages of the oldest components in the Galaxy. The Galaxy (near-field cosmology) provides a link to the distant universe (far-field cosmology).¹¹

EGGEN, LYNDEN-BELL & SANDAGE (1962) were the first to show that it is possible to study galactic archaeology using stellar abundances and stellar dynamics; this is probably the most influential paper on the subject of galaxy formation. ELS studied the motions of high velocity stars and discovered that, as the metal abundance decreases, the orbit energies and eccentricities of the stars increased while their orbital angular momenta decreased. They inferred that the metal-poor stars reside in a halo that was created during the rapid collapse of a relatively uniform, isolated protogalactic cloud shortly after it decoupled from the universal expansion. ELS is widely viewed as advocating a smooth monolithic collapse of the protocloud with a timescale of order 10^8 years.

The ELS picture was challenged by SEARLE (1977) and SEARLE & ZINN (1978) who noted that Galactic globular clusters have a wide range of metal abundances essentially independent of radius from the Galactic Center. They suggested that this could be explained by a halo built up over an extended period from *independent* fragments with masses of $\sim 10^8 M_{\odot}$. In contrast, in the ELS picture, the halo formed in a rapid free-fall collapse. But halo field stars, as well as globular clusters, are now believed to show an age spread of 2–3 Gyr (Marquez & Schuster 1994); for an alternative view, see Sandage & Cacciari (1990). The current paradigm, that the observations argue for a halo that has built up over a long period from infalling debris, has developed after many years of intense debate.

This debate paralleled the changes that were taking place in theoretical studies of cosmology (e.g. Peebles 1971; Press & Schechter 1974). The ideas of galaxy formation via hierarchical aggregation of smaller elements from the early universe fit in readily with the Searle & Zinn view of the formation of the galactic halo from small fragments. The possibility of identifying debris from these small fragments was already around in Eggen's early studies of moving groups, and this is now an active field of research in theoretical and observational stellar dynamics. It offers the possibility to reconstruct at least some properties of the protogalaxy and so to improve our basic understanding of the galaxy formation process.

We can extend this approach to other components of the Galaxy. We will argue the importance of understanding the formation of the galactic disk, because this is where most of the baryons reside. Although much of the information about the properties of the protogalactic baryons has been lost in the dissipation that led to the galactic disk, a similar

¹¹ These terms were first used in Bland-Hawthorn (1999).

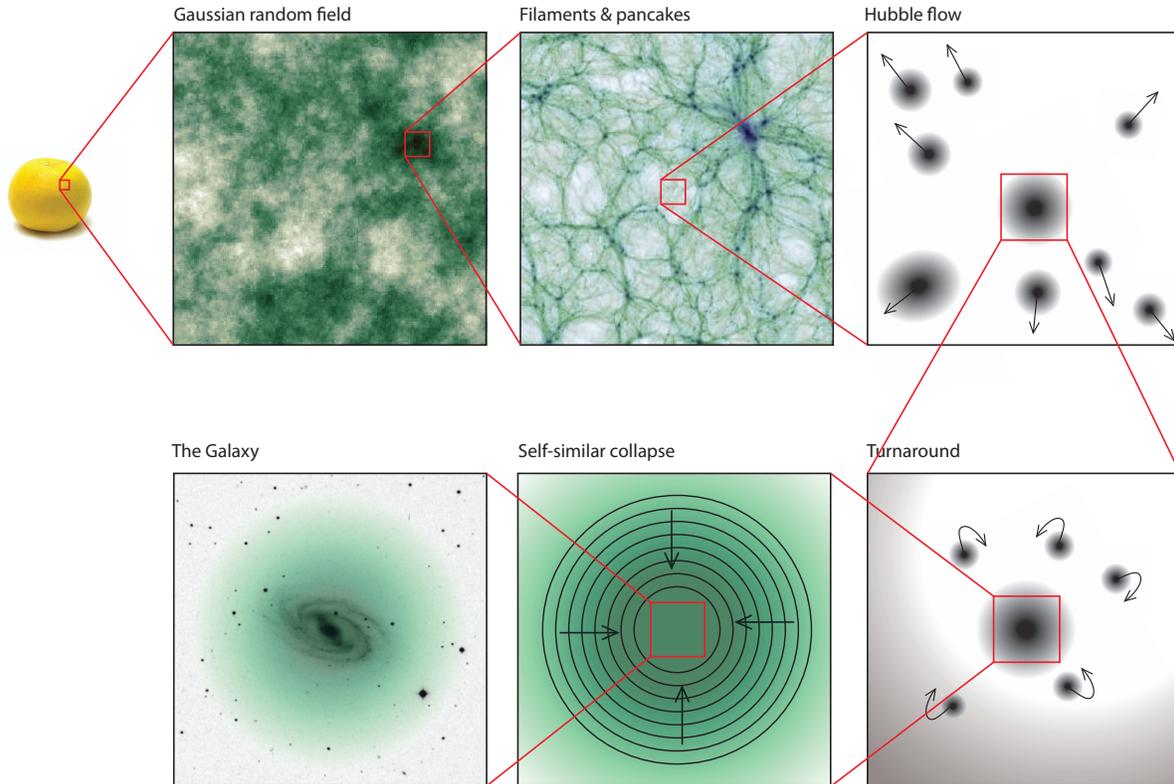


FIG. 15.— A sequence of images depicting the major stages in the evolution of the Galaxy from the initial fluctuation spectrum — see text. This was preceded by inflation when the Universe went from sub-atomic scales to the size of a grapefruit in less than a picosecond.

dynamical probing of the early properties of the disk can illuminate the formation of the disk, at least back to the epoch of last significant dissipation. It is also clear that we do not need to restrict this probing to stellar dynamical techniques. A vast amount of fossil information is locked up in the detailed stellar distribution of chemical elements in the various components of the Galaxy, and we will discuss the opportunities that this offers.

Before we embark on a detailed overview of the relevant data, we give a descriptive working picture of the sequence of events involved in galaxy formation. For continuity, the relevant references are given in the main body of the review where these issues are discussed in more detail. **In Appendix B, we provide extra information on stellar data with an emphasis on how to find important source material. We also summarise techniques and methods to assist the aspiring galactic archaeologist (near field cosmologist).**

5.2. A working model of how the Galaxy formed

Shortly after the Big Bang, cold dark matter began to drive baryons towards local density enhancements. The first stars formed after the collapse of the first primordial molecular clouds. The universe was ionized again in a process that is thought to have commenced ~ 200 million years after the Big Bang and is observed to be complete by the time the universe was 10 times that age. The study of this reionization is a topic for current research, but we do know that the first generations of massive stars likely played a major role. Observations of distant galaxies, seen as they were in the past because of the light travel time, show that many young galaxies were strong sources of ionizing radiation, but observations must reach still greater distances to show us the earliest generations of stars. This is a key science goal of the next generation of large telescopes, including the James Webb Space Telescope due to launch at the end of the decade.

In Fig. 15, we provide a cartoon of how the Galaxy formed from the initial quantum fluctuations. By a redshift of about $z \sim 10$, sufficient dark matter clumps (and associated baryons) were able to fall out of the Hubble flow to form modern day L_* galaxies. Within the context of Λ CDM, the dark halo of the Galaxy assembled first, although it is likely that its growth continues to the present time. In some galaxies, the first episodes of gas accretion established the stellar bulge, the central black hole, the first halo stars and the globular clusters. In the Galaxy and similar systems, the small stellar bulge may have formed later from stars in the inner disk.

The early stages of the Galaxy's evolution were marked by violent gas dynamics and accretion events, leading to the high internal densities of the first globular clusters, and perhaps to the well-known 'black hole mass – stellar bulge dispersion' relation. The stellar bulge and massive black hole may have grown up together during this active time.

We associate this era with the ‘Golden Age’, the phase before $z \sim 1$ when star formation activity and accretion disk activity were at their peak.

At that time, there was a strong metal gradient from the bulge to the outer halo. The metal enrichment was rapid in the core of the Galaxy such that, by $z \sim 1$, the mean metallicities were as high as $[\text{Fe}/\text{H}] \sim -1$ or even higher. In these terms, we can understand why the inner stellar bulge that we observe today is both old and moderately metal rich. The first halo stars ($[\text{Fe}/\text{H}] \approx -5$ to -2.5) formed over a more extended volume and presumably date back to the earliest phase of the protocloud. The first globular clusters formed over a similar volume from violent gas interactions ($[\text{Fe}/\text{H}] \approx -2.5$ to -1.5). We believe now that many of the halo stars and globulars are remnants of early satellite galaxies which experienced *independent* chemical evolution before being accreted by the Galaxy.

The spread in $[\text{Fe}/\text{H}]$, and the relative distribution of the chemical elements, is a major diagnostic of the evolution of each galactic component. If the initial mass function is constant, the mean abundances of the different components give a rough indication of the number of SN II enrichments which preceded their formation, although we note that as time passes, an increasing fraction of Fe is produced by SN Ia events. For a given parcel of gas in a closed system, only a few SN II events are required to reach $[\text{Fe}/\text{H}] \approx -3$, 30 to 100 events to get to $[\text{Fe}/\text{H}] \approx -1.5$, and maybe a thousand events to reach solar metallicities. We wish to stress that $[\text{Fe}/\text{H}]$ is not a clock: rather it is a measure of supernova occurrences and the depth of the different potential wells that a given parcel of gas has explored.

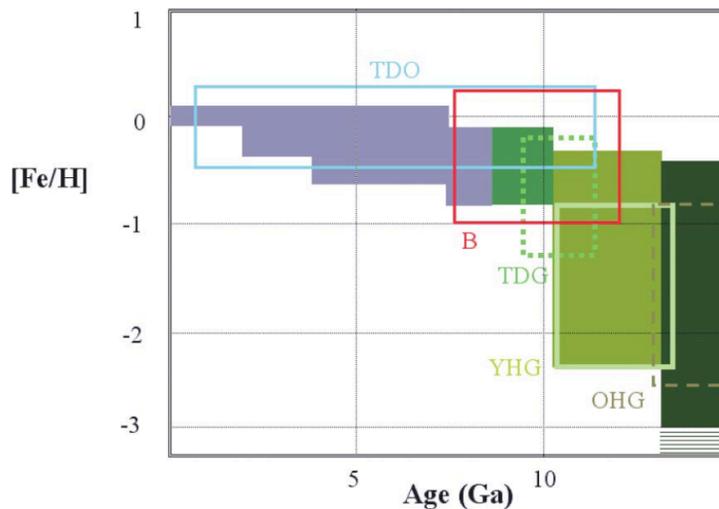


FIG. 16.— The age-metallicity relation of the Galaxy for the different components (see text): TDS – thin disk stars; TDO – thin disk open clusters; ThDS – thick disk stars; ThDG – thick disk globulars; B – bulge; YHG – young halo globulars; OHG – old halo globulars.

During the latter stages of the Golden Age, most of the baryons began to settle to a disk for the first time. Two key observations emphasize what we consider to be the mystery of the main epoch of baryon dissipation. First, there are no stars with $[\text{Fe}/\text{H}] < -2.2$ which rotate with the disk. Secondly, despite all the activity associated with the Golden Age, at least 80% of the baryons appear to have settled gradually to the disk over many Gyr; this fraction could be as high as 95% if the bulge formed after the disk.

About 10% of the baryons reside in a ‘thick disk’ which has $[\text{Fe}/\text{H}] \approx -2.2$ to -0.5 , compared to the younger thin disk with $[\text{Fe}/\text{H}] \approx -0.5$ to $+0.3$. It is striking how the globular clusters and the thick disk have similar abundance ranges, although the detailed abundance distributions are different. There is also a similarity in age: globular clusters show an age range of 12 to 14 Gyr, and the thick disk appears to be at least 12 Gyr old. Both the thick disk and globulars apparently date back to the epoch of baryon dissipation during $z \sim 1 - 5$. Fig. 16 summarises our present understanding of the complex age–metallicity distribution for the various components of the Galaxy.

It is a mystery how the thick disk and the globulars should have formed so early *and* over such a large volume from material which was already enriched to $[\text{Fe}/\text{H}] \sim -2$. Could powerful winds from the central starburst in the evolving core have distributed metals throughout the inner protocloud at about that time?

While half of all baryons were residing within the Galaxy by $z \sim 1$, we emphasize that 90% of the *disk baryons* have settled quiescently to the thin disk since $z \sim 1$. The disk star formation rate during this extended period was roughly constant (Binney et al 2000) or possibly in slow decline (Aumer & Binney 2009). The rapid decline in ρ_* observed in deep galaxy surveys since $z \sim 1$ is not in evidence over the volume of the Local Group.

5.3. Timescales and fossils

The oldest stars in our Galaxy are of an age similar to the look-back time of the most distant galaxies in the Hubble Deep Field (Fig. 14). For the galaxies, the cosmological redshift measured from galaxy spectra presently takes us to within a few percent of the origin of cosmic time. For the stars, their upper atmospheres provide fossil evidence of the available metals at the time of formation. The old Galactic stars and the distant galaxies provide a record of

conditions at early times in cosmic history, and both harbor clues to the sequence of events which led to the formation of galaxies like the Milky Way. The near field also provides important evolutionary timescales for individual stars and groups of stars, as we discuss in the next section.

The key timescale provided by far-field cosmology is the look-back time with the prospect of seeing galaxies at an earlier stage in their evolution. However, this does *not* imply that these high-redshift objects are unevolved. We know that the stellar cores of galaxies at the highest redshifts ($z \sim 5$) observed to date exhibit solar metallicities, and therefore appear to have undergone many cycles of star formation (Hamann & Ferland 1999). Much of the light we detect from the early universe probably arises from the chemically and dynamically evolved cores of galaxies.

Near-field cosmology provides a dynamical timescale, $\tau_D \sim (G\rho)^{-\frac{1}{2}}$, where ρ is the mean density of the medium. The dynamical timescale at a radial distance of 100 kpc is of order several Gyr, so the mixing times are very long. Therefore, on larger scales, we can expect to find dynamical and chemical traces of past events, even where small dynamical systems have long since merged with the Galaxy.

We note that the Λ CDM hierarchy reflects a wide range of dynamical timescales, such that different parts of the hierarchy may reveal galaxies in different stages of evolution. In this sense, the hierarchy relates the large-scale density to the morphology and evolution of its individual galaxies; this is the so-called ‘morphology-density relation’ (Dressler 1980; Hermit et al 1996; Norberg et al 2001). Over a large enough ensemble of galaxies, taken from different regions of the hierarchy, we expect different light-weighted age distributions because one part of the hierarchy is more evolved than another (see Sec. 3.1).

5.4. Stellar age dating

Stellar ages let us evaluate when events occurred. They are important for measuring the star formation history and for understanding how the metallicity and dynamics of different groups of stars have evolved: e.g. how has the star formation rate and the kinematics and the metallicity of the thin disk near the sun changed from 10 Gyr ago to the present time? Current evidence, based on chromospheric stellar ages, indicates that the star formation rate in the galactic thin disk near the sun has been roughly uniform over the last 10 Gyr, with episodic increases and decreases of about a factor 2 (e.g. Rocha-Pinto et al 2000). Stellar ages are still difficult to measure and lead to much uncertainty about the evolution of the Galaxy. Measuring stellar ages is one of the most important and challenging goals for the future.

Nucleo-cosmochronology (or cosmochronometry), or the aging of the elements through radioactive decay, has a long history (Fowler & Hoyle 1960; Butcher 1987). A related technique is widely used in solar system geophysics. Independent schemes have aged the oldest meteorites at 4.53 ± 0.04 Gyr (Guenther & Demarque 1997; Manuel 2000). The small uncertainties reflect that the age dating is direct. Element pairs like Rb and Sr are chemically distinct and freeze out during solidification into different crystalline grains. The isotope ^{87}Rb decays into ^{87}Sr which can be compared to ^{86}Sr , a non-radiogenic isotope, measured from a control sample of Sr-rich grains. This provides a direct measure of the fraction of a ^{87}Rb half-life ($\tau_{1/2} = 47.5$ Gyr) since the meteorite solidified.

It appears that, until we have a precise understanding of BBNS and the early chemical evolution history of the Galaxy, geophysical precision will not be possible for stellar ages. The major problem is that, as far as we know, there is no chemical differentiation which requires that we know precisely how much of each isotope was originally present. Modern nucleo-cosmochronology compares radioactive isotope strengths to a stable r-process element (e.g. Nd, Eu, La, Pt). The thorium method (^{232}Th , $\tau_{1/2} = 14.0$ Gyr) was first applied by Butcher (1987) and refined by Pagel (1989). Other radioactive chronometers include ^{235}U ($\tau_{1/2} = 0.70$ Gyr) and ^{238}U ($\tau_{1/2} = 4.47$ Gyr) although Yokoi et al (1983) have expressed concerns about their use (cf. Cayrel et al 2001). Arnould & Goriely (2001) propose that the isotope pair ^{187}Re – ^{187}Os ($\tau_{1/2} = 43.5$ Gyr) may be better suited for future work.

With the above caveats, we point out that several groups are now obtaining exquisite high-resolution data on stars with enhanced r-process elements (Cayrel et al 2001; Sneden et al 2000; Burris et al 2000; Westin et al 2000; Johnson & Bolte 2001; Cohen et al 2002; Hill et al 2002). For a subset of these stars, radioactive ages have been derived (Truran et al 2001) normalized to the heavy element abundances observed in meteorites.

There are few other direct methods for deriving ages of *individual* stars. A promising field is asteroseismology which relies on the evolving mean molecular weight in stellar cores (Christensen-Dalsgaard 1986; Ulrich 1986; Gough 1987; Guenther 1989). The stellar oscillation frequencies depend on density distribution in stellar interior, which changes as the star ages. The asteroseismology space missions (MOST, CoRoT, Kepler) will contribute greatly to deriving stellar ages with 5-10% errors. Gough (2001) has determined 4.57 ± 0.12 Gyr for the Sun which should be compared with the age of meteorites quoted above.

There are age indicators based on stellar rotation and chromospheric activity which are particularly useful for main sequence stars (see review by Barnes 2007). These indicators are calibrated on star clusters and the Sun. Stars spin down with age: their rotation periods are usually measured photometrically (Kepler, COROT) and can be used to estimate age – this technique is known as gyrochronology. Chromospheric Ca K emission has been much used as an age indicator in the past, but is believed to be less accurate for older stars. This method is based on the decrease of chromospheric activity with age. Barnes (2007) presents a figure of chromospheric ages against gyrochronology ages for a sample of well studied stars (see Fig. 17).

Other traditional methods rely on aging a population of stars that are representative of a particular component of the Galaxy. For example, Gilmore et al (1989) use the envelope of the distribution in a color-abundance plane to show that all stars more metal poor than $[\text{Fe}/\text{H}] = -0.8$ are as old as the globular clusters. Similarly, the faint end

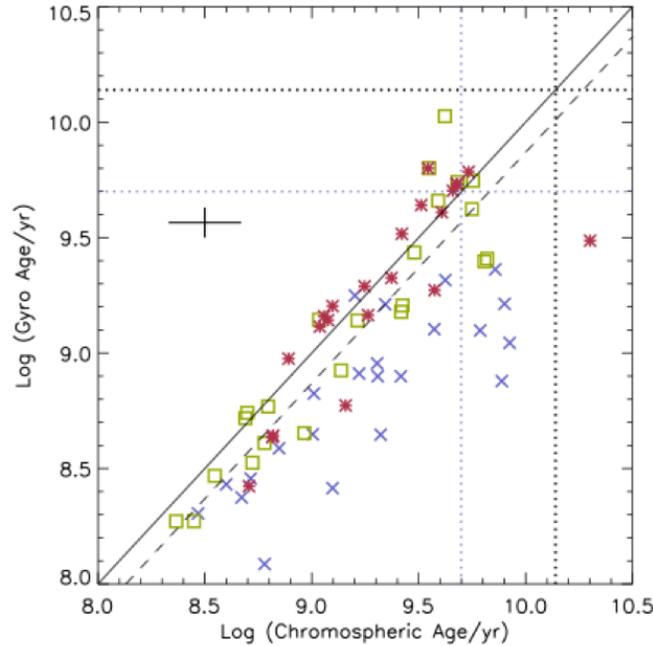


FIG. 17.— Gyro ages vs chromospheric ages for a sample of well-studied stars (Barnes 2007). The blue crosses are for stars bluer than $B - V = 0.6$ and should be disregarded.

of the white dwarf luminosity function is associated with the coolest, and therefore the oldest, stars (Oswalt et al 1996). The white dwarf luminosity function is useful for deriving ages of a population (eg the disk or the halo) or for a globular cluster. This method uses the cooling and fading of white dwarfs as an age estimator for the population. For example, Leggett et al (1998) used the luminosity function of white dwarfs in the nearby Galactic disk to derive an age of about 9 Gyr for the oldest stars of the disk. A comparison of the white dwarf luminosity function for the Galactic disk and two globular clusters, M4 and NGC 6397, by Hansen et al 2007, show that the disk is clearly a few Gyr younger than the clusters.

It has long been known that disk stars span a wide range of ages from the diversity of main-sequence stars. Edvardsson et al (1993) derived precise stellar evolution ages for nearby individual post main-sequence F stars using Strömgen photometry, and showed that the stars in the Galactic disk exhibit a large age spread with ages up to roughly 10 Gyr (Fig. 16). Using the ‘inverse age-luminosity relation’ for RR Lyrae stars, Chaboyer et al (1996) found that the oldest globular clusters are older than 12 Gyr with 95% confidence, with a best estimate of 14.6 ± 1.7 Gyr (Chaboyer 1998). But Hipparcos appears to show that the RR Lyr distances are underestimated leading to a downward revision of the cluster ages: 8.5 – 13.3 Gyr (Gratton et al 1997); 11 – 13 Gyr (Reid 1998); 10.2 – 12.8 Gyr (Chaboyer et al 1998). For a coeval population (e.g. open and globular clusters), isochrone fitting is widely used. The ages of the galactic halo and globular clusters, when averaged over eight independent surveys, lead to 12.2 ± 0.5 Gyr (Lineweaver 1999).

Parallaxes provide the stellar luminosity L after bolometric correction, and photometry gives the effective temperature T_{eff} . When the abundance $[\text{Fe}/\text{H}]$ is known, comparison of the location of the star in the $L - T_{\text{eff}}$ plane with theoretical isochrones can give an estimate of the age in regions where the L and T_{eff} depend on the age. For example, this method cannot be used on the unevolved main sequence, and is not ideal on the upper giant branch. The region near the main sequence is much used for isochrone ages, because these stars are common, but the isochrones in this region are complex and can give multi-valued age estimates. The subgiant branch is particularly well suited to isochrone ages. The Gaia mission will provide a huge increase in accurate parallaxes (errors of order 1% for bright stars with $V < 14$) and hence in precise age estimates.

If no trigonometric parallax is available, isochrones in the $\log-g - T_{\text{eff}}$ plane can be used to estimate ages, where $\log-g$ and T_{eff} can be derived from a combination of high resolution spectroscopy or low resolution spectrophotometry and broad-band photometry. Large biases can occur for isochrone ages if the errors in the observed quantities (e.g L and T_{eff}) are significant, because the underlying distributions of stellar mass and abundances are not uniform. Bayesian techniques can include these underlying distributions as priors (e.g. Pont & Eyer 2004).

For star clusters, the color-magnitude plane is often well populated, and isochrone fits to the main sequence, turnoff region and giant branch can be used to estimate the age and distance if the abundance $[\text{Fe}/\text{H}]$ is known. The recent ACS survey of Galactic globular clusters shows that old coeval clusters are present over the whole range of cluster metallicities, but that some of the outer clusters are younger and show an age-metallicity relation (Sarajedini et al 2007).

As Reid et al (2007) and Barnes (2007) showed, there is still much disagreement about estimating the ages of individual stars, because measuring accurate stellar ages is difficult. It remains a vital goal for unravelling the dynamical and chemical evolution of the Galaxy. The present estimate for the age of the old thin disk population when averaged

over five independent surveys is 8.7 ± 0.4 Gyr (Lineweaver 1999), although Oswalt et al argue for $9.5_{-0.8}^{+1.1}$ Gyr.

For a world model with ($\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$), the Big Bang occurred 13.7 Gyr ago – in our view, there is no compelling evidence for an age crisis from a comparison of estimates in the near and far field. But the inaccuracy of age dating relative to an absolute scale does cause problems. At present, the absolute ages of the oldest stars cannot be tied down to better than about 2 Gyr, the time elapsed between $z = 6$ and $z = 2$. This is a particular handicap to identifying specific events in the early Universe from the stellar record.

5.5. Goals of near field cosmology

We believe that the major goal of near field cosmology is to tag individual stars with elements of the protocloud. Some integrals of motion are likely to be preserved while others are scrambled by dissipation and violent relaxation. We suspect that complete tagging is impossible. However, some stars today may have some integrals of motion which relate to the protocloud at the epoch of last dissipation.

As we review, different parts of the Galaxy have experienced dissipation and phase mixing to varying degrees. The disk, in contrast to the stellar halo, is a highly dissipated structure. The bulge may be only partly dissipated. To what extent can we unravel the events that produced the Galaxy as we see it today? Could some of the residual inhomogeneities from prehistory have escaped the dissipative process at an early stage?

Far field cosmology currently takes us back to the ‘epoch of last scattering’ as seen in the microwave background. Cosmologists would like to think that some vestige of information has survived from earlier times (cf. Peebles, Seager & Hu 2000). In the same spirit, we can hope that fossils remain from the ‘epoch of last dissipation’, i.e. the main epoch of baryon dissipation that occurred as the disk was being assembled.

To make a comprehensive inventory of surviving inhomogeneities would require a vast catalog of stellar properties that is presently out of reach. The Gaia space astrometry mission, set to launch at the end of the decade, will acquire detailed phase space coordinates for about one billion stars, within a sphere of diameter 20 kpc (the Gaiasphere). In Sec. 8, we look forward to a time when all stars within the Gaiasphere have complete chemical abundance measurements (including all heavy metals). Even with such a vast increase in information, there may exist fundamental – but unproven – limits to unravelling the observed complexity.

The huge increase in data rates from ground-based and space-based observatories has led to an explosion of information. Much of this information from the near field is often dismissed as ‘weather’ or unimportant detail. But in fact fundamental clues are already beginning to emerge. In what is now a famous discovery, a large photometric and kinematic survey of bulge stars revealed the presence of the disrupting Sgr dwarf galaxy (Ibata et al 1994), now seen over a large region of sky and in a variety of populations (see Sec. 7.3.8). Perhaps the most important example arises from the chemical signatures seen in echelle spectroscopy of bulge, thick disk and halo stars. In Sec. 9, we envisage a time when the analysis of thousands of spectral lines for a vast number of stars will reveal crucial insights into the sequence of events early in the formation of the Galaxy.

In this review, we discuss fossil signatures in the Galaxy. A key aspect of fossil studies is a reliable time sequence. In Sec. 5.4, we discuss methods for age-dating individual stars and coeval groups of stars. In Sec. 6, we describe the main components of the Galaxy. In Sec. 7, we divide the fossil signatures of galaxy formation into three parts: zero order signatures that preserve information since dark matter virialized; first order signatures that preserve information since the main epoch of baryon dissipation; second order signatures that arise from major processes involved in subsequent evolution. In Sec. 8, we look forward to a time when it is possible to measure ages, phase space coordinates and chemical properties for a vast number of stars in the Galaxy. Even then, what are the prospects for unravelling the sequence of events that gave rise to the Milky Way? We conclude with some experimental challenges for the future.

6. STRUCTURE OF THE GALAXY

How would our Galaxy appear from the perspective of an external observer? This vexing question has a long history. Little progress has been made since de Vaucouleurs (1970b) classified the Milky Way as an SAB(rs)bc galaxy which recognizes it as weakly barred/ringed system. Several candidates have been put forward as surrogates of the Milky Way: NGC 891 (edge on) and NGC 3124, NGC 3992 or NGC 2336 (face on) are possible examples (see Fig. 18; Efremov 2011). There is ongoing debate about whether the Galaxy has 2 or 4 spiral arms, and whether these are asymmetric or not. Even the structure of the central bar is debated (Sec. 6.1).

Like most spiral galaxies, our Galaxy has several recognizable structural components that probably appeared at different stages in the galaxy formation process: a central bulge/bar, a thin disk, a thick disk and a tenuous stellar halo. These components will retain different kinds of signatures of their formation. We will describe these components in the context of other disk galaxies, and use images of other galaxies in Fig. 19 to illustrate the components.

6.1. The Bulge

First compare images of M104 and IC 5249 (Figs. 19(c,g)): these are extreme examples of galaxies with a large bulge and with no bulge. Large bulges like that of M104 are structurally and chemically rather similar to elliptical galaxies: their surface brightness distribution follows an $r^{1/4}$ law (e.g. Pritchet & van den Bergh 1994) and they show similar relations of [Fe/H] and [Mg/Fe] with absolute magnitude (e.g. Jablonka et al 1996). These properties lead to the view that the large bulges formed rapidly. The smaller bulges are often boxy in shape, with a more exponential surface brightness distribution (e.g. Courteau et al 1996). The current belief is that they may have arisen from the stellar disk through bending mode instabilities.

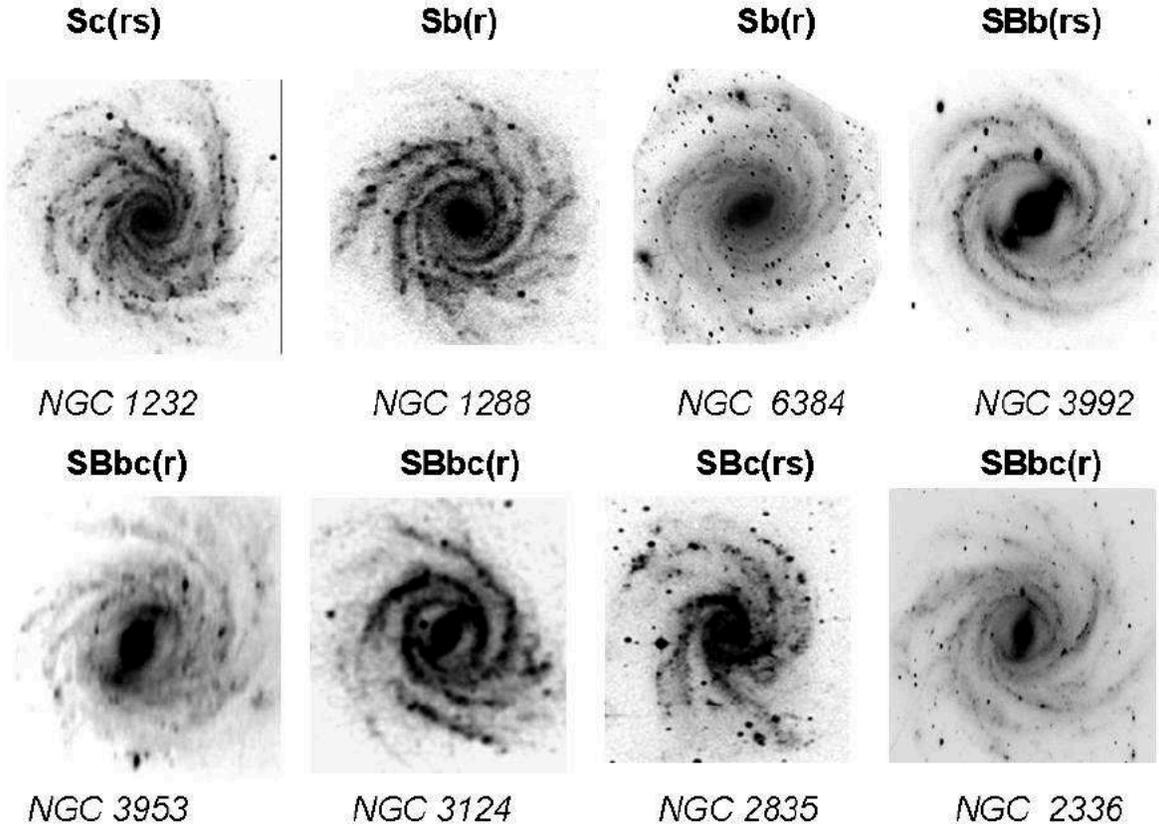


FIG. 18.— A selection of galaxies that bear resemblance to the Milky Way (Efremov 2011); NGC 3992 is arguably the closest match to the Galaxy.

Spiral bulges are usually assumed to be old but this is poorly known, even for the Galaxy. The presence of bulge RR Lyrae stars indicates that at least some fraction of the galactic bulge is old (Rich 2001). Furthermore, the color-magnitude diagrams for galactic bulge stars show that the bulge is predominantly old. McWilliam & Rich (1994) measured $[\text{Fe}/\text{H}]$ abundances for red giant stars in the bulge of the Milky Way. They found that, while there is a wide spread, the abundances ($[\text{Fe}/\text{H}] \approx -0.25$) are closer to the older stars of the metal rich disk than to the very old metal poor stars in the halo and in globular clusters, in agreement with the abundances of planetary nebulae in the Galactic bulge (e.g. Exter et al 2001).

The COBE image of the Milky Way (Fig. 19(b)) shows a modest somewhat boxy bulge, typical of an Sb to Sc spiral. Fig. 19(d) shows a more extreme example of a boxy/peanut bulge. If such bulges do arise via instabilities of the stellar disk, then much of the information that we seek about the state of the early galaxy would have been lost in the processes of disk formation and subsequent bulge formation. Although most of the more luminous disk galaxies have bulges, many of the fainter disk galaxies do not. Bulge formation is not an essential element of the formation processes of disk galaxies.

Galactic bulges show a wide range of morphology (e.g. Kormendy & Kennicutt 2004). At one extreme are the large classical bulges like the bulge of the Sombrero galaxy, which are believed to be generated by major mergers. They share some structural, dynamical and population properties with the lower-luminosity bright elliptical galaxies. At the other extreme are the pure disk galaxies which show no evidence for any kind of bulge. Smaller bulges like the Galactic bulge are often boxy or peanut in shape. They are believed to be associated with bar structures and to arise from the stellar disk through bar-forming and bar-buckling instabilities (Combes et al 1990; Athanassoula 2007).

There are also the pseudobulges (Kormendy 1993) which appear to be density enhancements which mimic the presence of a bulge in the radial surface brightness distribution, but lie entirely within the disk. These are currently believed to come from secular processes which transfer angular momentum and rearrange the surface density distribution within the disk. Forming large galaxies with small or no bulges is currently difficult within Λ CDM theory, because of the relatively active ongoing merger history in Λ CDM simulations.

In their recent survey of nearby galaxies, Kormendy et al (2010) find that essentially all of the disks are ‘pure’; while some have bulges, most of these bulges could have arisen from disk instabilities rather than infall. These authors state that “it is hard to understand how bulgeless galaxies could form as the quiescent tail of a distribution of merger histories. Recognition of pseudobulges makes the biggest problem with cold dark matter galaxy formation more acute: How can hierarchical clustering make so many giant, pure-disk galaxies with no evidence for merger-built bulges?”

Establishing the merger history of the Milky Way observationally is a major goal for Galactic archaeology, so we

need to understand how the Galactic bulge formed. Is it even partly a merger product, or did it form entirely through internal processes such as the disk and bar instabilities? The way in which the Galactic bulge rotates is a key indicator of its formation route.

The rotation field of small boxy bulges like the Galactic bulge are characteristically cylindrical (i.e. the rotational velocity is roughly constant with height above the galactic plane). This is seen in the dynamical simulations, observations of other boxy bulges (e.g. Falcon-Barroso et al 2006), and also in the Galactic bulge (brightest giants – Howard et al 2009; red clump giants – Ness et al 2012, 2013; Freeman et al 2013). This kind of observation is relatively straightforward for the bulges of other galaxies, via their integrated light. In our own Galaxy, the morphology of the rotation field is measured from large samples of individual bulge stars, and it is important to ensure that the sample is not contaminated by stars of the Galactic disk. The available evidence supports the view that the Galactic bulge formed primarily via disk/bar instabilities and not via mergers.

The stars of the Galactic bulge appear to be old and enhanced in alpha elements; it is not yet clear whether the bulge is more alpha enhanced than the local thick disk (e.g. Fulbright et al 2007, Melendez et al 2008). The significance of the alpha enrichment is that the chemical evolution of the stars in the bulge must have been very rapid, taking no more than about a Gyr. The bar-forming and bar-buckling process takes a few Gyr to act after the disk has settled. In the bar-buckling scenario, the bulge structure is then probably younger than the bulge stars themselves, which would have originally been part of the inner disk. The alpha enrichment of the bulge stars would come from the rapid star formation and chemical evolution which took place in the inner disk before the instability had time to act.

Λ CDM simulations make a prediction about the distribution of the first stars which form in the environment of a large galaxy like the Milky Way. These stars form in the more prominent overdensities of the mass distribution at redshifts > 12 (e.g. Moore et al 2006), long before the large galaxy itself becomes visible. As the galaxy comes together, the first stars are concentrated to the central regions of the galaxy, and inhabit the bulge region. They are not necessarily the most metal-poor stars in the galaxy, because of the chemical evolution which takes place in these overdensities before they become part of the larger system. The first stars are interesting archaeologically, but it will be a challenge to recognize them from among the stars of the inner halo which will also be concentrated towards the Galactic bulge.

There are repeated claims that the inner short, boxy bar is embedded within a long bar, where the major axes are misaligned with one another (Lopez-Corredoira et al 2005). This combination of short and long bars may be a common feature of many spirals (e.g. Efremov 2011). Evidence for the long bar has been claimed from infrared star counts (Benjamin et al 2005; Lopez-Corredoira et al 2007; Cabrera-Lavers et al 2008) and stellar velocities (Gardner & Flynn 2010). The dimensions of the long bar are $x : y : z \approx 8.0 : 1.2 : 0.2$ kpc and it is inclined at about 45° to our line of sight to the Galactic Centre (Lopez-Corredoira et al 2007). But an argument against the long bar is that there is no kinematic signature of it in the solar neighbourhood (Siebert et al 2011). It is possible that the infrared star counts are due to small spiral arms that extend from the ends of a bar aligned with the inner short bar (e.g. NGC 3992 in Fig. 18; Efremov 2011). Future kinematic surveys of the inner bulge should clarify this issue.

6.2. *The Disk*

Now look at the disks of these galaxies. The exponential thin disk, with a vertical scale height of about 300 pc, is the most conspicuous component in edge-on disk galaxies like NGC 4762 and IC 5249 (Figs. 19(e,g)). The thin disk is believed to be the end product of the quiescent dissipation of most of the baryons and contains almost all of the baryonic angular momentum. For the galactic disk, which is clearly seen in the COBE image in Fig. 19(b), we know from radioactive dating, white dwarf cooling and isochrone estimates for individual evolved stars and open clusters that the oldest disk stars have ages in the range 10 to 12 Gyr (see Sec. 5.4).

The disk is the defining stellar component of disk galaxies, and the universe (Driver et al 2007), and understanding its formation is in our view the most important goal of galaxy formation theory. Although much of the information about the pre-disk state of the baryons has been lost in the dissipative process, some tracers remain, and we will discuss them in the next section.

Many disk galaxies show a second fainter disk component with a larger scale height (typically about 1 kpc); this is known as the thick disk. Deep surface photometry of IC 5249 shows only a very faint thick disk enveloping the bright thin disk (Abe et al 1999): compare Figs. 19(g,h). In the edge-on S0 galaxy NGC 4762, we see a much brighter thick disk around its very bright thin disk (Tsikoudi 1980): the thick disk is easily seen by comparing Figs. 19(e,f). The Milky Way has a significant thick disk (Gilmore & Reid 1983): its scale height (~ 1 kpc) is about three times larger than the scale height of the thin disk, its surface brightness is about 10% of the thin disk's surface brightness, its stellar population appears to be older than about 12 Gyr, and its stars are significantly more metal poor than the stars of the thin disk. The galactic thick disk is currently believed to arise from heating of the early stellar disk by accretion events or minor mergers (see Sec. 7.2.3).

The thick disk may be one of the most significant components for studying signatures of galaxy formation because it presents a 'snap frozen' relic of the state of the (heated) early disk. Although some apparently pure-disk galaxies like IC 5249 do have faint thick disks, others do not (Fry et al 1999): these pure-disk galaxies show no visible components other than the thin disk. As for the bulge, formation of a thick disk is not an essential element of the galaxy formation process. In some galaxies the dissipative formation of the disk is clearly a very quiescent process. We direct the reader to a recent review of the Galactic disk by Rix & Bovy (2013).

6.3. *The Stellar Halo*

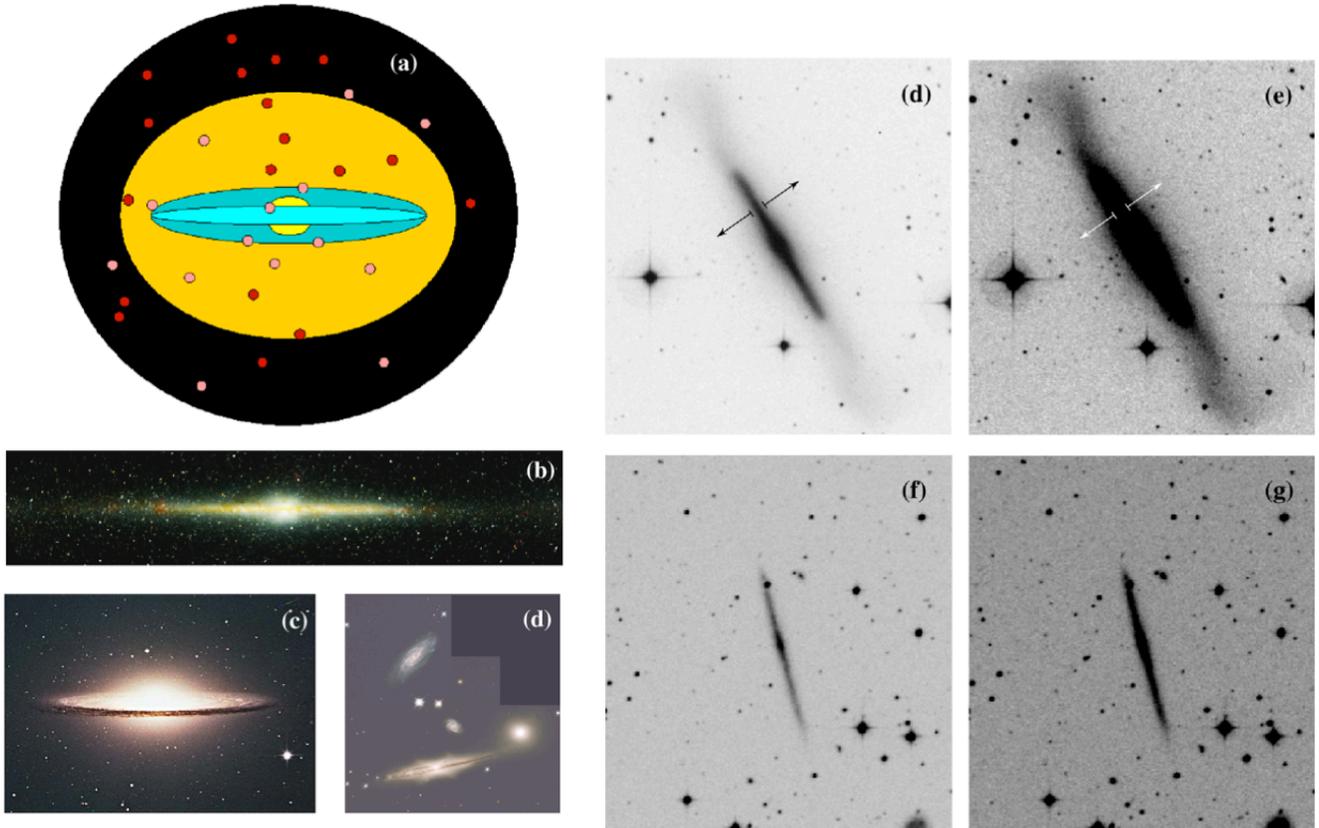


FIG. 19.— (a) Sketch of Milky Way showing the stellar disk (light blue), thick disk (dark blue), stellar bulge (yellow), stellar halo (mustard yellow), dark halo (black) and globular cluster system (filled circles). The radius of the stellar disk is roughly 15 kpc. The baryon and dark halos extend to a radius of at least 100 kpc. (b) Infrared image of the Milky Way taken by the DIRBE instrument on board the Cosmic Background Explorer (COBE) Satellite. [We acknowledge the NASA Goddard Space Flight Center and the COBE Science Working Group for this image.] (c) M104, a normal disk galaxy with a large stellar bulge (from AAO). (d) Hubble Heritage image of the compact group Hickson 87; one galaxy has a peanut-shaped stellar bulge due to dynamical interaction with other group members. (e) Image of the SO galaxy NGC 4762 (Digital Sky Survey) shows its thin disk and stellar bulge. The base of the arrows in (e) and (f) shows the height above the plane at which the thick disk becomes brighter than the thin disk (Tsikoudi 1980). (f) A deeper image of NGC 4762 (DSS) shows its more extended thick disk. (g) Image of the pure disk galaxy IC 5249 (DSS) shows its thin disk and no stellar bulge. (h) A deeper image of IC 5249 (DSS) shows no visible thick disk, although a very faint thick disk has been detected in deep surface photometry.

The Galactic halo (Fig. 19(a)) is a vast ancient repository that takes us back to a time when dark matter collapsed into bound structures and the Galaxy was seeded for the first time. Helmi (2008), Tolstoy, Hill & Tosi (2009) and Klement (2010) provide excellent overviews of the many stellar systems and fragments that inhabit the halo. These include field halo stars (Christlieb et al 2002; Cayrel et al 2004; Frebel et al 2005; Cohen et al 2007), globular clusters (Gratton, Sneden & Carretta 2004), dwarf spheroidals (Mateo 1999; Venn et al 2004), ultra-faint dwarf galaxies (Simon & Geha 2007; Kirby et al 2008), stellar streams (Ibata et al 1995; Chou et al 2010), stellar associations (Walsh et al 2007) and satellites to dwarf galaxies (Coleman et al 2004; Belokurov et al 2009).

A formal definition of the halo is made difficult by the presence of substructure, particularly if the entire halo is ultimately shown to be made up of distinct kinematic subsystems. But we do recognise a roughly smooth component from deep photometric surveys (e.g. Bell et al 2008). Its mass is only about 1% of the total stellar mass (about $10^9 M_{\odot}$: e.g. Morrison 1993). The surface brightness of the galactic halo, if observed in other galaxies, would be too low to detect from its diffuse light. It can be seen in other galaxies of the Local Group in which it is possible to detect the individual evolved halo stars.

The metal-poor halo of the Galaxy is very interesting for galaxy formation studies because it is so old and likely constitutes some of the first Galactic objects to form. The halo may have formed over a few billion years. Kalirai (2012) has recently determined an age 11.4 ± 0.7 Gyr from the white dwarf halo population in the solar neighbourhood. His sample is demonstrably younger than the oldest globular clusters in the halo.

The galactic halo has a power law density distribution $\rho \propto r^{-3.5}$ although this appears to depend on the stellar population (Vivas et al 2001, Chiba & Beers 2000). Unlike the disk and bulge, the angular momentum of the halo is close to zero (e.g. Freeman 1987), and it is supported almost entirely by its velocity dispersion; some of its stars are very energetic, reaching out to at least 100 kpc from the galactic center (e.g. Carney et al 1990).

The current view is that the galactic halo formed at least partly through the accretion of small metal-poor satellite galaxies which underwent some independent chemical evolution before being accreted by the Galaxy (Searle & Zinn 1978; Freeman 1987; Bell et al 2007). Although we do still see such accretion events taking place now, in the apparent

tidal disruption of the Sgr dwarf (Ibata et al 1995), most of them must have occurred long ago. Accretion of satellites would dynamically heat the thin disk, so the presence of a dominant thin disk in the Galaxy means that most of this halo-building accretion probably predated the epoch of thick disk formation ~ 12 Gyr ago. We can expect to see dynamically unmixed residues or fossils of at least some of these accretion events (e.g. Helmi & White 1999).

Of all the galactic components, the stellar halo offers the best opportunity for probing the details of its formation. There is a real possibility to identify groups of halo stars that originate from common progenitor satellites (Eggen 1977; Helmi & White 1999; Harding et al 2001; Majewski et al 2000). However, if the accretion picture is correct, then the halo is just the stellar debris of small objects accreted by the Galaxy early in its life. Although it may be possible to unravel this debris and associate individual halo stars with particular progenitors, this may tell us more about the early chemical evolution of dwarf galaxies than about the basic issues of galaxy formation. We would argue that the thin disk and thick disk of our Galaxy retain the most information about how the Galaxy formed. On the other hand, we note that current hierarchical Λ CDM simulations predict many more satellites than are currently observed. It would therefore be very interesting to determine the number of satellites that have already been accreted to form the galactic stellar halo.

We should keep in mind that the stellar halo accounts for only a tiny fraction of the galactic baryons and is dynamically distinct from the rest of the stellar baryons. There is mounting evidence that the stellar halo may harbour distinct abundance populations (Nissen & Schuster 2012; Schuster et al 2012) and/or comprise discrete rotating/counter-rotating components (Carollo et al 2007; Kafle et al 2012, 2013; cf. Schönrich et al 2012). We should note that the stellar halo of the Galaxy may not be typical: the halos of disk galaxies are quite diverse. The halo of M31, for example, follows the $r^{-1/4}$ law (Pritchett & van den Bergh 1994) and is much more metal rich in the mean than the halo of our Galaxy (Durrell et al 2001) although it does have stars that are very metal weak. It should probably be regarded more as the outer parts of a large bulge than as a distinct halo component.

6.4. The Dark Halo

The second inconspicuous component is the dark halo, which is detected only by its gravitational field. The dark halo contributes at least 90% of the total galactic mass and its $\rho \sim r^{-2}$ density distribution extends at least to 100 kpc (e.g. Kochanek 1996; Battaglia et al 2005). Beyond here, at some stage, the halo density must decline rapidly, as expected from theory and observed in numerical simulations. **The story of how dark matter in galaxies was established is given in Appendix A.**

Near the Sun, the density of the dark halo is about $0.008 M_{\odot} \text{ pc}^{-3}$ compared to the density of luminous matter of about $0.01 M_{\odot} \text{ pc}^{-3}$. Its contribution to the local surface density is more significant. The total surface density of matter near the sun, within $|z| = 1.1$ kpc, is about $71 M_{\odot} \text{ pc}^{-2}$, and the surface density of luminous matter alone is about $48 M_{\odot} \text{ pc}^{-2}$: near the Sun, the dark matter contributes about a third of the Galactic surface density (Kuijken & Gilmore 1991).

The Galactic dark halo is believed to be spheroidal rather than disk-like (Crézé et al 1998; Ibata et al 2001b; see Pfenniger et al 1994 for a contrary view). The tidal streams of the Sgr dwarf galaxy suggest that the halo is triaxial (Law et al 2009), with axial ratios of about $b/a = 0.83$ and $c/a = 0.67$. The minor axis appears to lie approximately along the Sun - Galactic Centre line and the major axis along the direction of Galactic rotation. The contribution of the dark matter in the inner Galaxy can be estimated using models of the Galactic mass distribution, based on the COBE near-infrared light distribution and normalised by HST star counts and by the kinematics of gas in the inner Galaxy. These models consistently give a good fit to the observed optical stellar microlensing depth as measured by the large microlensing surveys like MACHO, EROS and OGLE, which implies that the gravitational field in the inner few kpc of the Galaxy is dominated by the baryonic component (e.g. Englmaier & Gerhard 2006).

It is expected that at least some satellite galaxies accreted by larger disk galaxies would be pulled down into the disk by dynamical friction before tidal disruption (e.g. Quinn & Goodman 1986; Walker et al 1996; Abadi et al 2003). These systems would deposit their dark matter and their stars into the disk of the large galaxy. The presence of such disk dark matter would be difficult to detect if its scale height is a few kpc, but its existence would be interesting in the context of the enhancement to direct detection of dark matter (e.g. Purcell et al 2009). The vast samples of precise distances and transverse velocities expected from the Gaia mission offer prospects for independent estimates of total local densities and surface densities from the kinematics of disk stars in the inner disk of the galaxy.

We should emphasise here the small fraction of the Galactic mass which is in the form of stellar baryons. The mass of the dark halo, from a range of estimators, is about $1 - 1.5 \times 10^{12} M_{\odot}$ (Smith et al 2007). The high value is preferred by the orbit of the Magellanic System (Nichols et al 2012) whereas the low value is derived from halo stars (e.g. Kafle et al 2012). With less uncertainty, the stellar mass of the Galaxy is about $6 \times 10^{10} M_{\odot}$ (see below). The stellar baryons provide only about 3% of the total mass which agrees well with estimates from weak lensing for the stellar fraction of L_{*} galaxies (e.g. Mandelbaum et al 2006). This 3% fraction is much less than the universal ratio of $f_b^o \approx 17\%$.

In the current picture of galaxy formation, the dark halo plays a very significant role. The disk is believed to form dissipatively within the potential of the virialized spheroidal halo which itself formed through the fairly rapid aggregation of smaller bodies (e.g. White & Rees 1978). Λ CDM simulations suggest that the halo may still be strongly substructured (see Sec. 7.1.5). If this is correct, then the lumpy halo would continue to influence the evolution of the galactic disk, and the residual substructure of the halo is a fossil of its formation. It remains a challenge to find direct observational evidence of this dark substructure.

Within the limitations mentioned above, each of these distinct components of the Galaxy preserves signatures of its

past and so gives insights into the galaxy formation process. We now discuss these signatures.

7. SIGNATURES OF GALAXY FORMATION

Our framework is that the Galaxy formed through hierarchical aggregation. We identify three major epochs in which information about the proto-hierarchy is lost:

1. the dark matter virializes – this could be a time of intense star formation but need not be, as evidenced by the existence of very thin pure disk galaxies;
2. the baryons dissipate to form the disk and bulge;
3. an ongoing epoch of formation of objects within the disk and accretion of objects from the environment of the galaxy, both leaving some long-lived relic.

We classify signatures relative to these three epochs. The role of the environment is presently difficult to categorize in this way. Environmental influences must be operating across all of our signature classes.

7.1. Zero order signatures – information preserved since dark matter virialized

7.1.1. Introduction

During the virialization phase, a lot of information about the local hierarchy is lost: this era is dominated by merging and violent relaxation. The total dark and baryon mass are probably roughly conserved, as is the angular momentum of the region of the hierarchy that went into the halo. The typical density of the environment is also roughly conserved: although the structure has evolved through merging and relaxation, a low density environment remains a low density environment (see White 1996 for an overview).

7.1.2. Signatures of the environment

The local density of galaxies (and particularly the number of small satellite systems present at this epoch) affects the incidence of later interactions. For the Local Group, the satellite numbers appear to be lower than expected from Λ CDM (Moore et al 1999; Klypin et al 1999). However there is plenty of evidence for past and ongoing accretion of small objects by the Milky Way and M31 (Ibata et al 1995; Ibata et al 2001b).

The thin disk component of disk galaxies settles dissipatively in the potential of the virialised dark halo (e.g. Fall & Efstathiou 1980). The present morphology of the thin disk depends on the numbers of small galaxies available to be accreted: a very thin disk is an indication of few accretion events (dark or luminous) after the epoch of disk dissipation and star formation (e.g. Freeman 1987; Quinn et al 1993; Walker et al 1996). The formation of the thick disk is believed to be associated with a discrete event that occurred very soon after the disk began to settle, at a time when about 10% of the stars of the disk had already formed. In a low density environment, without such events, thick disk formation may not occur. Since the time of thick disk formation, the disk of the Galaxy appears to have been relatively undisturbed by accretion events. This is consistent with the observation that less than 10% of the metal-poor halo comes from recent accretion of star forming satellites (Unavane et al 1996).

The existence and structure of the metal-poor stellar halo of the Galaxy may depend on accretion of small objects. This accretion probably took place after the gaseous disk had more or less settled – the disk acts as a resonator for the orbit decay of the small objects. So again the environment of our proto-galaxy may have a strong signature in the very existence of the stellar halo, and certainly in its observed substructure. We would not expect to find a stellar halo encompassing pure disk galaxies, consistent with the limited evidence now available (Freeman et al 1983; Schommer et al 1992).

7.1.3. Signatures of global quantities

During the process of galaxy formation, some baryons are lost to ram pressure stripping and galactic winds. Most of the remaining baryons become the luminous components of the galaxy. The total angular momentum J of the dark halo may contribute to its shape, which in turn may affect the structure of the disks. For example, warps may be associated with misalignment of the angular momentum of the dark and baryonic components. The dark halo may have a rotating triaxial figure: the effect of a rotating triaxial dark halo on the self-gravitating disk has not yet been seriously investigated (see Bureau et al 1999).

The binding energy E at the epoch of halo virialization affects the depth of the potential well and hence the characteristic velocities in the galaxy. It also affects the parameter $\lambda = J|E|^{1/2}G^{-1}M^{-5/2}$, where M is the total mass: λ which critical for determining the gross nature of the galactic disk as a high or low surface brightness system (e.g. Dalcanton et al 1997).

The relation between the specific angular momentum J/M and the total mass M (Fall 1983) of disk galaxies is well reproduced by simulations (Zurek et al 1988). Until recently, ellipticals and disk galaxies appeared to be segregated in the Fall diagram: from the slow rotation of their inner regions, estimates of the J/M ratios for ellipticals were about 1 dex below those for the spirals. More recent work (e.g. Arnaboldi et al 1994) shows that much of the angular momentum in ellipticals appears to reside in their outer regions, so ellipticals and spirals do have similar locations in the Fall diagram. Internal redistribution of angular momentum has clearly occurred in the ellipticals (Quinn & Zurek 1988).

The remarkable Tully-Fisher law (1977) is a correlation between the HI line-width and the optical luminosity of disk galaxies. It appears to relate the depth of the potential well and the baryonic mass (McGaugh et al 2000). Both of these quantities are probably roughly conserved after the halo virialises, so the Tully-Fisher law should be regarded as a zero-order signature of galaxy formation. The likely connecting links between the (dark) potential well and the baryonic mass are (i) a similar baryon/dark matter ratio from galaxy to galaxy, and (ii) an observed Faber-Jackson law for the dark halos, of the form $M \propto \sigma^4$: i.e. surface density independent of mass for the dark halos.

7.1.4. Signatures of the internal distribution of specific angular momentum

The internal distribution of specific angular momentum $j(h)$ of the baryons (i.e. the mass with specific angular momentum $< h$) largely determines the shape of the surface brightness distribution of the disk rotating in the potential well of the dark matter. Together with $j(h)$, the total angular momentum and mass of the baryons determine the scale length and scale surface density of the disk. Therefore the distribution of total angular momentum and mass for protodisk galaxies determines the observed distribution of the scale length and scale surface density for disk galaxies (Freeman 1970; de Jong & Lacey 2000).

Many studies have assumed that $j(h)$ is conserved through the galaxy formation process, most notably Fall & Efstathiou (1980). It is not yet clear if this assumption is correct. Conservation of the internal distribution of specific angular momentum $j(h)$ is a much stronger requirement than the conservation of the *total* specific angular momentum J/M . Many processes can cause the internal angular momentum to be redistributed, while leaving the J/M ratio unchanged. Examples include the effects of bars, spiral structure (Lynden-Bell & Kalnajs 1972) and internal viscosity (Lin & Pringle 1987).

The maximum specific angular momentum h_{\max} of the baryons may be associated with the truncation of the optical disk observed at about four scale lengths (de Grijs et al 2001; Pohlen et al 2000). This needs more investigation. The truncation of disks could be an important signature of the angular momentum properties of the early protocloud, but it may have more to do with the critical density for star formation or the dynamical evolution of the disk. Similarly, in galaxies with very extended HI, the edge of the HI distribution may give some measure of h_{\max} in the protocloud. On the other hand, it may be that the outer HI was accreted subsequent to the formation of the stellar disk (van der Kruit 2002), or that the HI edge may just represent the transition to an ionized disk (Maloney 1993; Bland-Hawthorn et al 1997).

This last item emphasizes the importance of understanding what is going on in the outer disk. The outer disk offers some potentially important diagnostics of the properties of the protogalaxy. At present there are too many uncertainties about significance of (a) the various cutoffs in the light and HI distributions, (b) the age gradient seen by Bell & de Jong (2000) from integrated light of disks but not by Friel (1995) for open clusters in the disk of the Galaxy, and (c) the outermost disk being maybe younger but not ‘zero age’, which means that there is no real evidence that the disk is continuing to grow radially. It is possible that the edge of the disk has something to do with angular momentum of baryons in the protocloud or with disk formation process, so it may be a useful zero-order or first-order signature.

7.1.5. Signatures of the Λ CDM hierarchy

Λ CDM predicts a high level of substructure which is in apparent conflict with observation. Within galaxies, the early N-body simulations appeared to show that substructure with characteristic velocities in the range $10 < V_c < 30$ km s⁻¹ would be destroyed by merging and virialization of low mass structures (Peebles 1970; White 1976; White & Rees 1978). It turned out that the lack of substructure was an artefact of the inadequate spatial and mass resolution (Moore et al 1996). Current simulations reveal 500 or more low mass structures within 300 kpc of an L_* galaxy’s sphere of influence (Moore et al 1999; Klypin et al 1999). This is an order of magnitude larger than the number of low mass satellites in the Local Group. Mateo (1998) catalogues about 40 such objects and suggests that, at most, we are missing a further 15 – 20 satellites at low galactic latitude. Kauffmann et al (1993) were the first to point out the ‘satellite problem’ and suggested that the efficiency of dynamical friction might be higher than usually quoted. However, without recourse to fine tuning, this would remove essentially all of the observed satellites in the Local Group today.

Since the emergence of the Λ CDM paradigm, an inevitable question is whether a ‘basic building block’ can be recognized in the near field. Moore et al emphasize the self-similar nature of Λ CDM sub-clustering and point to the evidence provided by the mass spectrum of objects in rich clusters, independent of the N-body simulations. The lure of finding a primordial building block in the near field has prompted a number of tests. If the dark mini-halos comprise discrete sources, it should be possible to detect microlensing towards a background galaxy (see Sec. 6.4).

The satellite problem appears to be a fundamental prediction of Λ CDM in the non-linear regime. Alternative cosmologies have been suggested involving the reduction of small-scale power in the initial mass power spectrum (Kamionkowski & Liddle 2000), warm dark matter (Hogan & Dalcanton 2001; White & Croft 2000; Colin et al 2000), or strongly self-interacting dark matter (Spergel & Steinhardt 2000). Several authors have pointed out that some of the direct dark-matter detection experiments are sensitive to the details of the dark matter in the solar neighbourhood. Helmi et al (2001) (cf. Moore et al 2001) estimate that there may be several hundred kinematically cold ‘dark streams’ passing through the solar neighbourhood.

If Λ CDM is correct in detail, then we have simply failed to detect or to recognize the many hundreds of missing objects throughout the Local Group. For example, the satellites may be dark simply because baryons were removed long ago

through supernova-driven winds (Dekel & Silk 1986; Mac Low & Ferrara 1999). In support of this idea, x-ray halos of groups and clusters are almost always substantially enriched in metals ($[\text{Fe}/\text{H}] \geq -0.5$; Renzini 2000; Mushotzky 1999). In fact, we note that up to 70% of the mass fraction in metals is likely to reside in the hot intracluster and intragroup medium (Renzini 2000). Another explanation may be that the absence of baryons in hundreds of dark satellites was set in place long ago during the reionization epoch. Many authors note that the accretion of gas on to low-mass halos and subsequent star formation is heavily suppressed in the presence of a strong photoionizing background (Ikeuchi 1986; Rees 1986; Babul & Rees 1992; Bullock et al 2000). This effect appears to have a cut-off at low galactic mass at a characteristic circular velocity close to 30 km s^{-1} (Thoul & Weinberg 1996; Quinn et al 1996), such that the small number of visible Local Group dwarfs are those which exceed this cutoff or acquired most of their neutral hydrogen before the reionization epoch.

Blitz et al (1999) suggested that the high-velocity HI gas cloud (HVC) population is associated with dark mini-halos on megaparsec scales within the Local Group. This model was refined by Braun & Burton (1999) to include only the compact HVCs. The HVCs have long been the subject of wide-ranging speculation. Oort (1966) realized that distances derived from the virial theorem and the HI flux would place many clouds at megaparsec distances if they are self-gravitating. If the clouds lie at about a megaparsec and are associated with dark matter clumps (Nichols & Bland-Hawthorn 2009), then they could represent the primordial building blocks. However H α distances (Bland-Hawthorn et al 1998) suggest that most HVCs lie within 50 kpc and are unlikely to be associated with dark matter halos (Putman et al 2003). We note that several teams have searched for but failed to detect a faint stellar population in HVCs.

Moore et al (1999; see also Bland-Hawthorn & Freeman 2000) suggested that ultrathin disks in spirals are a challenge to the Λ CDM picture in that disks are easily heated by orbiting masses. However, Font et al (2001) find that in their Λ CDM simulations, very few of the Λ CDM sub-halos come close to the optical disk which may allow a very cold disk to form. At present there are real problems in reconciling the predictions of Λ CDM simulations with observations on scales of the Local Group (e.g. Kormendy et al 2010).

7.2. First order signatures – information preserved since the main epoch of baryon dissipation

7.2.1. The structure of the disk

At what stage in the evolution of the disk are its global properties defined? In part, we have already discussed this question in Sec. 7.1.4. The answer depends on how the internal angular momentum distribution $j(h)$ has evolved as the disk dissipated and various non-axisymmetric features like bars and formation and spiral structure came and went. Viscous processes associated with star formation, as suggested by Lin & Pringle (1987), may also contribute to the evolution of the $j(h)$ distribution.

The global structure of disks is defined by the central surface brightness I_o and the radial scalelength h of the disk. de Jong & Lacey (2000) evaluated the present distribution of galaxies in the (I_o, h) plane. If $j(h)$ has indeed remained roughly constant, as is often assumed for discussions of disk formation (e.g. Fall & Efstathiou 1980; Fall 1983), then the global parameters of the disk – the scale length, central surface brightness and the Tully-Fisher relation – are relics of the main epoch of baryon dissipation.

The edges of disks may hold clues to important physical processes. Until recently, it was believed that all galaxy disks undergo truncation near the Holmberg radius (26.5 mag arcsec $^{-2}$) and that this was telling us something important about the collapsing protocloud of gas that formed the early disk (van der Kruit 1979). This picture was challenged by our earlier finding (Bland-Hawthorn et al 2005) of a classic exponential disk (Freeman 1970) with no break in NGC 300, out to the distances corresponding to 10 disk scale lengths in this low-luminosity late-type spiral (see also Irwin et al 2005). The study of NGC 300 is particularly notable because the surface brightness profile was traced with resolved stars down to an effective surface brightness well below the levels achieved with surface photometry.

A recent study of the SDSS archives suggests that there are three basic classes of surface brightness profiles in galaxies (Pohlen et al 2008). Truncations in surface brightness profiles, characterized by the smooth break between the inner shallower and the outer steeper exponential, have been known for more than two decades and detected in a large number of galaxies (van der Kruit 1979; de Grijs et al 2001; Pohlen et al 2002). Despite this, the origin of the break remains unclear.

The currently favoured model for the explanation of these sub-exponential truncations is the star formation threshold scenario (Kennicutt 1989; Schaye 2004; Roskar et al 2008a,b; but see van der Kruit 2007). The threshold gas densities predicted in these models ($3 - 10 M_{\odot} \text{ pc}^{-2}$; Schaye 2004) broadly agree with the observed surface brightnesses at the radius of the break which correspond to stellar mass densities of $10 - 20 M_{\odot} \text{ pc}^{-2}$ (Bakos et al 2008). The second class of surface brightness profiles, so-called upbending or super-exponential breaks, may be due to minor mergers, supported both observationally (Ibata et al 2005) and by N-body/smoothed particle hydrodynamics simulations (Younger et al 2007). In the next section, we show how disk edges are being exploited to test new ideas on disk evolution.

7.2.2. Can disks preserve fossil information?

Here, we consider radial and vertical fossil gradients in the disk, in particular of abundance and age. Our expectation is that much of the information will be diluted through the dynamical evolution and radial mixing of the disk.

For spirals, different mechanisms may be at work to establish gradients (Molla et al 1996): (a) a radial variation of the yield due either to the stellar metal production or to the initial mass function, (b) a radial variation of the timescale for star formation, (c) a radial variation for the timescale of infall of gas from outside the disk. Once gradients are established, these can be amplified or washed out by radial mixing (Edmunds 1990; Goetz & Koeppen 1992).

Most stars are born in large clusters numbering hundreds or even thousands of stars. Some clusters stay together for billions of years, whereas others become unbound shortly after the initial starburst, depending on the star formation efficiency. When a cloud disperses, each star suffers a random kick superimposed on the cloud's mean motion. Thereafter, stars are scattered by transient spiral arm perturbations and star-cloud encounters.

These perturbations allow the star to migrate in integral space. During interaction with a single spiral event of pattern speed Ω_p , a star's energy and angular momentum change while it conserves its Jacobi integral: in the (E, J) plane, stars move along lines of constant $I_J = E - \Omega_p J$. The star undertakes a random walk in the (E, J) plane, perturbed by a series of spiral arm events (Sellwood 1999; Dehnen 2000). N-body models of disk evolution indicate that radial mixing is strong (Sellwood & Binney 2002; Lynden-Bell & Kalnajs 1972). This is believed to be driven by *transient* spiral waves that heat the in-plane motions, although the process is not yet well understood. Long-term spiral arms produce no net effect. Remarkably, a *single* spiral wave near co-rotation can perturb the angular momentum of a star by $\sim 20\%$ without significant heating: the star is simply moved from one circular orbit to another, inwards or outwards, by ~ 2 kpc (Sellwood & Binney 2002). Substantial variations in the angular momentum of a star are possible over its lifetime.

In addition to radial heating, stars experience vertical disk heating: their vertical velocity dispersion increases as they age. This is believed to occur through a combination of in-plane spiral-arm heating and scattering off giant molecular clouds (e.g. Spitzer & Schwarzschild 1953; Carlberg & Sellwood 1985). The in-plane heating is most effective at the inner and outer Lindblad resonances and vanishes at corotation. In the vertical direction, an 'age-velocity dispersion' relation is observed for stars younger than about 3 Gyr, but older disk stars show a velocity dispersion that is independent of age. Thus, the vertical structure does depend on the mean age of the population for $\tau < 3$ Gyr (Edvardsson et al 1993, confirmed from Hipparcos data by Gomez et al 1997).

As the amplitude of the random motions increases, the star becomes less vulnerable to heating by transient spiral waves, and the heating process is expected to saturate. This probably happens after about 3 Gyr (Binney & Lacey 1988; Jenkins & Binney 1990), consistent with observation. This is important for our purpose here. It means that dynamical information is preserved about the state of the thin disk at an early epoch, or roughly $\tau_L - 3 \approx 7$ Gyr ago, for which τ_L is the look-back time when the disk first began to form.

The survival of old open clusters like NGC 6791, Berkeley 21 and Berkeley 17 (Friel 1995; van den Bergh 2000) is of interest here. The oldest open clusters exceed 10 Gyr in age and constitute important fossils (Phelps & Janes 1996). Both old and young open clusters are part of the thin disk. If the heating perturbations occur over a lengthscale that significantly exceeds the size of an open cluster, it seems likely that the cluster will survive. A large spiral-arm heating event will heat many stars along their I_J trajectories. The trace of the heating event is likely to survive for a very long time but be visible only in integral space (Sellwood 2001a). We note that vertical abundance gradients have not been seen among the open clusters (Friel & Janes 1993).

About 4% of disk stars are super metal-rich (SMR) relative to the Hyades (Castro et al 1997). SMR stars of intermediate age appear to have formed a few kpc inside of the Solar circle from enriched gas. The oldest SMR stars appear to come from the Galactic Center: their peculiar kinematics and outward migration may be associated with the central bar (Carraro et al 1998; Grenon 1999).

In summary, our expectation is that fossil gradients within the disk are likely to be weak. This is borne out by observations of both the stars and the gas (Chiappini et al 2001). The vertical structure of the disk preserves another fossil – the 'thick disk' – which we discuss in the next section. Like the open clusters, this component also does not show a vertical abundance gradient (Gilmore et al 1995). In later sections, we argue that this may be the most important fossil to have survived the early stages of galaxy formation.

7.2.3. Disk heating by accretion: the thick disk

Heating from discrete accretion events also imposes vertical structure on the disk (Quinn & Goodman 1986; Walker et al 1996). Such events can radically alter the structure of the inner disk and the bulge (see Fig. 19(d) for an example), and are currently believed to have generated the thick disk of the Galaxy.

The galactic thick disk was first recognized by Gilmore & Reid (1983). It includes stars with a wide range of metallicity, from $-2.2 \leq [\text{Fe}/\text{H}] \leq -0.5$ (Chiba & Beers 2000): most of the thick disk stars are in the more metal-rich end of this range. The velocity ellipsoid of the thick disk is observed to be $(\sigma_R, \sigma_\phi, \sigma_z) = (46 \pm 4, 50 \pm 4, 35 \pm 3)$ km s⁻¹ near the sun, with an asymmetric drift of about 30 km s⁻¹. For comparison, the nearby halo has a velocity ellipsoid $(\sigma_R, \sigma_\phi, \sigma_z) = (141 \pm 11, 106 \pm 9, 94 \pm 8)$ km s⁻¹ and its asymmetric drift is about 200 km s⁻¹.

The mean age of the thick disk is not known. From photometric age-dating of individual stars, the thick disk appears to be as old as the globular clusters. Indeed, the globular cluster 47 Tuc (age 12.5 ± 1.5 Gyr: Liu & Chaboyer 2000) is often associated with the thick disk.

After Quinn & Goodman (1986), Walker et al (1996) showed in detail that a low mass satellite could substantially heat the disk as it sinks rapidly within the potential well of a galaxy with a 'live' halo. The conversion of satellite orbital energy to disk thermal energy is achieved through resonant scattering. Simulations of satellite accretion are important for understanding the survival of the thin disk and the origin of the thick disk. This is particularly relevant within the context of Λ CDM. The satellites which do the damage are those that are dense enough to survive tidal disruption by the Galaxy. We note that even dwarf spheroidals which appear fluffy are in fact rather dense objects dominated by their dark matter.

It is fortuitous that the Galaxy has a thick disk, since this is not a generic phenomenon. The disk structure may

be vertically stepped as a consequence of past discrete accretion events. The Edvardsson et al (1993) data appears to show an abrupt increase in the vertical component of the stellar velocity dispersion at an age of 10 Gyr: see also Strömberg (1987). Freeman (1991) argued that the age–(velocity dispersion) relation shows three regimes: stars younger than 3 Gyr with $\sigma_z \sim 10 \text{ km s}^{-1}$, stars between 3 and 10 Gyr with $\sigma_z \sim 20 \text{ km s}^{-1}$, and stars older than 10 Gyr with $\sigma_z \sim 40 \text{ km s}^{-1}$. The first regime probably arises from the disk heating process due to transient spiral arms which we described in the previous section. The last regime is the thick disk, presumably excited by an ancient discrete event.

Can we still identify the disrupting event that led to the thick disk? There is increasing evidence now that the globular cluster ω Cen is the stripped core of a dwarf elliptical (see Sec. 7.3.6). It is possible that the associated accretion event or an event like it was the event that triggered the thick disk to form.

In summary, it seems likely that the thick disk may provide a snap-frozen view of conditions in the disk shortly after the main epoch of dissipation. Any low level chemical or age gradients would be of great interest in the context of dissipation models. In this regard, Hartkopf & Yoss (1982) argued for the presence of a vertical abundance gradient in the thick disk, although Gilmore, Wyse & Jones (1995) find no such effect. Because stars of the thick disk spend relatively little time near the galactic plane, where the spiral arm heating and scattering by giant molecular clouds is most vigorous, radial mixing within the thick disk is unlikely to remove all vestiges of a gradient. If our earlier suggestions are right (§ 7.1.4), we might expect to see a different truncation radius for the thick disk compared to the thin disk.

7.2.4. *Is there an age–metallicity relation?*

Some fossil information has likely been preserved since the main epoch of baryon dissipation. The inner stellar bulge is a striking example. It is characterized by old, metal-rich stars which seems to be at odds with the classical picture where metals accumulate with time (Tinsley 1980). However, the dynamical timescales in the inner bulge are very short compared to the outer disk, and would have allowed for rapid enrichment at early times. This is consistent with the frequent occurrence of metal-rich cores of galaxies observed at high redshift (Hamann & Ferland 1999). The dynamical complexity of the Galactic bulge may not allow us to determine the sequence of events that gave rise to it. We anticipate that this will come about from far-field cosmology (Robertson et al 2010).

The existence of an age–metallicity relation (AMR; Fig. 16) in stars is a very important issue, about which there has long been disagreement. Twarog (1980) and Meusinger et al (1991) provide evidence for the presence of an AMR, while Carlberg et al (1985) find that the metallicity of nearby F stars is approximately constant for stars older than about 4 Gyr. More recently it has become clear that an AMR is apparent only in the solar neighbourhood and is strictly true only for stars younger than 2 Gyr and hotter than $\log T_{\text{eff}} = 3.8$ (Feltzing et al 2001). Edvardsson et al (1993) demonstrate that there is no such relation for field stars in the old disk. Similarly, Friel (1995) shows that there is no AMR for open clusters (see Sec. 7.3.5): she goes on to note that

Apparently, over the entire age of the disk, at any position in the disk, the oldest clusters form with compositions as enriched as those of much younger objects.

In fact, it has been recognized for a long time (e.g. Arp 1962; Eggen & Sandage 1969; Hirshfeld et al 1978) that old, metal-rich stars permeate the galaxy, throughout the disk, the bulge and the halo. We regard the presence of old metal-rich stars as a first-order signature. An age–metallicity relation which applies to all stars would have been an important second-order signature, but we see no evidence for such a relation, except among the young stars.

7.2.5. *Effects of environment and internal evolution*

Environmental influences are operating on all scales of the hierarchy and across all stages of our signature classification, so our attempts to classify signatures are partly artificial. Within Λ CDM, environmental effects persist throughout the life of the galaxy.

The parameters that govern the evolution of galaxies are among the key unknowns of modern astrophysics. Are the dominant influences internal (e.g. depth of potential) or external (e.g. environment) to galaxies? We consider here the effects of environment and internal evolution on the validity of the first-order signatures of galaxy formation (i.e. the properties that may have been conserved since the main epoch of baryon dissipation).

The well-known ‘G dwarf problem’ indicates that external influences are important. A simple closed box model of chemical evolution predicts far too many metal-poor stars in the solar neighbourhood (Tinsley 1980). This problem is easily remedied by allowing gas to flow into the region (Lacey & Fall 1983, 1985; Clayton 1987, 1988; Wyse & Silk 1989; Matteucci & Francois 1989; Worthey et al 1996). In the context of Λ CDM, this is believed to arise from the continued accretion of gas-rich dwarfs (e.g. Cole et al 1994; Kauffmann & Charlot 1998).

Environment is clearly a key factor. Early type galaxies are highly clustered compared to late type galaxies (Hubble & Humason 1931; Dressler 1980). Trager et al (2000) find that for a sample of early-type galaxies in low-density environments, there is a large spread in the $H\beta$ index (i.e. age), but little variation in metallicity. For galaxies in the Fornax cluster, Kuntschner (2000) finds the opposite effect: a large spread in metallicity is present with little variation in age. This probably reflects strong differences in environment between the field and the cluster.

Another likely environmental effect is the fraction of S0 galaxies in clusters, which shows a rising trend with redshift since $z \approx 0.4$ (Jones et al 2000). Furthermore, S0 galaxies in the Ursa Major cluster show age gradients that are inverted compared to field spirals, in the sense that the cores are young and metal-rich (Tully et al 1996; Kuntschner

& Davies 1998). Both of these effects involve more recent phenomena and would be properly classified as second-order signatures.

Internal influences are also at work. A manifestation is the color-magnitude relation (CMR) in early type (Sandage & Visvanathan 1978) and late type (Peletier & de Grijs 1998) galaxies. The CMR does not arise from dust effects (Bell & de Jong 2000) and must reflect systematic variations in age and/or metallicity with luminosity. In the case of ellipticals, the CMR is believed to reflect a mass-metallicity dependence (Faber 1973; Bower et al 1998). The relation is naturally explained by supernova-driven wind models in which more massive galaxies retain supernova ejecta and thus become more metal rich and redder (Larson 1974; Arimoto & Yoshii 1987). The CMR is presumably established during the main phase of baryon dissipation and is a genuine first-order signature.

Concannon et al (2000) analyzed a sample of 100 early type galaxies over a large range in mass. They find that lower mass galaxies exhibit a larger range in age than higher mass galaxies. This appears to show that smaller galaxies have had a more varied star formation history, which is at odds with the naive Λ CDM picture of low-mass galaxies being older than high mass galaxies (Baugh et al 1996; Kauffman 1996). The work of Concannon et al (2000) shows the presence of a real cosmic scatter in the star formation history. It is tempting to suggest that this cosmic scatter relates to different stages of evolution within the hierarchy. In this sense, we would regard the Concannon et al result as a first-order manifestation of galaxy formation (see Sec. 5.3).

Spiral galaxies commonly show color gradients that presumably reflect gradients in age and metallicity (Peletier & de Grijs 1998). Faint spiral galaxies have younger ages and lower metallicities relative to bright spirals. In a study of 120 low-inclination spirals, Bell & de Jong (2000) found that the local surface density within galaxies is the most important parameter in shaping their star formation and chemical history. However, they find that metal-rich galaxies occur over the full range of surface density. This fact has a remarkable resonance with the distribution of the metal rich open clusters that are found at any position in the Galactic disk (see Sec. 7.2.4). Bell & de Jong argue that the total mass is a secondary factor that modulates the star formation history. Once again, these authors demonstrate the existence of cosmic scatter that may well arise from variations in environment.

7.3. *Second order signatures – major processes involved in subsequent evolution*

7.3.1. *Introduction*

Here we consider relics of processes that have taken place in the Galaxy since most of the baryonic mass settled to the disk. There are several manifestations of these processes, probably the most significant of which is the star formation history of the disk, for which the open clusters are particularly important probes. There is a wealth of detail relating to anomalous populations throughout the Galaxy, discussed at length by Majewski (1993). Examples include an excess of stars on extreme retrograde orbits (Norris & Ryan 1989; Carney et al 1996), metal-poor halo stars of intermediate age (Preston et al 1994) and metal-rich halo A stars (Rodgers et al 1981).

In an earlier section, we discussed observational signatures of the Λ CDM hierarchy in the Galactic context. In fact, detailed observations in velocity space are proving to be particularly useful in identifying structures that have long since dispersed in configuration space. In external galaxies, related structures are showing up as low surface brightness features. We do not know what role globular clusters play in the galaxy formation picture, but we include them here because at least one of them appears now to be the nucleus of a disrupted dwarf galaxy.

7.3.2. *Star formation history*

The star formation history (SFH) of the Galaxy has been very difficult to unravel. This is intimately tied up with how gas accretion takes place over billions of years – see Sec. 4. So how do galaxies accrete their gas? Is the infalling gas confined by dark matter? Does the gas arrive cold, warm or hot? Does the gas rain out of the halo onto the disk or is it forced out by the strong disk-halo interaction? These issues have never been resolved, either through observation or through numerical simulation, but we do see evidence today for all of the above processes. For example, the Magellanic System is direct evidence for the accretion of stars, multiphase gas and dark matter onto the Galaxy within the next few Gyr. But as we have seen (Sec. 3.5), the Milky Way - Magellanic relationship appears to be relatively unusual.

Derived star formation histories range from a roughly uniform star formation rate over the history of the disk (Binney et al 2000) to a SFH that was highly peaked at early times (e.g. Twarog 1980; Rocha-Pinto et al 2000; Just 2001; Aumer & Binney 2009). Galaxies of the Local Group show a great diversity in SFH (Grebel 2001), although the average history over the Local Group appears consistent with the mean cosmic history (Hopkins et al 2001). The present emphasis is on star formation studies that make use of the integrated properties of external galaxies, but it should be noted that this is necessarily weighted towards the most luminous populations. Key results for external galaxies are reviewed in Sec. 7.2.5. It was concluded that environmental effects are very significant in determining the SFH for individual galaxies.

The conventional approach to the study of chemical evolution in galaxy disks is to consider the solar neighbourhood a closed box, and to assume that it is representative of all disks. Simple mathematical formulations have developed over the past 40 years (van den Bergh 1962; Schmidt 1963; Pagel & Patchett 1975; Talbot & Arnett 1971; Tinsley 1980; Twarog 1980; Pitts & Tayler 1989). Most observations are interpreted within this framework. The SFH is quantified in terms of stellar age, stellar (+ gas) metallicity and, to a lesser extent, the existing gas fraction.

The use of broadband photometry coupled with stellar population synthesis is a well-established technique for probing the SFH of galaxy populations from integrated light. The power of the method is its simplicity, although it cannot uniquely disentangle the age-metallicity degeneracy (Bica et al 1990; Charlot & Silk 1994).

Another widely used technique is the Lick index system (Burstein et al 1984) further refined in Worthey et al (1994) and Trager et al (1998). In this system, the $H\beta$ index is the primary age-sensitive spectral indicator, whereas the Mg and Fe indices are the primary metallicity indicators. The Lick indices have well known limitations: they correspond to low spectroscopic resolution ($8\text{--}9\text{\AA}$), require difficult corrections for internal galaxy motions, and are not calibrated onto a photometric scale. Furthermore, two of the most prominent Lick indices – Mg_2 $\lambda 5176$ and Fe $\lambda 5270$ – are now known to be susceptible to contamination from other elements, in particular Ca and C (Tripicco & Bell 1995).

How best to measure galaxy ages is a subject with a long history. The most reliable methods to date involve the low order transitions ($n < 4$) of the Balmer series. Ages derived from the $H\gamma$ equivalent width have been used by Jones & Worthey (1995). Rose (1994) and Caldwell & Rose (1998) have pioneered the use of even higher order Balmer lines to break the age-metallicity degeneracy (Worthey 1994). These higher order lines are less affected by Balmer line emission from the interstellar medium. They develop a line ratio index Hn/Fe which is a sum over $H\gamma$, $H\delta$ and $H8$ lines with respect to local Fe lines (e.g. Concannon et al 2000). Ultimately, full spectrum fitting matched to spectral synthesis models holds the most promise (Vazdekis 1999). The new models, which have a fourfold increase in spectroscopic resolution compared to the Lick system, show that the isochrone or isochemical grid lines overlaid on a plot of two Lick indices are more orthogonal than the Worthey models.

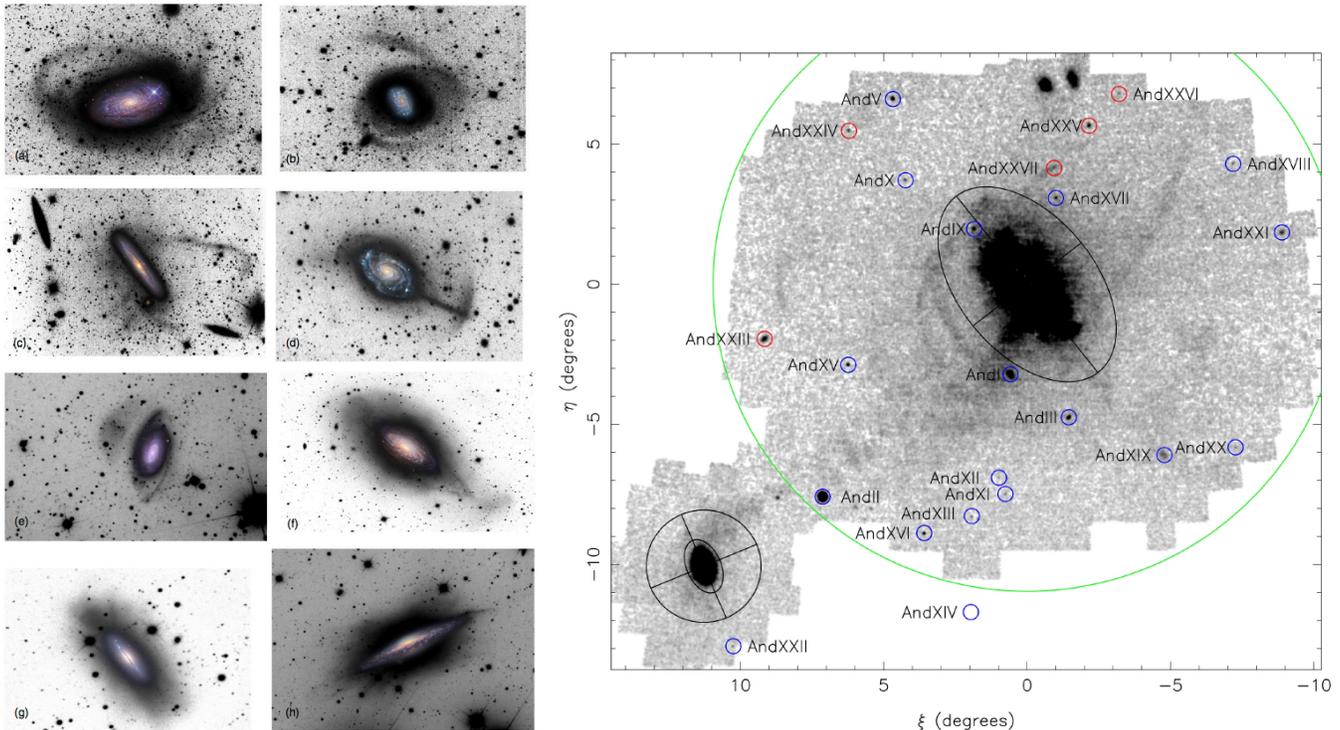


FIG. 20.— (Left) Examples of normal spirals with faint stellar streamers in the outer halo taken from a recent survey by Martinez-Delgado et al (2010). (Right) M31 and M33 system (PANDAs survey; Richardson et al 2011); many associated dwarfs are also shown.

7.3.3. Low surface brightness structures in galaxies

Dynamical interaction between galaxies lead to a range of structures including stellar shells (Malin & Carter 1980; Quinn 1984), fans (Weil et al 1997), and tidal streamers (Gregg & West 1998; Calcanero-Roldan et al 2000; Zheng et al 1999; Zucker et al 2004). Some excellent examples are shown in Fig. 20. We see evidence of multiple nuclei, counter-rotating cores, and gas in polar orbits. At low light levels, the outermost stellar contours of spiral disks appear frequently to exhibit departures from axisymmetry (Rix & Zaritsky 1995). The same is true for spiral arms in all Hubble types (Schoenmakers et al 1997; Sanchez-Gil et al 2011).

The stellar streamers are particularly interesting as these may provide important constraints on galaxy models, particularly as kinematic measurements become possible through the detection of planetary nebulae. More than a dozen stellar streams are already known and this is probably indicative of a much larger population at very low surface brightness. Johnston et al (2002) show that stellar streamers can survive for several Gyr and are only visible above the present optical detection limit (~ 30 V mag arcsec $^{-2}$) for roughly 4×10^8 yr. A few galaxy groups (e.g. the Leo group) do show largescale HI filaments that can remain visible for many Gyr.

Deep CCD imaging has revealed a stellar loop around NGC 5907 (Shang et al 1998) and a stellar feature extending from NGC 5548 (Tyson et al 1998). The technique of photographic amplification has revealed stellar streamers in about ten sources (Malin & Hadley 1997; Calcanero-Roldan et al 2000; Weil et al 1997). For these particular

observations, the limiting surface brightness is $\mu_V \approx 28.5 \text{ mag arcsec}^{-2}$. For all of these systems, we estimate that the total stream luminosities are in the range $3 - 20 \times 10^7 L_\odot$.

Wide-field CCD cameras have revealed stellar streamers through multiband photometry of millions of individual sources. A pointillist image can then be reconstructed in narrow color intervals so as to enhance features with respect to the field. This has led to the discovery of stellar streams in M31 (Ibata et al 2001a; McConnachie et al 2009) and tidal tails extending from the globular cluster Pal 5 (Odenkirchen et al 2001). This technique has the potential to push much deeper than the direct imaging method described above. The low surface brightness universe is notoriously difficult to observe. Modern telescope and instrument designs are simply not optimized for this part of parameter space. Many claims of diffuse light detections in the neighbourhood of galaxies have been shown to arise from internal scattered light.

Looking farther afield, we see evidence for discrete accretion events in the making. The Galaxy is encircled by satellite galaxies that appear confined to one or two great streams across the sky (Lynden-Bell & Lynden-Bell 1995). The most renowned of these are the Magellanic Clouds and the associated HI Magellanic stream. All of these are expected to merge with the Galaxy in the distant future, largely due to the dynamical friction from the extended halo.

In Sec. 7.3.8, we discuss ‘moving groups’ identified within the Galaxy from proper motion and spectroscopic surveys. Their projected surface brightness is $\mu_r = 30 - 34 \text{ mag arcsec}^{-2}$, below the limit of modern imaging techniques.

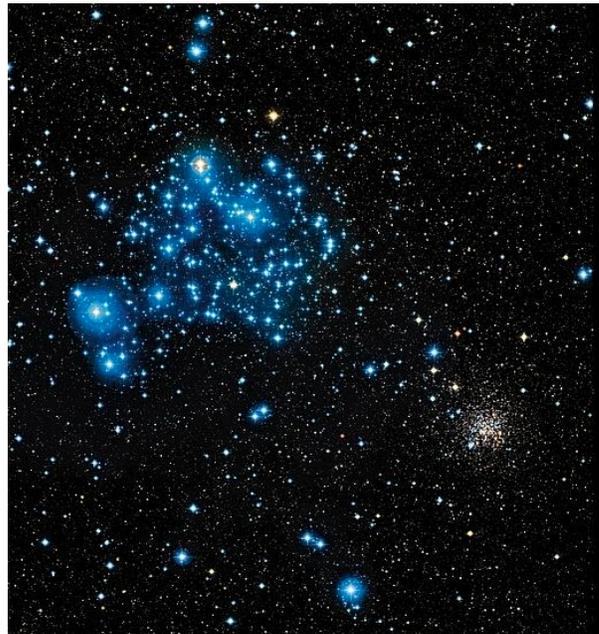
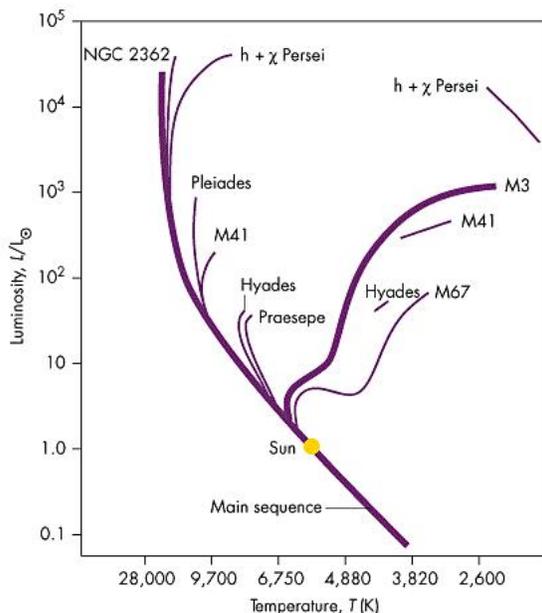


FIG. 21.— (Left) Hertzsprung-Russell diagram for stellar clusters over a wide range in age. (Right) CFHT imaging of two open clusters M35 (NE) and NGC 2158 (SW) in Gemini. The young cluster M35 (NGC 2168) has a mass of roughly $2500 M_\odot$ and is 850 pc from the Sun. The massive cluster NGC 2158 is four times further away and about 1 Gyr old.

7.3.4. Star clusters as a probe of the environment

A key characteristic of star formation is the creation of star clusters. The initial cluster mass function (ICMF), if it can be determined reliably, provides important information on the physical conditions under which the clusters formed (Elmegreen 2010; Larsen 2009; Bastian 2012). In principle, this function describes the creation of all clusters, from globulars to open clusters, to loose associations. The ICMF gives the number of clusters formed per unit mass, N_c , where present-day cluster formation is often described with a power-law with index γ : $dN_c/dm_c \propto m_c^\gamma$. A flat ICMF ($\gamma \gtrsim -1.5$), which favours massive clusters, appears to be indicative of high-pressure environments (e.g. starbursts, mergers), whereas a steep ICMF ($\gamma \lesssim -2$) is typically associated with quiescent environments (e.g. solar neighbourhood).

In recent years, much of the discussion on cluster formation has focussed on how many stars are born in bound vs. unbound systems. In galaxies with low star formation rates, the fraction born in bound clusters appears to be close to 10% but this increases dramatically in starburst systems (Bastian 2012). But we stress that the fraction born in bound vs. unbound clusters is secondary to the question of what fraction of stars are born in *instantaneous* clusters. As long as the collapsing cloud is homogenised before the first stars are formed, and therefore carries a unique chemical imprint, the long-term survival of the cloud is of lesser importance. It is possible that a much higher fraction of stars in all environments form in *instantaneous* clusters, most of which have long since diffused into the general background.

As we show in Sec. 8, the ICMF is of fundamental importance to chemical tagging, and unravelling the original star formation history. A few large clusters are easier to ‘tag’ than many smaller clusters, assuming that these are

all formed in homogenised collapsing clouds. Thus, the success of chemical tagging depends critically on how a particular component of the Galaxy formed. But first we discuss some specific types of bound stellar populations found throughout the Local Group – open clusters, globular clusters, dwarf galaxies.

7.3.5. *Open clusters*

In the context of near-field cosmology, we believe that the thick disk and the old open clusters of the thin disk are among the most important diagnostics. The open clusters are the subject of an outstanding and comprehensive review by Friel (1995). Here, we summarize the properties that are most important for our purpose.

Both old and young clusters are part of the thin disk; see Fig. 21 for some excellent examples of these beautiful objects. Their key attribute is that they provide a direct time line for investigating change, which we explore in the next section. The oldest open clusters exceed 10 Gyr in age and constitute important fossils (Phelps & Janes 1996). The survival of these fossil clusters is an interesting issue in its own right. Friel (1995) finds no old open clusters within a galactocentric radius of 7 kpc; these are likely to have disrupted or migrated out of the central regions (van den Bergh & McClure 1980). It has long been recognized that open clusters walk a knife edge between survival and disruption (King 1958).

Like field stars in the disk, Janes & Phelps (1994) find that the old cluster population (relative to Hyades) is defined by 375 pc scale height exponential distribution, whereas young clusters have a 55 pc scale height. Again, like the field stars, vertical abundance gradients have not been seen in open clusters (Friel & Janes 1993), although radial gradients are well established (Friel 1995; van den Bergh 2000). For old open clusters, Twarog et al (1997) claim evidence for a stepped radial metallicity distribution where $[\text{Fe}/\text{H}] \approx 0$ within 10 kpc, falling to $[\text{Fe}/\text{H}] \approx -0.3$ in the outer disk. However, this effect is not seen in young objects, e.g. HII regions and B stars (Henry 1998).

Both the old and young open clusters show essentially the same radial trend in metallicity. After reviewing the available observations, Friel (1995) finds no evidence for an age-metallicity relation for open clusters. In agreement with Eggen & Sandage (1969), she notes that over the entire age of the disk, at any position in the disk, the oldest clusters form with compositions as enriched as those of much younger objects.

These remarkable observations appear to indicate that shortly after the main epoch of baryon dissipation, the thin disk was established at least as far out as 15 kpc. The oldest open clusters approach the age of the thick disk. Since, in Sec. 7.2.3, we noted that the thick disk is likely to be a ‘snap frozen’ picture of the thin disk shortly after disk formation, we would expect the truncation of the thick disk to reflect the extent of the thin disk at the epoch of the event that puffed up the thick disk.

7.3.6. *Globular clusters*

We have long suspected that globular clusters are the fossil remnants of violent processes in the protogalactic era (Peebles & Dicke 1968; Conroy & Spergel 2011). But there is a growing suspicion that globulars are telling us more about globulars than galactic origins (Harris 2002). The Milky Way has about 150 globular clusters with 20% lying within a few kiloparsecs of the Galactic Center. They constitute a negligible fraction of the light and mass (2%) of the stellar halo today. Their significance rests in their age. The oldest globular clusters in the outer halo have an age of 13 ± 2.5 Gyr (90% confidence).

The ages of the oldest globular clusters in the inner and outer halo, the Large Magellanic Cloud and the nearby Fornax and Sgr dwarf spheroidal galaxies show a remarkable uniformity. To a precision of ± 1 Gyr, the onset of globular cluster formation was well synchronized over a volume centered on our Galaxy with a radius > 100 kpc (Da Costa 1999).

Globular cluster stars are older than the oldest disk stars, e.g. white dwarfs and the oldest red giants. These clusters are also more metal poor than the underlying halo light in all galaxies and at all radii (Harris 1991), but again there are exceptions to the rule. Since Morgan’s (1959) and Kinman’s (1959) classic work, we have known that there are two distinct populations of globular clusters in the Galaxy. The properties that we associate with these two populations today were derived by Zinn (1985) who showed that they have very different structure, kinematics and metallicities. The halo population is metal-poor ($[\text{Fe}/\text{H}] < -0.8$) and slowly rotating with a roughly spherical distribution; the disk population is metal-rich ($[\text{Fe}/\text{H}] > -0.8$) and in rapid rotation.

A major development has been the discovery of young globular clusters in disturbed or interacting galaxies, e.g., NGC 1275 (Holtzman et al 1992), NGC 7252 (Whitmore et al 1993) and the Antennae (Whitmore & Schweizer 1995). Schweizer (1987) first suspected that globular clusters were formed in mergers. Later, Ashman & Zepf (1992) predicted that the HST would reveal young globular clusters through their compact sizes, high luminosities and blue colors. The very high internal densities of globular clusters today must partly reflect the conditions when they were formed. Harris & Pudritz (1994) present a model for globular clusters produced in fragmenting giant molecular clouds, which are of the right mass and density range to resemble accretion fragments in the Searle-Zinn model.

Globular clusters have been elegantly referred to as ‘canaries in a coal mine’ (Arras & Wasserman 1999). They are subject to a range of disruptive effects, including two-body relaxation and erosion by the tidal field of their host galaxy, and the tidal shocking that they experience as their orbits take them through the galactic disk and substructure in the dark halo. In addition to self-destruction through stellar mass loss, tidal shocking may have been very important in the early universe (Gnedin et al 1999). If globular clusters originally formed in great numbers, the disrupted clusters may now contribute to the stellar halo (Norris & Ryan 1989; Oort 1965). Halo field stars and globular clusters in the Milky Way have similar mean metallicities (Carney 1993); however the metallicity distribution of the halo field stars

extends to much lower metallicity ($[\text{Fe}/\text{H}] \simeq -5$) than that of the globular clusters ($[\text{Fe}/\text{H}] \simeq -2.2$). We note again the remarkable similarity in the metallicity range of the globular clusters and the thick disk ($-2.2 \lesssim [\text{Fe}/\text{H}] \lesssim -0.5$).

In the nucleated dwarf elliptical galaxies (Binggeli et al 1985), the nucleus typically provides about 1% of the total luminosity; globular clusters could be considered as the stripped nuclei of these satellite objects without exceeding the visible halo mass (Zinnecker & Cannon 1986; Freeman 1993). It is an intriguing prospect that the existing globular clusters could be the stripped relicts of an ancient swarm of protogalactic stellar fragments, i.e. the original building blocks of the Universe.

In the Searle-Zinn picture, globular clusters are intimately linked to *gas-rich*, protogalactic infalling fragments. Multiple stellar populations have recently been detected in ω Cen, the most massive cluster in the Galaxy (Lee et al 1999). How did ω Cen retain its gas for a later burst? It now appears that it was associated with a gas-rich dwarf, either as an *in situ* cluster or as the stellar nucleus. The present-day cluster density is sufficiently high that it would have survived tidal disruption by the Galaxy, unlike the more diffuse envelope of this dwarf galaxy. The very bound retrograde orbit supports the view that ω Cen entered the Galaxy as part of a more massive system whose orbit decayed through dynamical friction.

If globular clusters are so ancient, why are the abundances of the most metal-poor population as high as they are? Because it does not take much star formation to increase the metal abundance up to $[\text{Fe}/\text{H}]=-1.5$ (Frayer & Brown 1997), the cluster abundances may reflect low levels of star formation even before the first (dark+baryon) systems came together.

Contrary to a common assumption, old age is not uniquely associated with low metallicity (cf. Sec. 5.3). In hindsight, we have known this for many years. Brown & Vanden Bout (1991) discovered a strong CO signal from a galaxy at $z \approx 2.3$. Molecule formation requires the presence of dust which in turn requires substantial enhancements in metallicity. Hamann & Ferland (1999) demonstrate that stellar populations at the highest redshifts currently observed ($z \gtrsim 5$) appear to have solar or super-solar metallicity; CO gas has also been detected at these times (e.g. Yun et al 2000).

We believe that there is no mystery about high abundances at high redshift. The dynamical times in the cores of these systems are short, so there has been time for multiple generations of star formation and chemical enrichment. In this sense, the cores of high redshift galaxies need not be relevant to the chemical properties of the globular clusters, although both kinds of objects were probably formed at about the same time.

The first generation of globular clusters may have been produced in merger-driven starbursts when the primordial fragments came together for the first time. If at least some fragments retained some of their identity while the halo was formed, a small number of enrichment events per fragment would ensure a Poissonian scatter in properties between globular clusters, and multiple populations within individual clusters (Searle & Zinn 1978).

7.3.7. Dwarf galaxies

The formation and long-term evolution of dwarf galaxies are topics of great interest today. Substantial populations are still being identified around the Galaxy and M31 (e.g. Zucker et al 2006, 2007). A defining characteristic is that these are baryons confined by a dark matter halo, although this definition may not cover ‘tidal dwarfs’ if these are ultimately shown to be free of dark matter (Barnes & Hernquist 1992). Very comprehensive reviews of the dwarf population can be found in Tolstoy et al (2009) and McConnachie (2012).

Dwarf galaxies are possibly the best probes of the first stars within the framework of near-field cosmology (Tolstoy et al 2009; Bland-Hawthorn et al 2010b; Karlsson et al 2012, 2013). A substantial fraction of the dwarf galaxy population of the Local Group is expected to have already merged with the halos of M31 and the Galaxy (see Fig. 22). This is supported by the evidence that the stellar Galactic halo has several distinct chemical similarities that it shares with the existing population of dwarf galaxies (Kirby et al 2008; Tolstoy et al 2009 and references therein; Starkenburg et al 2010; Frebel et al 2010a,b; Norris et al 2010a,b). Despite these similarities, the systems are not chemically identical and exactly how the merging process progressed, how the dwarf galaxies that survived are connected to their disrupted counterparts, and how these still-surviving galaxies relate to each other in a cosmological context, is not understood. The tidal streams in Sgr and Carina (Ibata et al 1994; Battaglia et al 2012) provide clear evidence that dwarfs are being disrupted on infall. Furthermore, all but the most massive dwarfs within the virial radii of M31 and the Galaxy show strong evidence of gas depletion (Grcevich & Putman 2009).

The study of old stellar populations in these galaxies provides us with a link to the high-redshift Universe. All Local Group dwarf galaxies, including the recently discovered ultra-faint dwarfs, are known to contain ancient, metal-poor stars (Grebel & Gallagher 2004; Tolstoy et al 2009). The star formation efficiency in dwarfs is very low which means that there have been relatively few star-formation events over cosmic time. The effects of stochasticity therefore plays a relatively big role (Audouze & Silk 1995; Karlsson 2005; Karlsson & Gustafsson 2005). This aspect must be addressed when seeking to understand the evolution of dwarf galaxies, particularly at early times.

Dwarf galaxies show clear evidence that star clusters have formed within them. At the present time, the most metal-poor star clusters are known have an iron abundance just below $[\text{Fe}/\text{H}] = -2$. One of the globular clusters (Cluster 1) in the Fornax dSph currently holds the record with a metallicity $[\text{Fe}/\text{H}] = -2.5$ (Letarte et al 2006). But not all clusters have survived as gravitationally bound objects to the present epoch. Much like the halo, dwarf galaxies have stars with metallicities well below $[\text{Fe}/\text{H}] = -3$ (Kirby et al 2008; Starkenburg et al 2010). The relatively simple environments of the low mass dwarfs raises the prospect of identifying disrupted star clusters at much lower metallicity than has been possible before. For example, Karlsson et al (2012) have identified through chemical tagging a disrupted

star cluster in Sextans with $[\text{Fe}/\text{H}] = -2.7$. If this result is confirmed by a larger survey sample, this will constitute the most metal poor star cluster ever observed. Cluster reconstruction of this kind provides us with a unique tool to probe the formation and early evolution of the present dwarf galaxy population and to determine the sequence of events that led to the formation of the stellar halo of the Milky Way (Bland-Hawthorn et al 2010b).

7.3.8. Structures in phase space

One class of systems that exhibit coherence in velocity space are the open clusters associated with the disk. Here the common space motion of the stars with respect to the Sun is perceived as a convergence of the proper motions to a single point (strictly speaking, minimum volume) on the sky (Boss 1908; see de Zeeuw et al 1999 for a recent application). More than a dozen such systems have been identified this way. However, these are all young open clusters largely associated with the Gould belt. With sufficiently precise kinematics, it may be possible to identify open clusters that have recently dispersed, particularly if the group is confined to a specific radial zone by resonances in the outer disk. For example, Feltzing & Holmberg (2000) show that the metal-rich ($[\text{Fe}/\text{H}] \approx 0.2$) moving group HR 1614, thought to be 2 Gyr old, can be identified in the Hipparcos data set.

Recently, attention has turned to a diverse set of ‘moving groups’ that are thought to be associated with the stellar halo and in some instances are clearly fossils associated with accretion events in the distant past. The evidence for these groups dates back to the discovery of the halo itself. Shortly before the publication of the landmark ELS paper, Eggen & Sandage (1959) discovered that the nearby high-velocity star, Groombridge 1830, belongs to a moving group now passing through the Galactic disk.

In a long series of papers, Eggen went on to ‘identify’ a number of moving groups, some of which appear to encompass the solar neighbourhood, and others that may be associated with the halo. The relevant references are given by Taylor (2000). Various authors have noted that many of the groups are difficult to confirm (Griffin 1998; Taylor 2000). More systematic surveys over the past few decades have identified a number of moving populations associated with the halo (Freeman 1987; Majewski 1993) although the reality of some of these groups is still debated. The reality of these groups is of paramount importance in the context of halo formation. Majewski et al (1996) suspect that much or all of the halo could exhibit phase-space clumping with data of sufficient quality.

In recent years, the existence of kinematic sub-structure in the galactic halo has become clear. Helmi et al (1999) identified 88 metal-poor stars within 1 kiloparsec of the Sun from the Hipparcos astrometric catalogue. After deducing accurate 3-D space motions, they found a highly significant group of 8 stars that appear clumped in phase space and confined to a highly inclined orbit. In fact, more than a dozen stellar streams have now been identified in the Galactic halo, the most recent being the Aquarius stream identified in the RAVE stellar survey (Williams et al 2011); for an updated list, see Helmi (2009).

The most dramatic evidence is surely the highly disrupted Sgr dwarf galaxy identified by Ibata et al (1994; 1995). These authors used multi-object spectroscopy to uncover an elongated stellar stream moving through the plane on the far side of the Galaxy. The Sgr dwarf is a low mass dwarf spheroidal galaxy about 25 kpc from the Sun that is presently being disrupted by the Galactic tidal field. The long axis of the prolate body (axis ratios $\sim 3:1:1$) is about 10 kpc, oriented perpendicular to the Galactic plane along $\ell = 6^\circ$ and centered at $b = -15^\circ$. Sgr contains a mix of stellar populations, an extended dark halo (mass $\geq 10^9 M_\odot$) and at least four globular clusters (Ibata et al 1997). The Sgr stream has since been recovered by several photometric surveys (Vivas et al 2001; Newberg et al 2002; Ibata et al 2001c).

N-body simulations have shown that stellar streams are formed when low mass systems are accreted by a large galaxy (e.g. Harding et al 2001). Streamers remain dynamically cold and identifiable as a kinematic substructure long after they have ceased to be recognizable in star counts against the vast stellar background of the galaxy (Tremaine 1993; Ibata & Lewis 1998; Johnston 1998; Helmi & White 1999).

Within the Galaxy, moving groups can be identified with even limited phase-space information (de Bruijne 1999; de Zeeuw et al 1999). This also holds for satellites orbiting within the spherical halo, since the debris remains in the plane of motion for at least a few orbits (Lynden-Bell & Lynden-Bell 1995; Johnston et al 1996). But a satellite experiencing the disk potential no longer conserves its angular momentum and its orbit plane undergoes strong precession (Helmi & White 1999). In Fig. 22, we show the sky projection of a satellite 8 Gyr after disruption. These more complex structures are usually highly localized, and therefore easy to recognize in the space of conserved quantities like energy and angular momentum for individual stars.

The evolution in phase space of a disrupting satellite is well behaved as its stars become phase mixed. Its phase space flow obeys Liouville’s theorem, i.e. the flow is incompressible. Highly intuitive accounts are given elsewhere (Carlberg 1986; Tremaine 1999; Hernquist & Quinn 1988). It should be possible to recognize partially phase-mixed structures that cover the observed space, although special techniques are needed to find them (e.g. Antoja et al 2012).

The astrometric mission Gaia will derive 6-dimensional phase space coordinates and spectrophotometric properties for millions of stars within a 20 kiloparsec sphere – the GaiaspHERE. The ambitious Gaia mission will obtain distances for up to 90 million stars with better than 5% accuracy, and measure proper motions with an accuracy approaching *microarcsec* per year. If hierarchical Λ CDM is correct, there should be thousands of coherent streamers that make up the outer halo, and hundreds of partially phase-mixed structures within the inner halo. A satellite experiencing the disk potential no longer conserves its angular momentum and its orbit plane undergoes strong precession (see Fig. 23(c) and (d)). In Fig. 23(a) and (b), Helmi et al (1999) demonstrate the relative ease with which Gaia will identify substructure within the stellar halo.

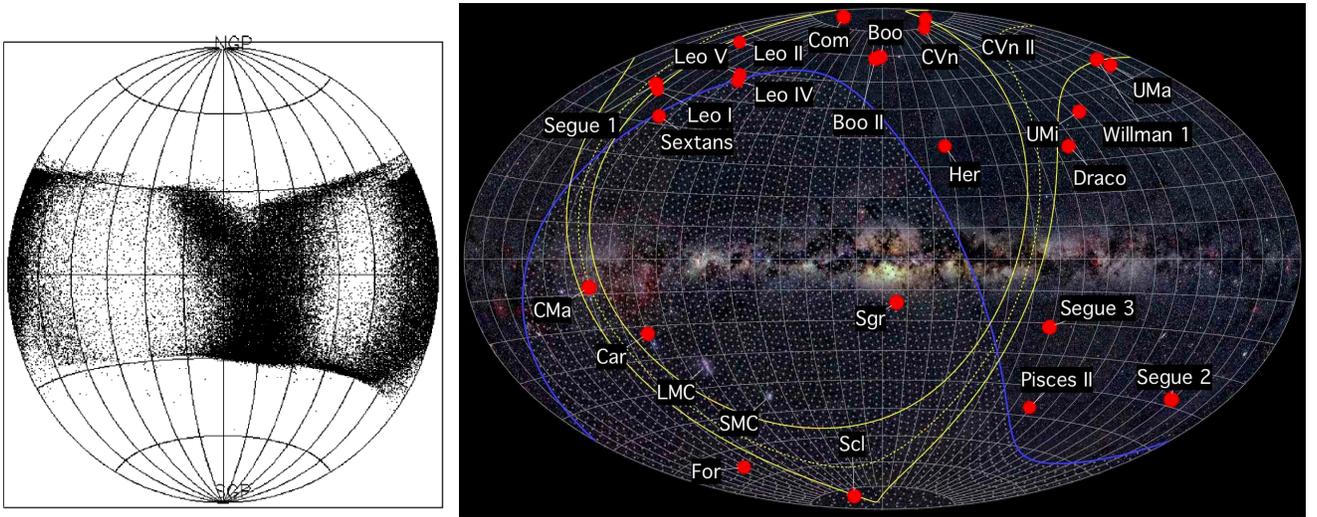


FIG. 22.— (Left) Aitoff projection of a satellite in orbit about the Milky Way as it would appear after 8 Gyr. While stars from the disrupted satellite appear to be dispersed over a very wide region of sky, it will be possible to deduce the parameters of the original event using special techniques (see text). [We acknowledge A. Helmi and S. White for this image.] (Right) Aitoff projection of most of the known dwarfs in orbit about the Galaxy. For decades, astronomers have argued that these are broadly confined to a plane rather than spherically distributed, an effect that is also seen in the PANDAs survey for the dwarfs in M31. [We acknowledge A. Frebel for this image.]

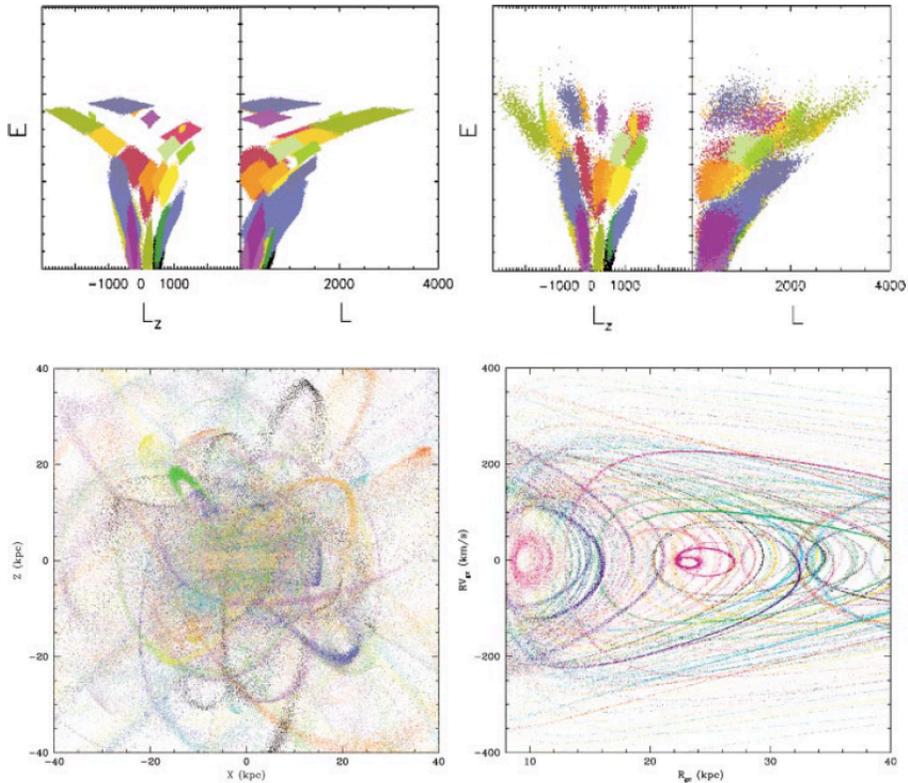


FIG. 23.— (a) Initial distribution of particles for 33 systems falling into the Galactic halo is integral of motion space. (b) The final distribution of particles in (a) after 12 Gyr after convolution with the errors expected for Gaia. [We acknowledge A. Helmi for these images.] A simulation of the baryon halo built up through accretion of 100 satellite galaxies. (c) The different colors show the disrupted remnants of individual satellites. (d) This is the same simulation shown in a different coordinate frame, i.e. the orbit radius (horizontal) plotted against the observed radial velocity (vertical) of the star. [We acknowledge P. Harding and H. Morrison for these images.]

8. RECONSTRUCTING THE PAST THROUGH CHEMICAL TAGGING

8.1. *Unravelling a dissipative process*

Most galaxies from about $z \sim 2$ through to the present day appear to be dominated by disks. During the most active period ($z \sim 1-3$), galaxies appear very clumpy and display turbulent kinematics. The early star-forming clumps which are particularly prominent in deep HST images produced stars that must reside within galaxies today (Elmegreen & Elmegreen 2006). Did they break up and form an old puffed-up thick disk (Kroupa 2002) or surviving fossil streams (FBH), did they spiral into the galaxy centre to form a bulge (Ceverino et al 2010), or did they migrate across the galaxy due to the formation of transient bars or spiral waves (Minchev et al 2012)? Answers to these questions continue to elude us: the most advanced simulations are unable to resolve individual star clusters and observations at many cosmic epochs are difficult to interpret. But the HERMES survey opens up a new window to the fate of these ancient star clusters by providing detailed chemical information for vast numbers of individual stars.

So how did the Galactic disk form and can the sequence of events ever be unravelled from the vast stellar inventory? This will require that some of the residual inhomogeneities from prehistory escaped the dissipative process at an early stage. Fossil hunting to date has concentrated mostly on the stellar halo, but a key source of information will be the thick disk. This is believed to be a 'snap frozen' relic which formed during or shortly after the last major epoch of dissipation, or it may have formed from infalling systems early in the life of the Galaxy.

In an earlier review (Freeman & Bland-Hawthorn 2002), we raised the prospect of 'chemical tagging' as a means to explore the early history of the halo and the thick disk by looking for discrete substructures in chemical \mathcal{C} -space. We proposed for the first time that this will require high signal to noise, echelle spectroscopy of up to a million stars. This idea is now central to major new and proposed stellar surveys, in particular, the HERMES project (Freeman & Bland-Hawthorn 2008) and the 4MOST project (de Jong et al 2012).

Our program has a short-term and a long-term goal. The short-term goal is to quantify the size and structure of the multi-dimensional chemical abundance space (\mathcal{C} -space) for all major components of the Galaxy. We seek to establish how many axes in \mathcal{C} -space are decoupled and have large intrinsic dispersions. A critical test of chemical tagging in the short term is that stellar streams in the halo, identified from detailed phase space information, are highly localized in \mathcal{C} -space, or are confined to chemical tracks. These trajectories presuppose that stars form in a closed box through progressive enrichments of the gas, leading to stars dispersed along a narrow track in a complex chemical space.

The long-term goal is to identify unique chemical signatures in the thick disk, originating from different formation sites, for star clusters which have long since dispersed. This will require precise chemical abundances for heavy elements such that a star can be localized to a discrete point in \mathcal{C} -space. If the star clusters originally formed outside the Galaxy in a bound infalling system, the stellar abundances may fall along a chemical track, rather than a discrete point in \mathcal{C} -space.

The ultimate goal of cosmology, both near and far, must be to explain how the Universe has arrived at its present state. It is plausible – although difficult to accept – that nature provides fundamental limits of knowledge, in particular, epochs where the sequence of events are scrambled. Our 'intuition' is that any phase dominated by relaxation or dissipation probably removes more information than it retains. But could some of the residual inhomogeneities from prehistory have escaped the dissipative process at an early stage? We may not know the answer to this question with absolute certainty for many years. In the absence of certainty, we consider what might be the likely traces of a bygone era prior to the main epoch of baryon dissipation.

In order to follow the sequence of events involved in dissipation, we propose that the critical components which need to be re-assembled are the individual star clusters which formed at each stage. Since most stars are born in dense clusters, the formation and evolution of galaxies today must involve millions of discrete cluster events throughout their history. We would like to establish the evolving mass function of star clusters, their chemical composition, formation and survival rate as a function of cosmic time. Galaxy-wide enrichment from the fall-out of nuclear winds or mergers would be evident in the fossil record of reconstructed star clusters, assuming these provide an unbiased sampling of cosmic time regardless of the star formation history.

8.2. How many star clusters ?

If we consider our volume of interest to be the Solar torus, defined as the annulus 2 kpc thick in radius centered on the Solar circle at $R_o \approx 8$ kpc, then the cluster birth rate within this region is

$$\frac{d^2 \mathcal{N}}{dt d \ln(M_*/ M_\odot)} = 3 \times 10^3 \dot{M}_0 \left(\frac{M_*}{10^4 M_\odot} \right)^{-1} \text{ Gyr}^{-1}, \quad (37)$$

where $\dot{M}_0 = \dot{M} / (M_\odot \text{ yr}^{-1})$ and \dot{M} is the total star formation rate in the Galaxy and the upper and lower cluster mass limits are $(M_{*,\min}, M_{*,\max}) = (50 M_\odot, 2 \times 10^5 M_\odot)$. Thus, for example, in the last Gyr we expect $\sim 10^4$ clusters whose masses are within a factor of three of $10^4 M_\odot$ to have formed within the Solar torus if the star formation rate over this time has been close to its present value of $\sim 3 M_\odot \text{ yr}^{-1}$ (McKee & Williams 1997). This result is quite insensitive to the assumed limits on the cluster mass function or the extent of the star-forming disk. Changing $M_{*,\min}$ or $M_{*,\max}$ by a factor of 10 alters the estimated cluster birth rate by only $\sim 20\%$.

All but a small fraction of clusters formed are unbound, and the stars born within them drift off to become part of the field star population. By comparing catalogs of open and embedded clusters, Lada & Lada (2003) estimate that less than 10% of clusters survive more than 10 Myr, and that only 4–7% survive for 100 Myr¹². Moreover, analysis of extragalactic cluster populations (albeit with masses significantly larger than the typical Milky Way cluster) suggests

¹² We point out, however, that the distribution must have a long tail because a few open clusters have ages $\gg 1$ Gyr (Friel 1995).

that the survival fraction is nearly independent of mass until two-body relaxation becomes the dominant destruction process at times $\gg 100$ Myr after cluster formation (Fall et al 2005, 2009; Fall 2006; Whitmore et al 2007). During this first 100 Myr, roughly 80 – 90% of clusters are disrupted independent of mass, while two-body relaxation at longer times preferentially destroys low mass clusters. Since we do not strive for better than 10% accuracy in our prediction of the disrupted cluster birth rate, we will not introduce the added complication of a mass- or time-dependent disruption rate. Instead we simply assume that a fixed 95% of all clusters that are born are disrupted, independent of mass and time. This will slightly overestimate the number of disrupted clusters that formed less than 100 Myr, but for a relatively quiescent star formation history such as the Galaxy’s, these clusters constitute a negligible fraction of the total population. Thus, if we target stars older than 100 Myr, the birth rate of now-disrupted clusters is given by equation 37 to within 10%.

8.3. Cluster chemistry

Since most stars are not born in clusters that remain bound, and thus do not end up in open clusters more than 10 Myr later, the clusters observed by De Silva and collaborators presumably represent a biased sample. They are either the innermost remnants of clusters that lost most of their members and produced only a small bound core (e.g. Kroupa et al 2001), or they are clusters that remained bound because they had star formation efficiencies significantly higher than the typical 10 – 30% (Lada & Lada 2003). We would therefore like to define some rough theoretical expectations about the chemical homogeneity of the disrupted cluster population.

Mixing of heavy elements in protoclusters is governed by several time scales. The dominant mode of chemical mixing in molecular clouds is turbulent diffusion (Murray & Lin 1990). The time scale associated with this process is $t_{\text{diff}} \sim H^2/K$ (Xie et al 1995), where $H = (H_i^{-1} - H_{\text{H}_2}^{-1})^{-1}$ is a composite between the scale heights of the H_2 molecules, $H_{\text{H}_2} = n_{\text{H}_2}/|\nabla n_{\text{H}_2}|$, and of the species i whose diffusion time we are computing, $H_i = n_i/|\nabla n_i|$. Here n_{H_2} and n_i are the number densities of the two species. The quantity K is the diffusion coefficient, which is of order σL , where σ is the turbulent velocity dispersion and L is the correlation length of the turbulence. Observations indicate that the turbulent motions in molecular clouds are correlated on the scale of the entire cloud (Heyer & Brunt 2004), so L is roughly the size of a cloud. If we consider a species that is distributed with an inhomogeneity comparable to or greater than that of the molecular hydrogen, $H_i \lesssim H_{\text{H}_2}$, and a composition gradient on the scale of the entire cloud, $H_i \sim L$, then $t_{\text{diff}} \sim L/\sigma = t_{\text{cr}}$, where t_{cr} is the cloud crossing time. Thus, the time required for turbulence to smooth out a large-scale chemical gradient in a protocluster gas cloud is of order the cloud crossing time; smaller scale gradients disappear more quickly, with the time required to smooth out the gradient varying as the square of its characteristic size scale.

Star formation in a molecular cloud also appears to begin in no more than a crossing time after its formation (Tamburro et al 2008), so there is no clear separation in time scales between chemical homogenization and the onset of star formation. This suggests that molecular clouds cannot be too far from homogeneous when they are assembled, but that the homogenization need not be total, since small residual chemical gradients will be wiped out as star formation proceeds. Thereafter, changes in element abundances, and inhomogeneities in the resulting stars, may be produced by supernovae within the cloud. In principle, even a single supernova is sufficient to change the chemical signature measurably. For example, De Silva et al (2007b) find that scatter in Fe abundance in the HR 1614 moving group is roughly 0.01 dex. The Fe content of the Sun is $\approx 10^{-3} M_{\odot}$ so the measured De Silva et al (2007b) scatter corresponds to a scatter in mass fraction of $\sim 10^{-5}$. A supernova from a 15 M_{\odot} star produces $\sim 10^{-1} M_{\odot}$ of Fe (Woosley & Weaver 1995), so if that supernova occurs inside a star-forming cloud that continues forming stars thereafter, the change in iron abundance will be measurable even if the supernova ejecta are mixed with $10^4 M_{\odot}$ of pre-supernova gas. The time required for a very massive star to evolve from formation to explosion defines the supernova time scale, $t_{\text{SN}} \approx 3$ Myr. Thus, while the “initial conditions” for cluster formation are well-mixed, we only expect star clusters to be homogeneous if they are assembled on time scales shorter than t_{SN} .

8.4. Chemical homogeneity

To determine under what conditions this requirement is satisfied, we must compare t_{SN} to the cluster formation timescale t_{form} . There is considerable debate about whether star cluster formation occurs in a single crossing time (Elmegreen 2000; Hartmann et al 2001; Elmegreen 2007; Pflamm-Altenburg & Kroupa 2007) or 3 – 4 crossing times (Tan et al 2006; Huff & Stahler 2006; Krumholz & Tan 2007; Jeffries 2007; Nakamura & Li 2007; Matzner 2007). Since the latter is more restrictive, we adopt it. This leaves the problem of estimating the crossing time. To do so, we use a convenient form that expresses it in terms of a cloud’s mass M and column density Σ (Tan et al 2006)

$$t_{\text{cr}} = \frac{0.95}{\sqrt{\alpha_{\text{vir}} G}} \left(\frac{M}{\Sigma^3} \right)^{1/4}, \quad (38)$$

where α_{vir} is the cloud’s virial ratio (Bertoldi & McKee 1992), essentially its ratio of kinetic to gravitational potential energy up to geometric factors. Observed star-forming clouds all have $\alpha_{\text{vir}} \approx 1 - 2$ (McKee & Ostriker 2007) so we take $\alpha_{\text{vir}} = 1.5$ as typical, but Σ is somewhat more difficult to estimate. Partly-embedded clusters detected in nearby star-forming regions using 2MASS and *Spitzer*/IRAC have surface densities of a few 0.01 g cm^{-2} (Adams et al 2006; Allen et al 2007), but because these samples are selected in the near-infrared they are strongly biased against more deeply-embedded and thus presumably younger clusters. Indeed, these samples are all dominated by stellar rather

than gas mass, and thus have likely undergone significant expansion. Observations of gas-dominated cluster-forming regions show column densities of $\sim 0.1 - 3 \text{ g cm}^{-2}$ (Plume et al 1997; Mueller et al 2002; Shirley et al 2003; Faundez et al 2004; Fontani et al 2005); the data are summarized in Fig. 1 of Fall et al (2010). We therefore adopt $\Sigma = 0.3 \text{ g cm}^{-2}$ as a fiducial value. This is toward the low end of the observed range, giving a high (and thus conservative) estimate of the crossing time. Together with an assumed formation time $t_{\text{form}} = 4t_{\text{cr}}$ gives

$$t_{\text{form}} \approx 3.0 \left(\frac{\epsilon}{0.2} \right)^{-1/4} \left(\frac{M_*}{10^4 M_\odot} \right)^{1/4} \text{ Myr}, \quad (39)$$

where $\epsilon = M_*/M$ is the star formation efficiency, i.e. the fraction of the initial cloud mass that is converted into stars. Observations suggest this is typically 10 – 30% (Lada & Lada 2003).

Since t_{form} is the time required to form all the stars, we conclude that for clusters $10^4 M_\odot$ or smaller, almost all of the stars will form before the first supernova occurs within the cluster, so the gas will be chemically homogeneous. This is probably somewhat conservative, since only the most massive stars go supernova in 3 Myr, and such a massive star is unlikely to be found in a cluster containing only $10^4 M_\odot$ of stars. Moreover, t_{SN} is the time delay between formation of a massive star and the resulting supernova, and a massive star is not necessarily the first to form in a given cluster. If we assume that massive star formation occurs on average halfway through the cluster formation process, then we only require $t_{\text{form}} < 2t_{\text{SN}}$ to ensure there is no self-pollution, which corresponds to a cluster mass $M_* = 1.6 \times 10^5 M_\odot$. We therefore conclude¹³ that clusters smaller than $10^4 M_\odot$ should essentially all be chemically homogeneous, while a significant fraction of clusters up to $\sim 10^5 M_\odot$ will also be homogeneous. This is to be expected if the surface density of their more massive cluster-forming regions are higher than for the lower-mass clusters.

8.5. Unique chemical signatures

The analysis in §3.2 allows us to derive an order-of-magnitude estimate of the number of unique chemical signatures that are needed to identify stars with their parent clusters. Within the Solar toroid, the total number of tags required comes from integrating equation 37 over the mass limits, viz.

$$\frac{d\mathcal{N}}{dt} = \int_{\ln M_{*,\text{min}}}^{\ln M_{*,\text{max}}} \frac{d^2\mathcal{N}}{dM dt} d(\ln M_*) = 6 \times 10^5 \text{ Gyr}^{-1} \quad (40)$$

Thus, assuming that we could select stars with an age similar to the Sun ($t_o \approx 4.57 \pm 0.5 \text{ Gyr}$), we would need of order 6×10^5 unique chemical signatures. Bland-Hawthorn & Freeman (2004) suggested going after dozens of independent elements that could be measured accurately enough to distinguish a “weak” line strength from a “strong” one. We can express this as $N_c = N_L^m$ where N_L is the number of distinct line strengths and m is the number of independent elements or element groups. After detailed consideration of what is possible in the context of a fibre-fed high-resolution spectrograph, the HERMES project has settled on 8 independent groups with 5 measurably distinct abundance levels, i.e. $N_c \sim 5^8 \approx 4 \times 10^5$ unique chemical signatures. This is the target specification for the million-star Galactic Archaeology survey at the Anglo-Australian Telescope (AAT) due to commence in 2012 (Freeman et al 2010). The independent groups (e.g. alpha elements) include one or more elements that show some variation with respect to [Fe/H].

The volume of the potential chemical space (\mathcal{C} -space; Bland-Hawthorn & Freeman 2004) accessible to the HERMES survey is barely adequate for our proposed study. But further constraints will come from stellar ages at least for some stars. An important population is the subgiants for which ages can be derived from differential spectroscopy. In particular, a differential precision in surface gravity of $\Delta \log g \lesssim 0.1$ can be used to determine ages to $\sim 1 \text{ Gyr}$ from isochrone fitting in the surface gravity vs. effective temperature plane. However the analysis does emphasize the potentially vast amount of information that awaits us in systematic high-resolution surveys of millions of stars.

8.6. Primary requirements of chemical tagging

But how are we to reconstruct star clusters which have long since dispersed? It will be necessary to tag individual stars to their parent cloud through unique chemical signatures shared by these stars, assuming these exist. We now discuss our basic strategy for ‘chemical tagging’ (Bland-Hawthorn & Freeman 2004; Bland-Hawthorn et al 2010a,b). High resolution spectroscopy at high signal-to-noise ratio of many stellar types reveals an extraordinarily complex pattern of spectral lines. The spectral lines carry key information on element abundances that make up the stellar atmosphere. Many of these elements cannot arise through normal stellar evolution, and therefore must reflect conditions in the progenitor cloud at the time of its formation.

Our long-term goal is to chemically tag stars into coeval groups, i.e. to identify individual members of star clusters which have long since dispersed. For unique chemical signatures to exist, there are several key requirements:

1. Most stars must be born in dense star clusters.
2. Most dense star clusters must be chemically uniform in key elements.
3. Key chemical elements must reflect the cloud composition of the progenitor cloud.

¹³ It is interesting to note what happens in the case of globular clusters. These form at a density of $\Sigma \sim 3 \text{ g cm}^{-2}$ which leads to a sixfold drop in the dynamical time and a uniformity mass limit of $\gtrsim 10^7 M_\odot$. Apart from a few light elements, globular clusters are chemically uniform, with few exceptions (e.g. ωCen ; Gratton et al 2004).

4. Key chemical elements must not be rigidly coupled (i.e. vary in lock step), and there must be sufficient abundance dispersion in key elements to allow for unique groups (reflecting unique sites of formation) to be readily identified.

5. There must be accessible spectral window(s) which contain the necessary information on key elements for chemical tagging.

We discuss each of these requirements, or conditions, in turn. Conditions 1, 2, 3 and 5 appear to be supported by observation. Condition 4 is the most uncertain largely because stellar abundance surveys to date target either too few stars or too few chemical elements.

Condition 1. Most stars are born within rich clusters of many hundreds to many thousands of stars (Clarke et al 2000; Carpenter 2000). This essential fact is supported by many studies from optical, infrared, millimeter and radio surveys.

Condition 2. The widely held view that open clusters are chemically uniform can be traced to early work on Hyades by Conti et al (1965). But, until recently, very little was known about detailed heavy element abundance work on open clusters with rigorous membership established by reliable astrometry or accurate radial velocities (< 0.5 km/s), so as to minimize 'pollution' from stars not associated with the cluster.

There have been many recent studies of open clusters which have established that open clusters are indeed highly uniform. In a detailed analysis of 55 F–K dwarfs in Hyades, De Silva et al (2007a) find that the abundance variations in Si, Ti, Na, Mg, Ca and Zn with respect to Fe, and in $[\text{Fe}/\text{H}]$, are within the measurement errors (< 0.04 dex). To date, all open clusters have been shown to be extremely chemically uniform within the errors.

Are we to expect chemical uniformity among stars in an open cluster from a theoretical perspective? This was addressed for the first time by Bland-Hawthorn et al (2010a) who derive a 'homogeneity equation' for a star cluster of a given mass which depends only on the surface density of the cluster. Their basic condition is that the gas cloud collapses before the first intercloud supernova explodes, thereby polluting the cloud. This sets a minimum collapse time of about 3 Myr. These authors find that most open clusters up to a mass of about $10^5 M_{\odot}$ are likely to be homogeneous, as will be most *globular* clusters up to a mass of about $10^7 M_{\odot}$, assuming the high mass extremes even exist. (Globular clusters are highly chemically uniform in $[\text{Fe}/\text{H}]$ but show marked anticorrelations between O, Mg and Al, Na.)

McKee & Tan (2002) propose that high-mass stars form in the cores of strongly self-gravitating and turbulent gas clouds. Two possible routes to chemical uniformity is that all stars form at the same instance from a chemically uniform cloud, or that the low mass stars form outside of the core shortly after the supernovae have uniformly enriched the cloud. The precise sequence of events which give rise to open clusters is a topic of great interest and heated debate in contemporary astrophysics (e.g. Stahler et al 2000).

Condition 3. It is generally believed that r-process elements (e.g. Sm, Eu, Gd, Tb, Dy, Ho) cannot be formed during quiescent stellar evolution. While some doubts remain, the most likely site for the r-process appears to be Type II supernova (SN II), as originally suggested by Burbidge et al (1957). Therefore, r-process elements measured from stellar atmospheres reflect conditions in the progenitor cloud. The same is believed to be true for most of the α elements since these are produced in the hydrostatic burning phase of the pre-supernova star.

In contrast, the s-process elements (e.g. Sr, Zr, Ba, Ce, La, Pb) are thought to arise from the He-burning phase of intermediate to low mass (AGB) stars ($M < 10 M_{\odot}$), although at the lowest metallicities, trace amounts are likely to arise from high mass stars (Burris et al 2000; Rauscher et al 2002).

Condition 4. During the past four decades, evidence has gradually accumulated for a large dispersion in metal abundances $[\text{X}/\text{Fe}]$ (particularly n-capture elements) in low to intermediate metallicity stars relative to solar abundances. Elements like Sr, Ba and Eu show 1 – 2 dex dispersion, although $[\alpha/\text{Fe}]$ dispersions are typically an order of magnitude smaller. Ting et al (2012) find evidence that certain element families are decoupled and provide useful axes in \mathcal{C} -space (see also Mitschang et al 2013).

These observations have been used to constrain detailed supernova models, which in turn show how different yields arise as a function of progenitor mass, progenitor metallicity, mass cut (what gets ejected compared to what falls back towards the compact central object), and detonation details. These models help to explain the smaller dispersions in α elements: the α yields are not dependent on the mass cut or details of the fallback/explosion mechanism, which leads to a smaller dispersion at low metallicity.

The large scatter of $[\text{n-capture}/\text{Fe}]$ in metal poor stars (Wallerstein et al 1997; Sneden et al 2000) means that n-capture element abundances in ultra-metal poor stars are products of one or very few prior nucleosynthesis events that occurred in the very early, poorly mixed galactic halo, a theme that has been developed by many authors (Gilroy et al 1988; Audouze & Silk 1995; Bland-Hawthorn et al 2010b).

The oversimplified schematic in Fig. 24 is illustrative of the expected abundance variation in successive generations of low mass stars in a closed box from gas enriched by successive generations of supernovae. The upper envelope is determined by enrichment from low mass supernovae, the lower envelope from high mass supernovae. From left to right, the black squares indicate the number of enrichment events, i.e. 1, 10, 100, 1000 events. Just ten enrichments from high mass supernovae are sufficient to enrich a cloud to $[\text{Fe}/\text{H}] = -2$. Note the rapid convergence in this simple model above $[\text{Fe}/\text{H}] = -1$.

In the simulation, we consider only core collapse supernovae; subsequent enrichment by Type Ia supernovae causes the converging stream to dip down above $[\text{Fe}/\text{H}] = -1$. It is not clear at what $[\text{Fe}/\text{H}]$ the r-process elements become swamped by the ubiquitous Fe-group and s-process elements. Travaglio et al (1999) suggest that the s-process does not become significant until $[\text{Fe}/\text{H}] \sim -1$ because of the need for pre-existing seed nuclei, although Pagel & Tautvaisiene (1997) argue for some s-process production at $[\text{Fe}/\text{H}] \sim -2.5$. In any event, Fig. 26 makes clear that, at sufficiently

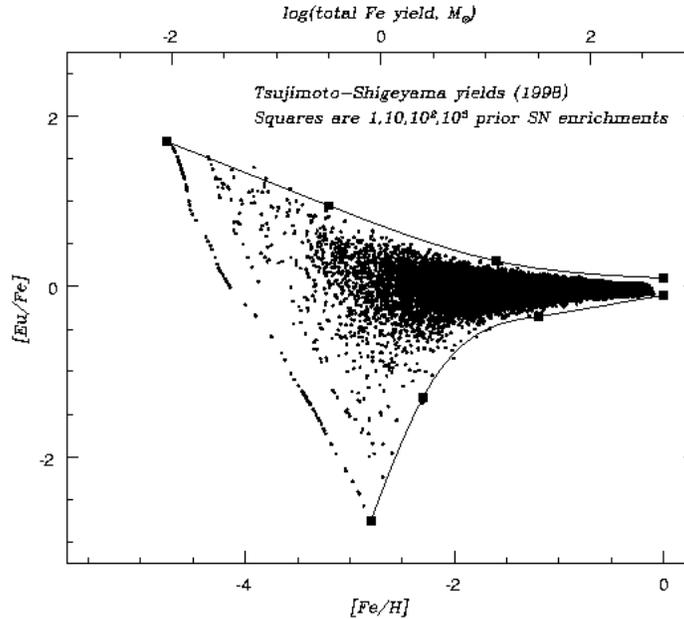


FIG. 24.— An illustration of how successive chemical enrichments by massive supernovae can lead to low mass stars with a wide scatter in chemical abundances which converge to a universal value. The model assumes a constant star formation rate in a closed box. The stars are formed according to a Salpeter mass function and the yields are taken from models by Tsujimoto & Shigeeyama (1998). Only 1% of the stars produced in the simulation are shown after the initial burst; we show all the stars at $t=0$. The contribution of Type Ia supernovae is not included for clarity.

high resolution, both the Sun and Arcturus reveal a rich network of heavy elements, many of which arise from the r-process.

An important point to appreciate is that the substantially smaller scatter in $[\alpha/\text{Fe}]$ compared to heavy element scatter (Carretta et al 2002) presumably argues that the scatter is unlikely to arise from chemical differentiation, i.e. where metal yields from a given supernova are distributed in different amounts, compared to Fe, to the surrounding ISM. In other words, the mixing is sufficiently uniform such that any given cell receives the same proportion of Eu/Fe from the supernova. This puts limits on the degree to which asymmetric supernovae must differentially enrich the surrounding gas (e.g. Maeda et al 2002).

Condition 5. Any practical experiment involving chemical signatures in millions of stars will require a wide-field, multi-fibre spectrograph operating at echelle resolutions. Since rayleigh scattering renders optical fibres rather lossy below 4000 Å, it is important to establish that there exist contiguous spectral windows which contain abundance information on a large number of chemical elements. Since it is necessary to maximize the number of object fibres at the detector, this information will need to exist within a single narrow band ($\sim \lambda/35$) echelle order. The optimal parameters for a 'galaxy genesis' machine are determined below.

Do there exist narrow band spectral windows, with chemical information on dozens of elements, which could be utilized by a fibre-fed echelle spectrograph? Until recently, this has not been possible to answer objectively. Fig. 25 shows such a window that was revealed by this study. An optimal window needs information on more than one ionization state for as many elements as possible in order to accurately determine the stellar surface gravity. Various ionic transitions in combination must also provide accurate information on the stars luminosity and surface temperature. Some regions of the spectrum are dominated by telluric features; in cool stars, molecular bands complicate the spectrum.

We concentrate our analysis on digitized atlases of the Sun ($[\text{Fe}/\text{H}]=0.0$) and for the luminous giant Arcturus ($[\text{Fe}/\text{H}]=-0.56$). Our solar line list derives from data supplied by P. Hall (NOAO) transcribed from the Moore et al (1996) solar spectrum (2935-8770Å). The atlas includes information on 25,000 absorption lines covering 63 elements, for which 28 are detected in two ionization states. The Arcturus line list comes from the digitized atlas (3570-7405Å) of Wallace et al (1998). The line list includes 7700 atomic transition for 42 elements, for which 18 are detected in two ionization states. Both lists are dominated by Fe I (45%) and light α elements (30%).

Fig. 26 shows the number of unique ions detectable as a function of wavelength for the Sun and for Arcturus. At each wavelength, the accessible window is $\Delta\lambda$ where $\lambda/\Delta\lambda=35$, such that at $\lambda=3500$ Å, the unique ion count is made over a 100 Å. The linear dependence of the window on wavelength is expected for spectral coverage in a single echelle order limited by the detector size. Note the rising count towards with decreasing wavelength down to 4000 Å, both in terms of the number of unique ions, and elements with two ionization states.

In the lower panels of Fig. 26, we demonstrate that this rise is largely due to a rising fraction of neutron capture elements; the rise in iron peak elements is not as dramatic. It is noteworthy that we obtain similar results for a solar metallicity main sequence star and a metal weak giant. In Fig. 27, we show the effect of resolving power on the fraction of lines which can be resolved as a function of wavelength. The optimal window at 4100Å requires a spectral resolving

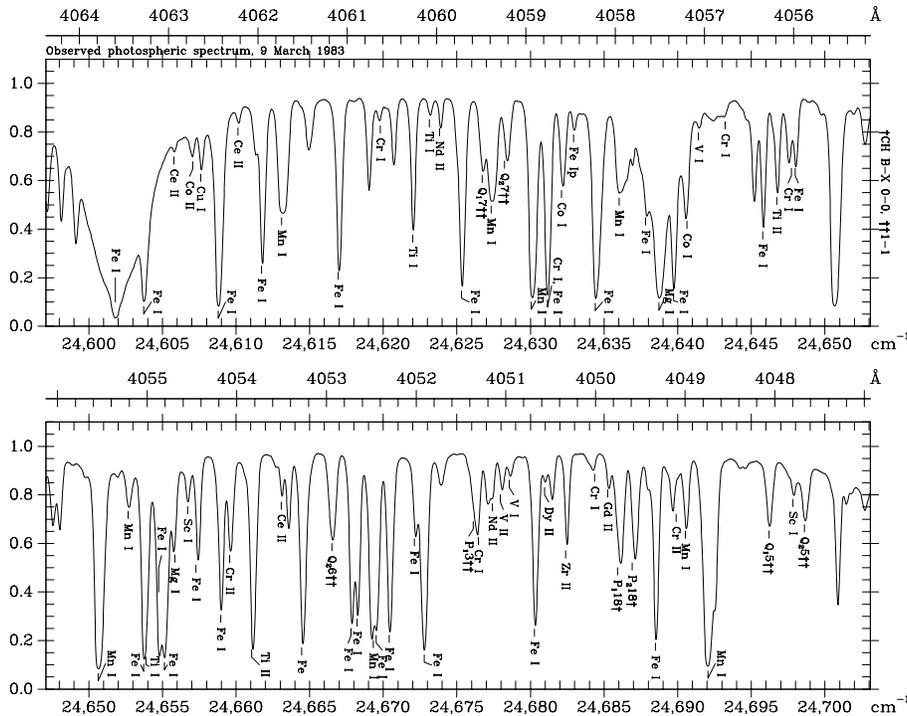


FIG. 25.— 17 Å spectral window (4047-4064Å) of the Sun obtained with the Fourier Transform Spectrograph at the National Solar Observatory revealing detailed chemistry on Fe, Cr, Ti, V, Co, Mg, Mn, Nd, Cu, Ce, Sc, Gd, Zr, Dy. The lower values are wavenumbers (cm^{-1}). The data are taken from <http://www.nso.edu/>.

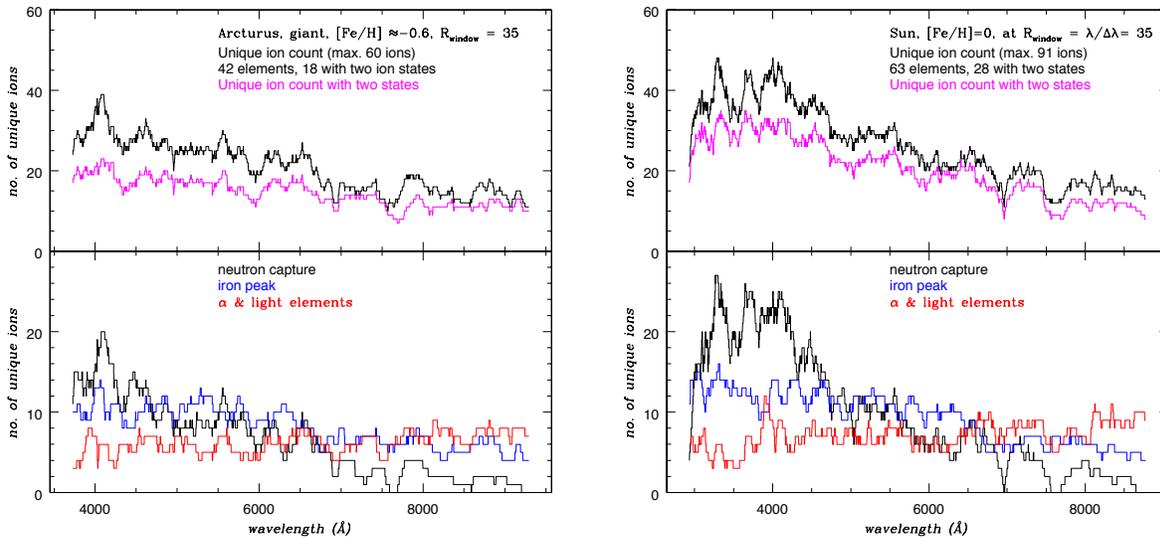


FIG. 26.— A demonstration of where most of the chemical information lies in a highly resolved spectrum of Arcturus (left) and the Sun (right). The black lines in the top plots give the total ion count in a spectral window $\Delta\lambda = \lambda/35$. The magenta lines show what fraction of this count is made up of elements with two ionization states. The lower plots separate these into element types. Note the rapid rise of heavy elements to the blue in both stars; note also that the solar spectral information extends to bluer wavelengths.

power of $R = 40,000$ in order to resolve 80 % of the lines. As the resolving power decreases, the loss of information on heavy elements is particularly dramatic.

An analysis of this kind led to the development of the first galactic archaeology machine – HERMES. This instrument (www.aao.gov.au/HERMES; Freeman & Bland-Hawthorn 2004) uses the existing 2dF positioner at the Anglo-Australian Telescope is used to obtain high-resolution spectroscopy on sufficiently bright stars ($g < 14$) over the two-degree field. The 400 fibres will allow up to a million stars to be observed over five years starting in 2013. The HERMES instrument exploits four distinct spectral windows: (blue) 4718 - 4903Å; (green) 5649 - 5873Å; (red1)

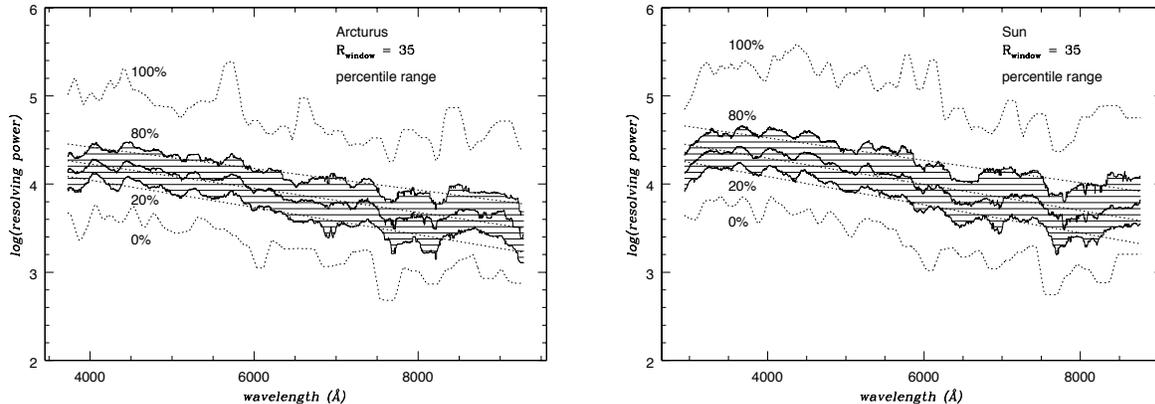


FIG. 27.— The effect of resolving power on the fraction of spectral lines which can be resolved as a function of wavelength. In order to resolve 80 % of lines at 4000 Å, we need $R = 40,000$; this falls to $R = 10,000$ at 8000 Å. Once again, both stars reveal the same general trends; note also that the solar spectral information extends to bluer wavelengths.

6481 - 6739Å; (red2) 7590 - 7890Å. Up to 30 distinct chemical elements have been identified across these windows, where the actual number depends on spectral type and luminosity class. We estimate that the HERMES data set will contain *at least* 8 distinct axes in \mathcal{C} -space.

8.7. Candidates for chemical tagging

Chemical tagging is not possible for all stars. In hot stars and young star clusters, our ability to measure abundances is reduced by the stellar rotation and lack of transitions for many ions in the optical. In very cool stars, very little light emerges in the optical or near-IR due to complex molecular and dust opacities. The ideal candidates are evolved FGK stars that are intrinsically bright, FGK subgiants and dwarfs. These populations probably account for 10% of all stars.

Giants can be observed at $R = 40,000$ over the full Gaiasphere, i.e. the 20 kpc diameter sphere centred on the Sun which will be surveyed in great detail by the GAIA satellite. Dwarf stars will only be observable within 1-2 kpc. While this is only a small fraction of the available volume, the local volume may include a representative sample of all old disk stars, regardless of their point of origin.

For many stars, certain spectral regions are dominated by thermally broadened profiles, particularly for light elements. In general, heavier atoms show narrow profiles appropriate to their mass, although odd numbered atomic isotopes are susceptible to hyperfine splitting (due to the non-zero nuclear magnetic moment) which produces a broadened asymmetric line profile.

Giants have deep, low density atmospheres that produce strong low-ionization absorption lines compared to higher gravity atmospheres. Even in the presence of significant line blending, with sufficient signal, it should be possible to derive abundance information by comparing the fine structure information with accurate stellar synthesis models. Detailed abundances of large numbers of F and G subgiants would be particularly useful, if it becomes possible to make such studies, because direct relative ages can be derived for these stars from their observed luminosities.

8.8. Short-term goal – size and structure in a multi-dimensional \mathcal{C} -space

An intriguing prospect is that reconstructed star clusters can be placed into an evolutionary sequence, i.e. a family tree, based on their chemical signatures. Let us suppose that a star cluster has accurate chemical abundances determined for a large number n of elements (including isotopes). This gives it a unique location in an n -dimensional space compared to m other star clusters within that space. We write the chemical abundance space as $\mathcal{C}(\text{Fe}/\text{H}, X_1/\text{Fe}, X_2/\text{Fe}, \dots)$ where $X_1, X_2 \dots$ are the independent chemical elements that define the space (i.e. elements whose abundances are not rigidly coupled to other elements).

Our simple picture assumes that a cloud forms with a unique chemical signature, or that shortly after the cloud collapses, one or two massive SN II enrich the cloud with unique yields which add to the existing chemical signature. The low-mass population forms with this unique chemical signature. If the star-formation efficiency is high ($\geq 30\%$), the star group stays bound although the remaining gas is blown away. If the star-formation efficiency is low, the star cluster disperses along with the gas. In a closed box model, the dispersed gas reforms a cloud at a later stage.

In the closed box model, each successive generation of supernovae produce stellar populations with progressive enrichments. These will lie along a trajectory in \mathcal{C} -space (see Fig. 8, Karlsson & Gustafsson 2001). The overall distribution of the trajectories will be affected by fundamental processes like the star formation efficiency, the star formation timescale, the mixing efficiency, the mixing timescale, and the satellite galaxy infall rate.

A critical test of chemical tagging is that stellar streams in the halo, identified from detailed phase space information, are highly localized in \mathcal{C} -space, or are confined to chemical tracks – this is a key short term goal. There may already be evidence for accreted halo stars from their distinct chemical signatures (e.g. suppressed Mg) since those with estimated

orbital parameters are found to have large apogalacticon distances (Carretta et al 2002). This may suggest that these stars originated in lower mass stellar systems with very different chemical histories from that of the Galaxy. Several moving groups have now been identified with unique chemical signatures: these include U Ma (Castro et al 1999), HR 1614 (De Silva et al 2007b), Sgr (Chou et al 2010), Wolf 630 (Bubar & King 2010) the Aquarius Stream (Wylie de Boer et al 2012) and the Argus association (De Silva et al 2013).

As we approach solar levels of metallicity in $[\text{Fe}/\text{H}]$, the vast number of trajectories will tend to converge. By $[\text{Fe}/\text{H}] \sim -2.5$, AGB stars will have substantially raised the s-process element abundances; by $[\text{Fe}/\text{H}] \sim -1$, Type Ia supernovae will have raised the Fe-group abundances. Star clusters that appear to originate at the same location in this \mathcal{C} -space may simply reflect a common formation site, i.e. the resolution limit we can expect to achieve in configuration space.

Even with a well established family tree based on chemical trajectories in the chemical \mathcal{C} -space, this information may not give a clear indication of the original location within the protocloud or Galactic component. This will come in the future from realistic baryon dissipation models. Forward evolution of any proposed model must be able to produce the observed chemical tree. The \mathcal{C} -space will provide a vast amount of information on chemical evolution history. It should be possible to detect the evolution of the cluster mass function with cosmic time (Kroupa 2002), the epoch of a starburst phase and/or associated mass ejection of metals to the halo (Renzini 2002), and/or satellite infall (Noguchi 1998). The chemical tracks could conceivably be punctuated by discontinuities due to dramatic events like galactic infall or large-scale winds (Chiappini et al 1999).

Early estimates of the size and dimensionality of \mathcal{C} -space have been made by Ting et al (2012) and Mitschang et al (2012). Roughly speaking, if we are able to measure 5 distinct abundance levels across 8 dimensions (see Sec. 8.6), then in principle we can measure 400,000 unique chemical signatures, assuming the space is uniformly filled.

As we go back in time to the formation of the disk, we approach the chemical state laid down by Pop. III stars (Karlsson et al 2012). These stars presumably had a top-heavy IMF and therefore no remnants are expected to have survived to the present day. To date, no distinctive signature of Pop. III has been clearly identified except that a large fraction of the most metal poor stars have now been shown to have a strong excess in $[\text{C}/\text{Fe}]$ (Norris et al 2007). The rarity of stars below $[\text{Fe}/\text{H}] \sim -5$ suggests that the protocloud was initially enriched by the first generation of stars (Argast et al 2000) or maybe that stars moving through the ISM today have a minimum threshold metallicity due to Bondi-Hoyle accretion. If one could unravel the abundances of heavy elements at the time of disk formation, this would greatly improve the precision of nucleo-cosmochronology. Important information is beginning to emerge from echelle observations of damped Ly α systems at high redshift (Pettini et al 2003; Cooke et al 2011).

8.9. Long-term goal – reconstructing ancient star groups from unique chemical signatures

The abundance dispersion in α and heavy elements provides a route forward for tagging groups of stars to common sites of formation. With sufficiently detailed spectral line information, it is feasible that the 'chemical tagging' will allow temporal sequencing of a large fraction of stars in a manner analogous to building a family tree through DNA sequencing. We provide a worked example below.

Consider the (extraordinary) possibility that we could put many coeval star groups back together over the entire age of the Galaxy. This would provide an accurate age for the star groups either through the color-magnitude diagram, or through association with those stars within each group that have $[\text{n-capture}/\text{Fe}] \gg 0$, and can therefore be radioactively dated. This would provide key information on the chemical evolution history for each of the main components of the Galaxy.

There is no known age-metallicity relation that operates over a useful dynamic range in age and/or metallicity. (This effect is only seen in a small subset of hot metal-rich stars). Such a relation would require the metals to be well mixed over large volumes of the ISM. For the foreseeable future, it seems that only a small fraction of stars can be dated directly (Freeman & Bland-Hawthorn 2002).

Ideally, we would like to tag a large sample of representative stars with a precise time and a precise site of formation. Can we identify the formation site? The kinematic signatures will identify which component of the Galaxy the reconstructed star group belongs, but not specifically where in the Galactic component (e.g. radius) the star group came into existence. For stars in the thin disk and bulge, the stellar kinematics will have been much affected by the bar and spiral waves; it will no longer be possible to estimate their birthplace from their kinematics. Our expectation is that the derived family tree will severely restrict the possible scenarios involved in the dissipation process. In this respect, a sufficiently detailed model may be able to locate each star group within the simulated time sequence.

Our ability to detect structure in \mathcal{C} -space depends on how precisely we can measure abundance differences between stars. It may be possible to construct a large database of differential abundances from echelle spectra, with a precision of 0.05 dex or better; differential abundances are preferred here to reduce the effects of systematic error. Differential precisions of 0.01 dex have been achieved on solar twin studies working at very high resolution and signal-to-noise (Ramirez et al 2010).

We have simulated in detail how many formation sites can be detected throughout the thin and thick disk in the Galaxy (Fig. 28). We have used the *Galaxia* model to simulate all components of the Galaxy (Sharma et al 2011; Appendix B) and have applied a realistic sampling appropriate to the HERMES survey. The results are shown for two different ICMF slopes, i.e. $\gamma = -2.5$ consistent with quiescent star formation, and $\gamma = -1.5$ consistent with high pressure activity.

The total number of signatures for the thick disk is an order of magnitude less than the thin disk, consistent with its lower mass. The impact of migration, as expected, is that many more signatures are required for strong migration,

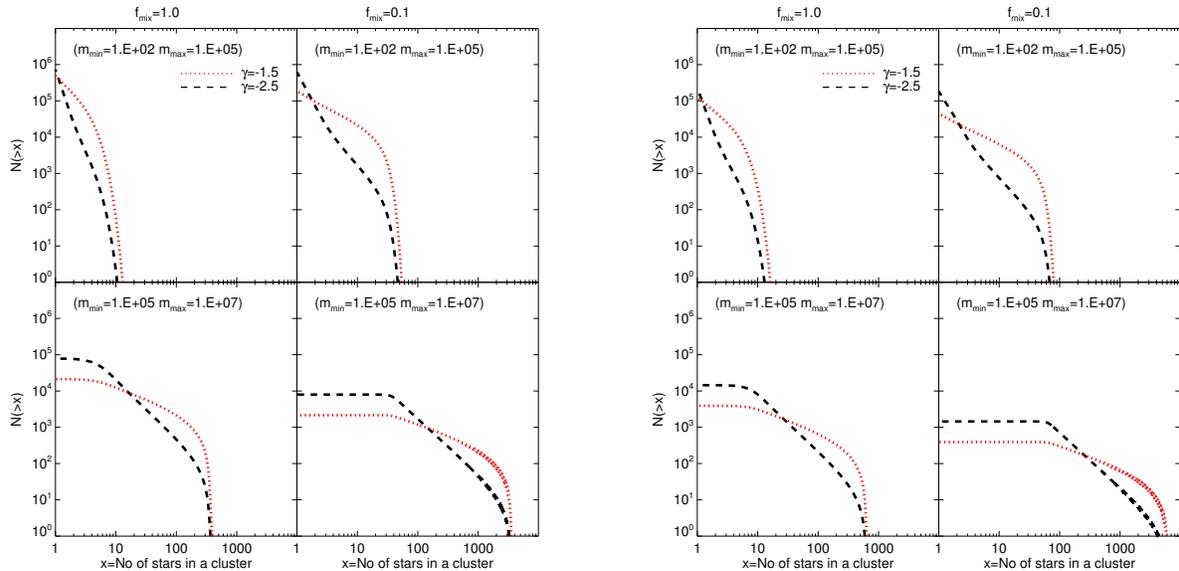


FIG. 28.— The expected number of chemical signatures needed to reconstruct stellar clusters in the thin (left) and thick (right) disk. In each sub-plot, the black dashed line refers to a cluster mass function with $\gamma = -2.5$, and the red dotted line to $\gamma = -1.5$. The x-axis is the number of reconstructed star clusters, and the y-axis is the cumulative total number of signatures required. In each panel of 4 sub-plots, the left hand figures assume that all stars are uniformly scattered throughout the disk through stellar migration; the right figures assume no migration. The top figures assume cluster mass limits of $m_c \in (10^2, 10^5) M_\odot$; the bottom figures assume $m_c \in (10^5, 10^7) M_\odot$.

and the sizes of the reconstructed clusters are characteristically smaller. The limits of the cluster mass function play an important role: high mass limits (lower figures) greatly reduces the number of required signatures.

Kroupa (2002) has argued that the thick disk formed from super star clusters as observed in the Hubble Ultra Deep Field (HUDF). In this case, a flat ICMF with high mass limits indicates that as few as 300–1000 signatures will be needed, although this range is an order of magnitude larger in the presence of strong migration.

It is an intriguing thought that one day we may be able to identify hundreds or thousands of stars throughout the Gaiasphere that were born within the same cloud as the Sun. We note from Fig. 28 that some of the reconstructed star clusters could have hundreds or even thousands of members. In this instance, assuming the reconstructed clusters have a well defined age from the Hertzsprung-Russell diagram, their spatial distribution throughout the Galaxy is a direct measure of stellar migration with cosmic time (Bland-Hawthorn et al 2010a).

9. EPILOGUE: CHALLENGES FOR THE FUTURE

9.1. The limitations of near field cosmology – are we really putting Λ CDM to the test?

In recent reviews, Silk & Mamon (2012) and Kroupa et al (2012) highlight the many challenges to Λ CDM due to the mismatch of observations and the most advanced numerical simulations. An issue that arises frequently is whether these discrepancies are a consequence of the lack of microphysics in numerical simulations or whether a new theory of gravity is justified (e.g. MOND). Λ CDM is widely cited as being extremely successful on the largest scales supported by “precision cosmology” measurements, particularly those that have come from CMB experiments. On scales smaller than galaxy groups, it becomes increasingly difficult to relate observations to the results of numerical simulations.

Λ CDM is a very useful hierarchical framework that acts over many orders of magnitude in physical scale (Davis et al 1985). In this respect, it has a lot in common with statistical theories of turbulence. But the turbulence description breaks down on the smallest scales due to microphysics. This does not lead us to argue against the usefulness of the statistical models. Rather, we simply recognize that the description breaks down on the smallest scales due to other mechanisms. Our views of Λ CDM fall along the same lines. It seems unlikely that any of the existing claims that purport to challenge Λ CDM will hold up in the decades to come as we treat the complex physical processes more carefully.

Inevitably, when we find ourselves swimming in a sea of complex data, one begins to worry about fundamental limits of knowledge. But in many applied fields, important clues to fundamental physics often emerge from complex data, particularly when complex physical processes are at work (cf. beam-line experiments in particle physics). In fields where we are far from understanding key physical processes, as in the study of galaxy formation and evolution, this effort must be worth it.

Astronomers now recognize this: every large survey has unveiled new lines of enquiry or revealed something important about our environment. We talk in terms of data mining, virtual observatories, and so forth. Furthermore, the numerical simulators push down to ever decreasing scales, and work to include new algorithms that capture an important process. At what stage do we declare that galaxy formation is basically understood? Such a declaration becomes possible when one is able to reproduce the salient features of galaxies today, in a host of different environments, with a theory that is firmly rooted (presumably) in Λ CDM. This same axiomatic theory should be able to reproduce observations of galaxies at different epochs out to high redshift, until we reach an epoch where objects no longer look

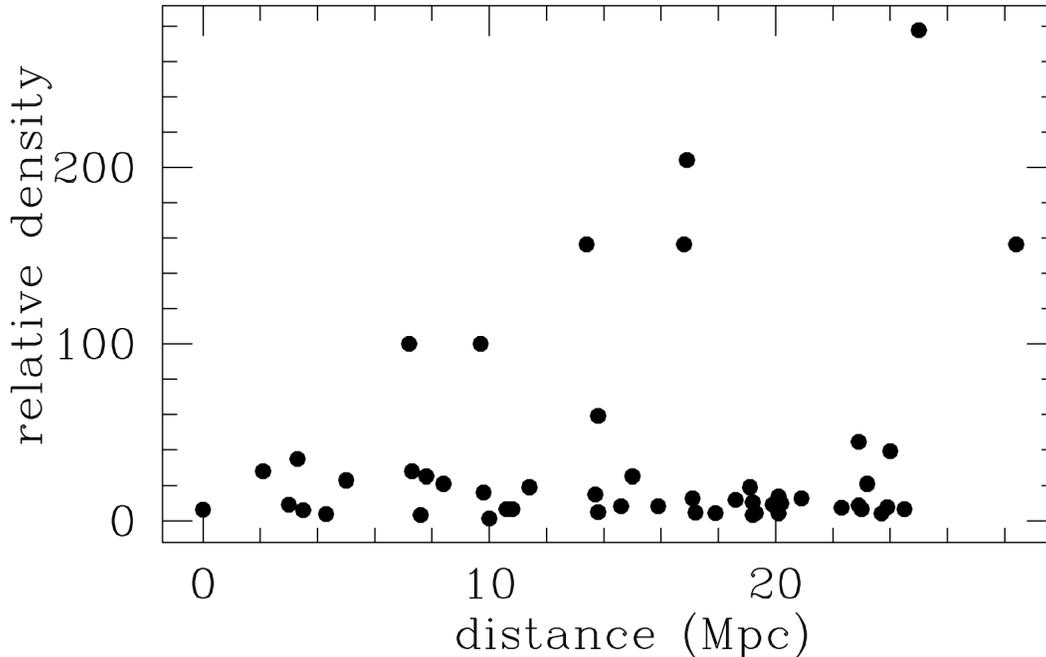


FIG. 29.— The dependence of linear group density with linear distance from a sample of 50 galaxy groups taken from Tully (1987). Most groups are much like the Local Group in their average density (Bland-Hawthorn & Freeman 2006).

like modern day galaxies. The HUDF indicates that this appears to happen at about $z \sim 2$.

Any complete theory of galaxy formation must seek to explain the properties of the Local Group and tell us if they are relatively unusual or typical of the Universe as a whole. Recent compilations from Tully’s Catalog of Galaxies (1987), and the GAMA survey (§3), suggest that “Local Group” collections of galaxies are relatively common throughout the local universe (see Fig. 29). Therefore, it would be a surprise to discover decades from now that our large-scale surveys were seriously misleading us in our quest to understand galaxy formation and evolution.

A word of caution however is that to properly survey the effects of environment (or universal overdensity), we need to study resolved stellar populations out to at least 20 Mpc (Bland-Hawthorn & Freeman 2006). In Fig. 29, over these distances, the universal density contrast increases by more than two orders of magnitude compared to the Local Group. A physical scale of 20 Mpc allows us to as to cover the full range of galaxy environments, from voids to massive groups and clusters. This is what we refer to as the Local Universe or Local Volume, now recognized by the International Astronomical Union (Division H). This sphere encloses all galaxies that have accurate distances determined from the tip-of-the-red giant (TRGB) and surface brightness fluctuation methods. We consider the Local Universe to be a useful and physically motivated working definition. Over this volume, all galaxies to be at roughly the same cosmic epoch. We have a very good picture of the 3D distribution of galaxies provided in part by Tully Catalog of Galaxies and detailed 3D flow models for all galaxies are now possible. In Fig. 30, this volume falls within the domain of the Constrained Local Universe Simulations (CLUES; <http://www.clues-project.org>). All galaxies with masses equivalent to the LMC or larger can be imaged in most wavelength bands (e.g. x-rays, infrared, radio). The most detailed and complete observations of galaxies will always come from this volume, and will therefore continue to dominate studies of physical processes in galaxies. In time, we fully expect near-field cosmological studies to extend to the Local Volume.

9.2. Future surveys

Throughout this review, we have identified fossil signatures at all cosmic epochs, with a focus on those relating to galaxy formation and evolution accessible within the Galaxy. We have raised many questions and challenges relating to Local Group archaeology throughout the text. Fossil signatures allow us to probe back to early epochs. We believe that the near-field universe has the same level of importance as the far-field universe for a comprehensive understanding of galaxy formation and evolution.

We have argued that understanding galaxy formation is primarily about understanding baryon dissipation within the Λ CDM hierarchy; to a large extent, this means understanding the formation of disks. The question we seek to address is whether this can ever be unravelled in the near or far field. Dynamical information was certainly lost at several stages of this process, but we should look for preserved signatures of the different phases of galaxy formation.

Far-field cosmology can show how the light-weighted, integrated properties of disks change with cosmic time. While light-weighted properties provide some constraint on simulations of the future, they obscure some of the key processes during dissipation. The great advantage of near field studies is the ability to derive ages and detailed abundances for individual stars within galaxies of the Local Group.

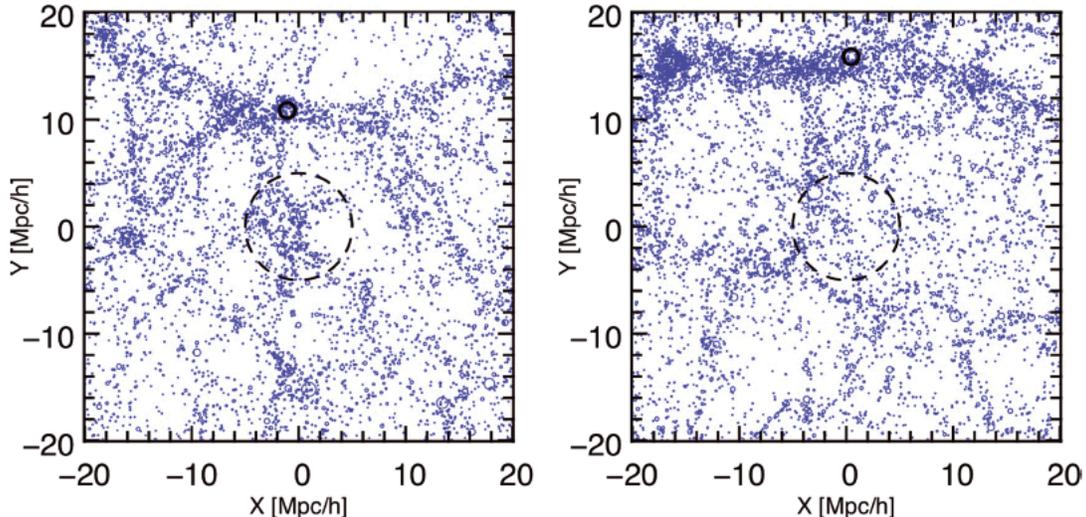


FIG. 30.— Two realizations from the Constrained Local Universe Simulations (CLUES) collaboration (Forero-Romero et al 2011). The small thick circle is identified with the Virgo Cluster. The $5 h^{-1}$ Mpc sphere associated with the Local Group is shown as a dashed circle. Only haloes with masses greater than halo mass of the LMC are shown. The slice has a depth of $25 h^{-1}$ Mpc centred at the Local Group.

It is hard to overstate the impact of ESA Gaia on the future of near-field cosmology, assuming all goes well later this year (2013). The anticipated orders-of-magnitude gain in astrometric accuracy and survey size is illustrated in Fig. 31. We have addressed the issue of information content within the Gaiasphere. The detailed information that is possible on ages, kinematics, and chemical properties for a billion stars – which we see as the limit of observational knowledge over the next two decades – may reveal vast complexity throughout the disk. We may not be able to perceive the sequence of events directly. However, we are optimistic that future dissipational models may provide unique connections with the observed complexity.

We believe that detailed high resolution abundance studies of large samples of galactic stars will be crucial for the future of fossil astronomy. Gaia will provide accurate distances, ages and space motions for a vast number of stars, separate with great precision the various Galactic components, and identify most of the substructure in the outer bulge and halo. High resolution spectrographs like VLT UVES, Subaru HDS and Keck HIRES have already begun to reveal the rich seam of information in stellar abundances. Instruments like AAT HERMES, SDSS APOGEE (and VISTA 4MOST) are now poised to undertake much larger stellar surveys with uniform chemical abundances.

It is an extraordinary fact that we can probe back to the first billion years from observations of the *local* universe. We can say with absolute certainty that stars existed at this time. These were responsible for the first chemical elements (Ryan Weber et al 2006) and for reionizing the fog of hydrogen that permeated the early Universe (Fan et al 2002). Precisely when the first star clusters formed is unknown. It seems likely, however, that gas was able to fragment at very high density even at primordial abundance levels (Clark et al 2008).

It may be possible to directly probe these environments in an era of the Atacama Large Millimetre Array and the James Webb Space Telescope. But we believe that some of the most important insights, particularly with regard to progenitor yields, will undoubtedly come from near-field cosmology. We have already made a case for the importance of dwarf galaxies, particularly the ultra faint dwarfs, in preserving ancient chemical signatures (Karlsson et al 2013). To this end, it will be necessary to equip the next generation of ELTs with wide-field multi-object spectrographs that operate at high spectroscopic resolution ($R \gtrsim 20,000$). The instrument will be expensive and technically challenging, but we believe this must be tackled if we are to ever unravel the formation of the Galaxy.

APPENDIX

A. THE DISCOVERY OF DARK MATTER IN GALAXIES

In the past year, there have been excellent reviews that revisit how we got to where we are today. Who discovered that the Universe is expanding (Peacock 2013)? Who came up with Big Bang cosmology (Belenkii 2012)? and so forth. Speakers introduce topics with a compressed narrative that tends to write out important contributions through the process of oversimplification. This is certainly the case for who discovered dark matter in galaxies.

One of the cornerstones of modern astrophysics is the cold dark matter paradigm. We know today that essentially all galaxies are encompassed by dark matter haloes that are an order of magnitude larger in diameter than what we observe with our telescopes. While the nature of "dark matter" remains a mystery, it is widely believed to be made up of one or more unidentified sub-atomic particles.

A question of more than historical interest is to ask how the existence of dark matter in galaxies was firmly established. Dark matter drives the large-scale dynamics of matter in the universe and is the scaffolding on which individual galaxies

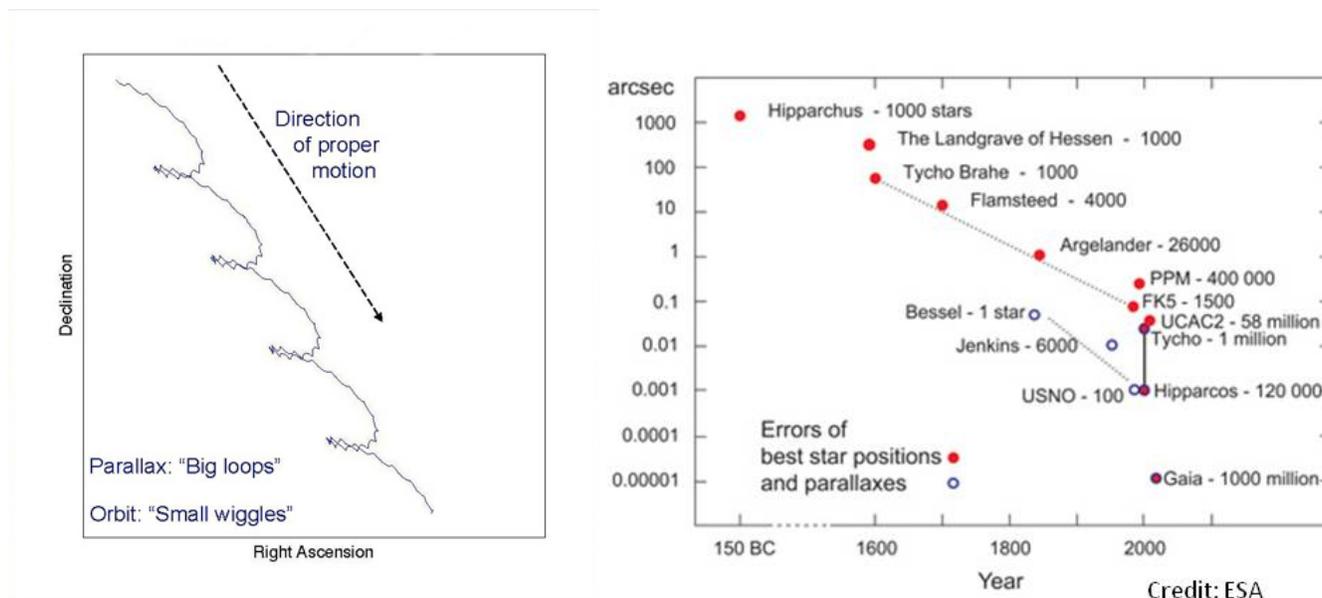


FIG. 31.— (Left) The components of proper motion, parallax and barycentric wobble in a star’s projected motion across the sky. [We acknowledge the NASA Gravity Probe B team for this figure.] (Right) The improvement in measurement accuracy of stellar positions (red dots) and parallaxes (blue dots) over the centuries. The size of each survey is also indicated. The huge expected gain in astrometric accuracy from ESA Gaia is evident. The satellite will determine proper motions for 1–2 billion stars, a few percent of all stars in the Galaxy. [We acknowledge the ESA Gaia team for this figure.]

are built. The role of dark matter in individual galaxies was finally recognized in the 1970s, as we discuss. We find that the history of this discovery is chequered with early hints, followed by definitive demonstrations, and ultimately a more complete understanding of its implications. In this respect, it is not easy to attribute the discovery of dark matter in galaxies to specific astronomers. The roll call includes Zwicky, Smith, Kahn & Woltjer for the discovery of dark matter in groups and clusters, followed by Freeman, Ostriker, Rubin & Bosma for the discovery of dark matter in galaxies. (These astronomers mostly published their results in multi-author papers, but credit is given to those that carried their arguments forward in later papers.)

The presence of some kind of “unseen matter” was first inferred by Zwicky (1933) in the Coma cluster, and by Smith (1936) in the Virgo cluster. Two decades on, Kahn & Woltjer (1959) showed that the total mass of the Local Group greatly exceeds the sum of the stellar masses of M31 and the Milky Way. In these early papers, it was not possible to establish how the unseen matter was distributed.

There were alternative interpretations that galaxy clusters were “birth sites” caught in the act of flying apart (Ambartsumian 1958; Neyman et al 1961). But the problem with these ideas is that the core of the Coma cluster, for example, is 250 kpc in diameter, its 3D velocity dispersion is 1500 km s^{-1} and hence its dispersal time is $\sim 10^8$ yrs. The stellar populations, however, are obviously much older. (At about this time, tidal arms in interacting galaxies were being argued as an ejection phenomenon, an interpretation that was dismissed for the same reasons.) On the basis of more reliable velocity data, Peebles (1970) and Rood et al (1972) both derived M/L values for Coma which exceeded those for any known stellar population.

Tremaine (1999) notes that deriving galaxy masses was a thriving industry in the 1960s, but these studies vastly underestimated the true masses due to their limited radial extent and because the connection to unseen matter in clusters had not been made. Ostriker et al (1974) and Einasto et al (1974) were some of the first to state clearly (cf. Freeman 1970; Roberts & Rots 1973) that unseen matter in clusters is probably distributed in extensive halos around individual galaxies. This made any interpretation of “dark matter” rather colder than, say, an invisible medium permeating clusters.

Ostriker in particular had fully understood the early signs from galaxian HI rotation curves that were more flat than Keplerian, and was therefore guided by this result, at least in part. It took several decades to establish the reality of super-Keplerian rotation. The effect was already clear in Babcock’s (1939) early observations confirmed decades later with more extensive observations by Roberts (1967). In hindsight, we can say that this was a necessary but not a sufficient condition because the rotation curves did not extend far enough in radius. Flat or slowly declining rotation over the inner 6 or so scale lengths is easily accounted for in terms of disk components (e.g. van Albada et al 1985). Kalnajs (1983) questioned evidence for DM for some of the optical rotation curves presented by Rubin at the Besancon conference, where he found the stellar components alone could explain the raised rotation curves.

Throughout the 1970s, few were thinking of unseen haloes in terms of elementary particles (cf. Tremaine & Gunn 1979). Essentially all papers discussed unseen haloes in terms of baryons, especially low mass stars extending down to brown dwarfs (e.g. Roberts 1975; White & Rees 1978). For example, a string of influential papers on galaxy formation considered how gas accretes and cools within a gravitating potential well (Rees & Ostriker 1977; Binney 1977; Silk

1977). Gunn (1977) and White & Rees (1978) argued that the inferred dark haloes formed naturally from the infall of collisionless dark matter onto bound structures in an expanding universe. Interestingly, at the same time, Searle & Zinn (1978) realized that the Galactic system of globular clusters must have accreted onto the halo over billions of years, rather than in the rapid collapse envisioned by Eggen, Lynden-Bell & Sandage (1962). As far as one can determine, these papers implicitly assumed that the gravitating haloes were baryonic in nature, to the extent that galaxy haloes were discussed by them.

Flat rotation curves. Flat rotation curves in individual galaxies, with their small velocity scales, provide some of the most compelling evidence we have for the existence of (cold) dark matter haloes. In her history of the discovery of dark matter, Trimble (1987) notes that “Freeman (1970) was among the first to notice that such non-Keplerian rotation curves were a widespread phenomenon and to deduce that there might be considerable gravitating mass outside the observed region.” This observation is in connection to his comment in his Appendix A that:

For NGC 300 and M33, the 21 cm data give turnover points near the photometric outer edges of these systems. These data have relatively low spatial resolution; if they are correct, then there must be in these galaxies additional matter which is undetected, either optically or at 21 cm. Its mass must be at least as large as the mass of the detected galaxy, and its distribution must be quite different from the exponential distribution which holds for the optical galaxy.

In the centennial issue of the *Astrophysical Journal*, Tremaine (1999) comments on the importance of the Ostriker et al (1974) paper but notes that “the first clear statement that rotation curves demand dark matter” is Freeman’s (1970) statement above. This point is made again in van Albada et al’s (1985) classic study of NGC 3198.

An influential paper in the same year as Freeman’s classic work was Rubin & Ford’s (1970) optical study of M31. They show the galaxy has a non-Keplerian (although declining) rotation curve to a radius of 20 kpc. No definitive conclusions were drawn from their careful study about unseen matter because their disk models were able to produce some of the slow decline in rotation over the inner disk. They state that “it does not appear possible, from the presently available data, to infer anything about the mass beyond 24 kpc in M31.”

In a published Research Note, on the basis of three extended rotation curves (M81, M101, and Rubin’s data for M31), Roberts & Rots (1973) noted that extended rotation curves indicate “a significant amount of matter at these large distances and imply that spiral galaxies are larger than found from photometric measurements.” These authors discussed the possible role of mass to light variations in galaxies and between galaxies. In a published Comment, Roberts (1976) drew attention to the striking phenomenon of flat rotation curves (see also Roberts 1975).

In the mid 1970s, there was some discussion that the few extended HI rotation curves that existed (e.g. M81, M31, M33) are possibly affected by dynamical interactions with neighbours (e.g. conference discussions by Sancisi, van der Kruit, van Albada, Allen). Another point of discussion was the contribution of poorly calibrated side lobes (Salpeter) and beam smearing arising from the Arecibo observations.

But the critical study that put the role of galaxian dark matter beyond doubt was Bosma’s (1978) PhD thesis which included 21 cm rotation curves for a large sample of galaxies, including about 20 disk galaxies to unprecedented radial scales. This thesis was widely circulated at the time and remains the 2nd most cited PhD in the history of astronomy. Bosma concludes “The mass models indicate that in the outer parts of a spiral the mass-to-light ratio is higher than in the inner parts. Perhaps a substantial fraction of the mass is not distributed in a disk at all.”

From 1978 to 1985, Rubin and collaborators wrote influential papers presenting a large sample of optical rotation curves, many of which exhibit flat rotation on radial scales comparable to Bosma’s HI data and consistent with dark matter (cf. Kalnajs 1983). Rubin considered both disk and spherical models and recognized the possible implication of massive halos around spirals to large radius, an insight that they credit to the Einasto and Ostriker papers. In the same year as Bosma’s thesis, Rubin et al (1978) state:

These results take on added importance in conjunction with the suggestion of Einasto et al (1974) and Ostriker et al (1974) that galaxies contain massive halos extending to large radius. Such models imply that the galaxy mass increases significantly with increasing radius which in turn requires that rotational velocities remain high for large radius. The observations presented here are thus a necessary but not sufficient condition for massive halos. The developments of the 1970s, in large part because of the phenomenon of flat rotation curves, paved the way for the cold dark matter paradigm in the following decade.

Today, the nature of dark matter remains unclear although one or more sub-atomic particles appear to be the most likely candidates. Examples include the hypothetical axion, neutralino, or a weakly interacting massive particle (WIMP), perhaps 100 times the mass of the proton, predicted by extensions of the Standard Model. WIMPs may be their own antiparticles and therefore capable of mutual annihilation, an effect that can be searched for with gamma ray satellites. The most recent Λ CDM simulations involving as many as 10¹⁰ particles reveal that dark haloes are likely to be highly structured and clumpy. Recent experiments with the Fermi gamma ray satellite to search for decaying dark matter (assuming a decay channel involving photons) has drawn a blank. A recent review by Abbasi et al (2012) puts stringent limits on all dark matter candidates to date.

In summary, Ostriker was arguably the most persistent champion of dark haloes in the early 1970s. While no single result was compelling, taken together, the case for dark haloes appeared to be strong, as discussed in his 1974 paper. Most but not all of the arguments have held up with the passage of time (Tremaine 1999). For example, Turner’s

(1976a,b) arguments based on binary galaxies were undermined by White (1981, 1983): the derived masses are plagued by uncertain orbits, contamination, selection effects and so on. In the same vein, Ostriker & Peebles (1973)¹⁴ wrote an influential paper on the stabilizing influence of dark haloes against bar modes. They fully understood that a disk could be stabilized by its internal velocity dispersion (e.g. Hohl 1970) but the Q parameter required is much larger than observed. Bars are not a solution either: the stable barred galaxies that they found were again much hotter than observed bar galaxies. (What could conceivably undermine their result is a rotation curve with a sharp bend near the centre, i.e. with a very small uniformly rotating core. This was demonstrated by Toomre and Sellwood in the 1980s.) Today we know that bars occur in a high fraction of disk galaxies (Eskridge et al 2000) so it seems unlikely that dark haloes are successful in stabilizing disks against bar modes.

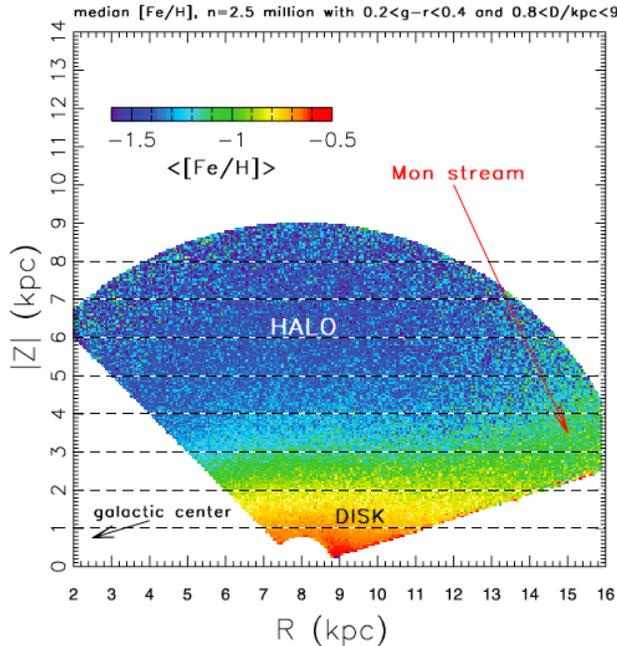


FIG. 32.— Distribution of metallicity in the Galaxy from SDSS photometry (From Ivezić et al 2008).

APPENDIX

B. STELLAR DATA – SOURCES AND TECHNIQUES

Data needed for galactic archaeology

Our goal is to use relic or fossil information to evaluate the state of the Galaxy and understand how it got to be in this state. We want to answer questions such as

- What were the major formation events ?
- How important was the role of mergers ?
- What was the infall history and the star formation history ?
- What were the major dynamical and chemical evolution processes ?

To answer these questions, we need stellar data to compare with theoretical predictions and to guide the theory. The basic stellar data that we need are the magnitudes and colours of the stars, their distances, motions and chemical properties, and their ages. We will look first at which each kind of data can do for us, and then at the techniques for acquiring the data.

¹⁴ This paper is routinely cited (incorrectly) as arguing for the existence of dark haloes but they only studied the dynamics of the Galaxy inside of the Solar Circle.

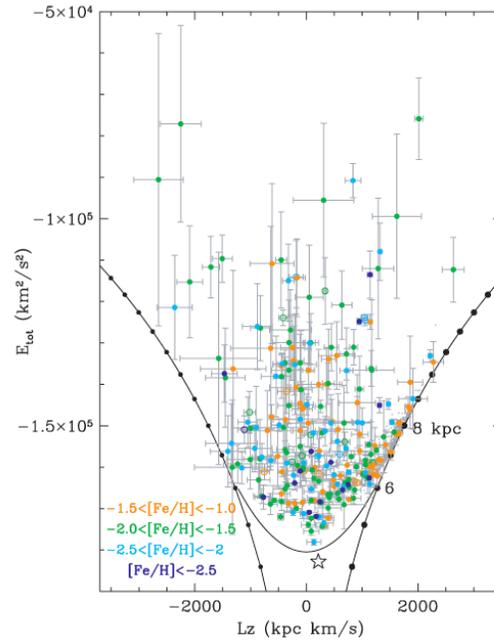


FIG. 33.— The Lindblad diagram shows orbital energy and angular momentum for Galactic disk and halo stars. Very energetic halo stars lie towards the top of the diagram; some halo stars are in retrograde orbits ($L_z < 0$). Disk stars lie near the prograde circular orbit locus (the black curve on the right). The sun is located near the “8” on the prograde circular orbit locus. The energy E and angular momentum component L_z are both integrals of the motion in a steady-state axisymmetric galaxy. (From Morrison et al 2009.)

Stellar photometry: magnitudes and colours of stars

Photometric catalogues are essential input data for stellar observational programs. They give magnitudes and colours for up to billions of stars over the whole sky or large fractions of the sky. The catalogues typically have stellar coordinates and photometry in two or more optical or near-IR bands, at different levels of accuracy.

They can be used to make preliminary estimates of stellar parameters like temperature and chemical abundance and distance for vast numbers of stars. From photometric catalogues alone, it is possible to derive useful information about the structure of the Galaxy. For example, the 5-band *ugriz* photometry from the Sloan Digital Sky Survey has been used to derive the distribution of stars and their overall chemical abundances in the Galaxy. Ivesic et al (2008) estimated photometric distances and $[\text{Fe}/\text{H}]$ abundance for 2.5 million FG stars from the SDSS to derive the distribution of abundance and of positions (R, z) out to distances of about 8 kpc from the Sun. Their map (Fig. 32) nicely shows the planar stratification of $[\text{Fe}/\text{H}]$ in the Galaxy, decreasing with height above the Galactic plane. The Monoceros concentration of relatively metal-poor stars at a Galactocentric distance of about 16 kpc can also be seen in their data.

Stellar distances: where do stars lie

We need stellar distances to

- measure transverse velocities of stars from their proper motions
- map substructure in the halo and disk
- calibrate the luminosities of different kinds of stars
- measure the structure and dynamics of the Galactic components

Stellar Motions

The basic data are the stellar radial velocities and proper motions. With an estimate of distance, it is possible to calculate the 3D space motions. The motions of stars are used to

- measure how energetic the orbits of particular kinds of stars are, and how far they are from circular motion
- measure the sense of their angular momentum: prograde or retrograde
- see how stellar orbits have evolved: how do the orbital properties correlate with age and metallicity
- detect kinematic substructure, such as moving stellar groups

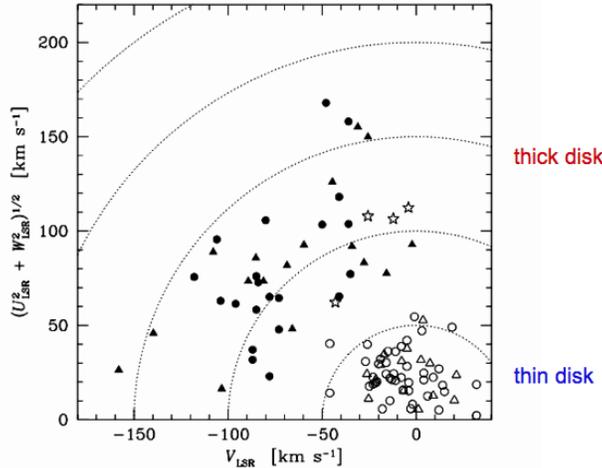


FIG. 34.— The Toomre diagram for nearby stars of the thin (open symbols) and thick (filled symbols) disk. The V component of the stellar velocity represents the stellar angular momentum and has an asymmetric distribution with a negative mean: this asymmetric drift increases with velocity dispersion and is a useful diagnostic. The U and W components have more or less symmetric distributions about zero mean, so their combination is a measure of the orbital energy. (From Bensby et al 2005)

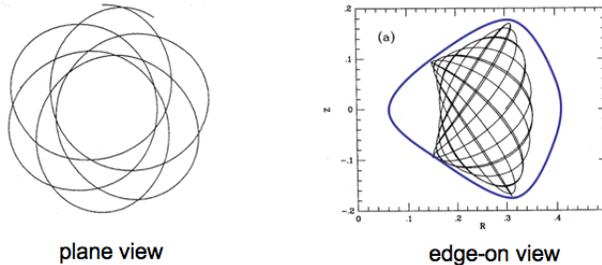


FIG. 35.— A typical stellar orbit in an axisymmetric potential (adapted from Binney & Tremaine 2008).

- compute stellar orbits to learn about the dynamical structure of the Galaxy.
- measure the properties of the Galactic potential from their spatial and kinematic distributions. In this way, the dark matter content of the Galaxy can be measured from the kinematics and distribution of distant halo stars, and the total density of luminous and dark matter near the sun can be estimated from the properties of nearer disk stars.

Fig. 33 shows the Lindblad diagram (orbital energy against angular momentum) for a sample of halo and disk stars. This diagram can be constructed for stars with known motions, and is a very useful diagnostic of the dynamical properties of different stellar populations. It is immediately clear if stars are in near-circular orbits, retrograde orbits, or highly energetic orbits. The Toomre diagram (Fig. 34) is another diagram derived from stellar kinematics and is much used to identify to which Galactic component (thin disk, thick disk, halo) a star belongs. Here (U, V, W) are components of the star's motion relative to the Local Standard of Rest: U is in the Galactic radial direction (towards $l = 0^\circ$ or 180° depending on the convention), V in the direction of rotation (towards $l = 90^\circ$) and W in the vertical direction towards the North Galactic Pole. Stars of the thin disk are clustered towards low velocities, while the stars of the thick disk are hotter kinematically.

If we know the 3D location and velocity for a star, and have a reliable model for the Galactic gravitational potential, then we can compute the Galactic orbit of the star. Fig. 35 shows the plane and edge-on view of a typical stellar orbit. Stars from the inner and outer Galaxy can pass through the solar neighborhood. Knowing the orbit is not always very useful for Galactic archaeology. There is no guarantee that the dynamical properties of a star have remained unchanged throughout the life of the star. The Galactic potential has evolved as the Galaxy's mass has gradually increased by accretion of baryonic and dark matter. Stellar orbits can be disturbed as the star interacts with spiral structure and giant molecular clouds, and resonances with the central bar and the spiral structure can flip a star from one near-circular orbit to another.

Stellar element abundances

The cosmic abundance distribution in Fig. 36 shows the outcome (in the solar neighborhood) of the chemical evolution of our Galaxy. The element abundances of stars come initially from the abundances in the gas from which they formed. This gas has been enriched by previous generations of evolving and dying stars. Different element groups come from different progenitors.

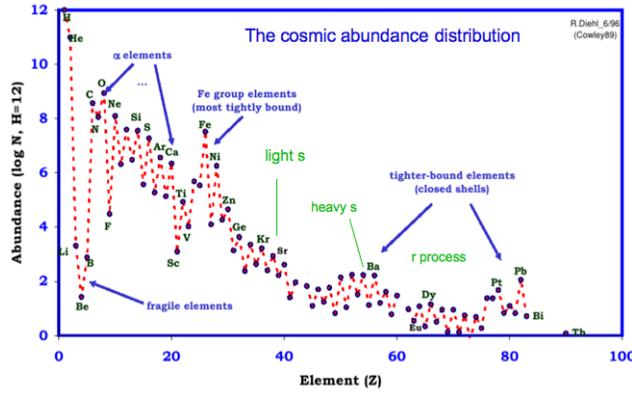


FIG. 36.— The cosmic abundance distribution, showing structure due to the major chemical evolution processes.

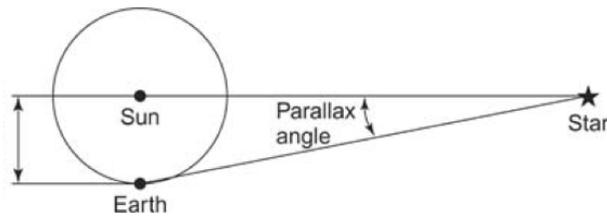


FIG. 37.— Geometry of parallax measurement.

- the Fe-peak elements come mainly from type Ia SN
- the α -elements (e.g. Mg, Si, Ca, Ti) and r-process elements come mainly from the more massive type II SN
- the s-process elements come mainly from thermally pulsing AGB stars

For most of the heavier elements, stars remember the abundances with which they are born. The abundances of different element groups in stars can tell us a lot about the star formation history which led to the formation of these particular stars. For example, α -enrichment relative to Fe indicates that SNI were important for the chemical evolution and the star formation history was fairly rapid, on a timescale of order 1 Gyr: Fe-enrichment from SNIa (which take ~ 1 Gyr to evolve) was less prominent.

Different components of the Galaxy (halo, bulge, thick disk, thin disk) each have different characteristic chemical properties. For example, the halo stars are mostly metal poor ($-1 > [\text{Fe}/\text{H}] > -5$), while the thick disk stars are more metal rich ($-0.2 > [\text{Fe}/\text{H}] > -2$). The halo and thick disk stars are both enriched in α -elements (Mg, Si, Ca, Ti). The $[\text{Fe}/\text{H}]$ range of the thin and thick disks overlap; the thick disk stars have higher $[\alpha/\text{Fe}]$ ratios than thin disk stars of the with similar $[\text{Fe}/\text{H}]$ abundances, indicating that the chemical evolution of the thick disk proceeded more rapidly.

Groups of stars born together, like open star clusters, usually have almost identical abundances, reflecting the abundances of the gas from which they formed (e.g. De Silva et al 2009). This is true also for most of the globular clusters: a few have heavy element variations from star to star, and many show correlated variations of lighter elements such as C, N, O, Na, Mg, Al, the origin of which are not full understood yet. Chemical signatures may allow us to recognize groups of stars which were born together but have dispersed and drifted apart. This technique is known as chemical tagging and will be used for analysing the products of large high resolution spectroscopic surveys like the HERMES survey on the AAT and the APOGEE near-IR survey with the Sloan telescope. Although it is readily possible to measure abundances of more than 30 elements, these elements do not all vary independently. The number of independently varying elements is 8 – 9 (Ting et al 2011): this is the dimensionality of the space defined by the abundances of the chemical elements.

Stellar ages

Stellar ages let us evaluate when events occurred in the evolution of the Galaxy. They are important for measuring the star formation history and for understanding how the metallicity and dynamics of different groups of stars have evolved. For example, how has the star formation rate in the disk of the Galaxy evolved since the disk began to form? The star formation rate is believed to have been roughly constant over time near the sun, with episodic star bursts over the past 10 Gyr (e.g. Rocha-Pinto et al 2000) but this remains uncertain because of uncertain ages. How have the kinematics and the metallicity of the thin disk near the sun changed from 10 Gyr ago to the present time under the effects of dynamical and chemical evolution? Again, the kinematics and metallicities are not difficult to measure, but the derivation of stellar ages remains problematic. Because stellar ages are still difficult to measure, there remains

much uncertainty about the evolution of the Galaxy. Measuring stellar ages is one of the most important goals for Galactic evolution for the future.

Sources of data

Now we turn to the sources of the various kinds of data needed for Galactic archaeology.

Photometric catalogs

Here is an incomplete list of some of the major photometric catalogs.

2MASS is a relatively shallow all-sky near-IR (JHK) survey and includes a point source catalog. This catalog is notable for its excellent astrometry and is an invaluable source of stars for brighter spectroscopic surveys. The UKIDSS survey (*YJHK*) is a deeper survey of 7500 square degrees of northern sky. The VISTA VHS survey (JK) is still in progress and will also go significantly deeper.

The Sloan Digital Sky Survey (SDSS) covers about 8000 square degrees, mainly in the northern sky. It uses a 5-filter system (*ugriz*) which has become a standard system. The Pan-STARRS project will survey about 30,000 square degrees, again mainly in the northern sky, using *grizy* filters. In the future, the LSST will provide a very deep survey of the whole southern sky with *ugrizy* filters. In the meantime, the SkyMapper survey will cover the southern sky several times, using a six-filter system (*uvgriz*), reaching somewhat deeper than the SDSS.

Techniques for measuring distances

Trigonometric parallaxes. The stellar distance scale (and the extragalactic distance scale) are based on fundamental trigonometric parallaxes. Stellar positions are measured from the extremes of the earth's orbit around the sun (see Fig. 37). Because this technique is used for relatively nearby stars, it is usually necessary to continue the observation for several seasons, in order to separate out the shifts in position due to parallax from the shifts due to the star's transverse or proper motion. From the ground, parallax errors of a few milli-arcseconds (mas) can be achieved, giving distances with 10% errors out to about 30 pc from the sun.

From space, higher precision is possible. The Hipparcos mission provided parallaxes with errors of about 1 mas. The Gaia mission, due for launch in 2013, will give parallax errors of about $10 \mu\text{as}$ at a *V*-magnitude of about 14. Parallaxes with this level of accuracy will be useful out to distances of about 10 kpc. For turnoff stars with $V \sim 14$, the distances will be known to within about 1%. The Gaia parallax errors are larger for fainter stars.

Photometric and spectroscopic parallaxes. Theoretical or empirical isochrones are used to estimate the absolute magnitude and hence the distance of the star. To make this work well, one needs to know the abundance [Fe/H] of the star, and an estimate of its effective temperature and surface gravity or luminosity. These can all be derived from multi-filter photometry or spectroscopy. An assumption of the star's age is needed if the star has evolved from the main sequence.

The more one knows, the better this works. If only broad-band colors are known, the distance errors can be very large. In the best cases, when the errors are ~ 0.1 in [Fe/H], 100 K in temperature and 0.1 in $\log g$, distance errors of about 15% can be achieved. For stars on the steep giant branch, the distance estimates are usually less accurate.

For some kinds of stars, like the He-core burning RR Lyrae stars, blue horizontal branch stars and red giant branch clump (RGBC) stars, accurate absolute magnitudes are known or can be estimated from periods and colors, and the errors in their photometric distances can be $< 10\%$. RR Lyr and BHB stars are usually found in more metal-poor populations like the Galactic halo and thick disk. RGBC stars are particularly useful for studies of galactic structure in the more metal-rich populations because they are so common and it is easy to measure their element abundances.

For star clusters, one can fit theoretical isochrones to their color-magnitude diagrams to derive ages and distances if their metallicities are known. Globular clusters mostly have horizontal branch stars or RR Lyr stars from which accurate distances can be derived.

Interstellar reddening and extinction are a problem for photometric parallaxes. This problem is less significant in the near-infrared. Multicolor photometry can give independent stellar reddening estimates if the wavelength dependence of reddening is known. For individual stars, the diffuse interstellar bands can be used to estimate the reddening directly (e.g. Munari et al 2008). The reddening can also be derived from reddening maps of the Galaxy, like the Schlegel et al (1998) maps derived from the COBE/DIRBE near-IR mapping of the Milky Way. These maps give the total reddening along each line of sight. For stars that are located within the reddening layer, correction is needed via models of the reddening distribution along the line of sight.

Techniques for measuring stellar velocities

Radial (line of sight) velocities are measured spectroscopically via the Doppler shift. Proper (transverse) motions are measured astrometrically from the shift of a star's position with time relative to a set of reference stars, and are usually measured from wide field photographic or CCD images.

Some large stellar radial velocity surveys are in progress for Doppler planet searches and for Galactic structure and dynamics. The fiber spectrograph surveys SEGUE, and RAVE surveys are observing several $\times 10^5$ stars for Galactic structure and chemical evolution. The LAMOST survey is one or two orders of magnitude larger. Radial velocities have typical accuracies ranging from about 1 m s^{-1} with special techniques at spectroscopic resolution $R = \lambda/\Delta\lambda \sim 50,000$ to 1 km s^{-1} at $R \sim 7000$ (RAVE) and 5 km s^{-1} at $R \sim 2000$ (SEGUE, LAMOST). For most Galactic programs, 5 km s^{-1} is good enough. For some programs, like finding the kinematically cold debris of tidally disrupting star clusters,

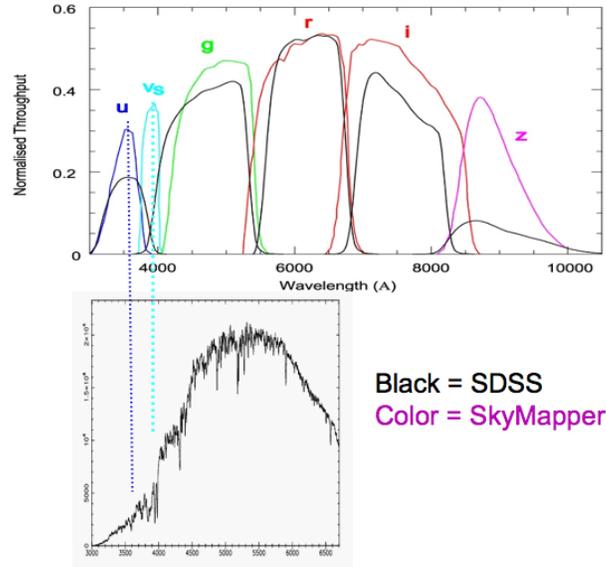


FIG. 38.— The filter passbands of the SkyMapper and SDSS filter system are aligned with the spectrum of a cooler star to show how the passbands are located relative to features in the spectrum of the star.

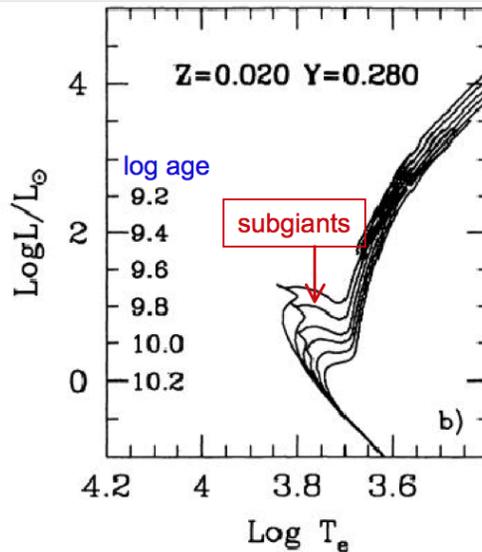


FIG. 39.— Isochrones from Bertelli (1994) for stars of near-solar abundance and a range of ages from 1.6 to 16 Gyr. The isochrones can overlap near the turnoff are well separated in the subgiant region.

an accuracy of 1 km s^{-1} can be very useful. Spectra acquired for radial velocities can also give useful estimates of the stellar parameters T_{eff} , $\log g$ and $[\text{Fe}/\text{H}]$.

From the ground, the proper motion accuracy can be a few mas yr^{-1} ($\sim 20 \text{ km s}^{-1}$ at a distance of 1 kpc). Very large samples of proper motions (10^5 to 10^9 stars) come from many ground-based surveys, such as USNO, UCAC2, SPM, SDSS, 2MASS, GSC, PM2000, PPMXL ... with more to come from the large imaging surveys Pan-STARRS, SkyMapper and LSST.

From space, the Hipparcos/Tycho mission provided proper motions of about 2 million stars, with an accuracy of about 1 mas yr^{-1} . The Gaia mission will give proper motions for about a billion stars: the accuracy depends on the brightness and color of the star, and is about $10 \mu\text{as yr}^{-1}$ at $V = 14$ (i.e about 0.7 km s^{-1} for a bright giant at a distance of 15 kpc). Gaia will really change Galactic astrophysics, with vast numbers of very precise parallaxes and proper motions. We should be prepared to get the most from this resource (launch is due in 2013). The JASMINE missions are smaller near-IR astrometric space projects, aimed particularly at the Galactic plane and bulge.

Techniques for measuring chemical abundances

Intermediate and broad band photometry, like the Strömgren photometry and the SDSS/SkyMapper photometry, can give estimates of stellar temperature, gravity and metallicity. The typical metallicity errors are about 0.2 dex but can be smaller. Fig. 38 shows how the passbands of the $uvgr$ filters of the SkyMapper and SDSS systems align with

features in FGK stars. Examples of photometric surveys that have generated large samples of stellar abundances are the Geneva-Copenhagen survey (GCS) of about 14,000 nearby FG dwarfs using Strömgen photometry complemented with Hipparcos astrometry and precise groundbased radial velocities (Nordström et al 2004), and the large study of the Galactic abundance distribution by Ivesic et al (2008) using SDSS data.

Medium resolution spectroscopy ($R \sim 2000 - 10,000$) uses the strengths of spectral features to estimate stellar parameters, including $[\text{Fe}/\text{H}]$ and sometimes $[\alpha/\text{Fe}]$ and a few other elements. Recent examples are the RAVE and SEGUE surveys of several $\times 10^5$ stars. Various techniques are used to measure the stellar parameters, including empirical calibration of individual line strengths and χ^2 matching of the spectra to grids of synthetic spectra. The $[\text{Fe}/\text{H}]$ errors are typically about 0.15 dex. Several of the medium resolution spectroscopy facilities can acquire spectra of many stars at once. For example, the AAOmega spectrometer on the AAT is fed by optical fibers and can measure medium resolution spectra of up to about 400 stars simultaneously in a 2-degree diameter field.

High resolution spectroscopy ($R > 20,000$) allows measurement of abundances for many elements, including the important neutron-capture elements. Many of these elements have relatively weak lines which are difficult to measure at lower resolutions. It is also possible to measure isotopic abundances for some elements, and these can be good diagnostics of nuclear processes. From high resolution spectra with high signal-to-noise ratios, it is possible to measure differential abundances with an error as low as 0.02 dex in the element abundance ratios $[\text{X}/\text{Fe}]$ when comparing stars of similar temperatures and gravities. Techniques include analysis of equivalent widths of individual spectral lines and matching synthetic spectra in detail to the observed spectra. The newly developed MATISSE method (Recio-Blanco et al 2006) involves projection of the spectra onto basis vectors to derive the individual parameters.

High resolution spectrographs are usually echelle spectrographs, some with a few hundred fibers for multi-object capability. Existing systems include Hectochelle on the MMT, MIKE and MIKE-fibers on Magellan, and the FLAMES/GIRAFFE/UVES system on the VLT. HERMES on AAT and APOGEE on the Sloan telescope are coming soon and will be used for large high-resolution spectroscopic surveys of 10^5 to 10^6 stars at resolutions of 20,000 (APOGEE, in the NIR H-band) to 30,000 (HERMES in optical bands). HERMES will also have a $R = 48,000$ multi-object capability. The analysis of high resolution spectra is currently laborious, but this will change as pipelines are developed for the coming high-resolution surveys of large samples of stars.

Currently available compilations of high resolution abundance data for nearby stars include those by Venn et al (2004: 781 stars) and Soubiran & Girard (2005: 743 stars).

For large optical high-resolution surveys, FGK stars with $T_{\text{eff}} = 4000 - 6500$ K are usually selected. These stars are cool enough to have plenty of lines and warm enough for analysis to be relatively straightforward. Hotter stars have mostly weaker metallic lines and are often younger and rotating which broadens the lines and makes weak lines difficult to measure. Cooler stars have complex atmospheres with molecules: they are more difficult to analyse from optical spectra, and are better studied in the near-IR, which is also much less affected by interstellar extinction. The infrared is not so good for neutron capture elements.

This is a very brief overview of high resolution spectroscopy and has not attempted to address the many issues involved in abundance analysis related to the physics of energy transfer in stellar atmospheres. The study of hotter and cooler stars are each major specialties in astrophysics.

Techniques for measuring stellar ages

Measuring ages for individual stars remains difficult; see Soderblom (2010) for a comprehensive review.

Nuclear cosmochronology is a fundamental technique for estimating ages: the observed ratios of radioactive and stable species (e.g ratio of Th to U) are compared to the expected production ratios from nucleosynthesis theory. The technique has been used on a few stars. The expected production ratios from theory are somewhat uncertain.

Asteroseismology uses the spectrum of stellar oscillations to estimate the stellar age. The frequency spectrum depends on the structure of the stellar interior, which changes as the star ages. Accurate photometry at the μm level is needed and is usually done from space. The asteroseismology space missions (CoRoT, Kepler) promise to derive stellar ages with 5–10% errors. The technique works for main sequence and giant stars for which it is otherwise difficult to determine reliable ages.

Stellar activity and rotation are useful estimators of stellar age. Stars spin down as they age, with rotation periods that typically increase roughly as $(\text{age})^{1/2}$. The rotation periods are measured photometrically or spectroscopically, and the relation between rotation period and age is calibrated empirically on star clusters and the Sun (see Barnes 2007). Chromospheric activity (usually measured from Ca K emission) is associated with rotation and decreases with age. It was much used in the past but is believed to be less accurate for older stars. Fig. 17 compares rotational (gyro) ages and chromospheric ages for a sample of well studied stars.

If the stellar temperature, metallicity and luminosity (or surface gravity) are known, it is possible to estimate the age of a star from isochrones. In its most direct form, the trigonometric parallax gives the absolute magnitude and hence the stellar luminosity L after bolometric correction. Photometry or spectroscopy gives the temperature T_{eff} and the metallicity. One then compares the location of the star in the $L - T_{\text{eff}}$ plane with theoretical isochrones for the appropriate metallicity. This works if the star is in a region of the $L - T_{\text{eff}}$ plane where the luminosity and temperature depend on age, so it does not work on the unevolved main sequence and not well on the upper giant branch. It can be used for evolved stars which are still close to the main sequence, but accurate temperatures are needed. As seen in Fig. 39, one needs to beware of regions near the turnoff where the isochrones cross and the estimated ages can be multi-valued. Subgiants are well suited to isochrone aging because the isochrones are well separated on the

subgiant branch: see Fig. 39. The Gaia mission will provide a huge increase in the numbers of stars for which accurate luminosities are known, and for which accurate isochrone ages can be derived. Recall that the Gaia distance errors will be about 1% at $V = 14$.

If parallaxes are not available, isochrone ages can be measured by using the isochrones in the surface gravity - temperature plane. The gravity g is related to the luminosity, mass and temperature by $L = 4\pi GMaT_{\text{eff}}^4/g$ where a is the Stefan-Boltzmann constant and M the stellar mass. The gravity can be measured spectroscopically with a typical error in $\log g$ of about 0.3 for medium resolution spectra and 0.1 – 0.2 for high resolution spectroscopy. Age errors of about 25% can be achieved in this way.

Large biases can occur in isochrone ages if the errors in L or $\log g$, temperature and abundance are significant. These occur because the underlying distributions of stellar mass and abundance are not uniform. They reflect the stellar initial mass function, Galactic density distribution and metallicity distribution function. Bayesian techniques can include these underlying mass distributions as priors: see Pont & Eyer (2004).

For populations of stars, such as white dwarfs of the disk, and for clusters of stars, other techniques can be used. The luminosity function of a population of white dwarfs can be used to estimate its age; as the white dwarfs cool and fade, their luminosity function evolves in a predictable way. At the faint end, the white dwarf luminosity function drops rapidly, and the luminosity at which it drops is a measure of the age of the white dwarf population. For example, Leggett et al (1998) use the luminosity function of white dwarfs in the disk to estimate that the oldest disk stars are about 9 Gyr old. For open and globular clusters, the color-magnitude diagram defines an empirical isochrone which can be dated by comparison with theoretical isochrones of the same metallicity.

Sources of models

In this section we will briefly discuss stellar atmosphere models, theoretical isochrones and Galactic models which are important tools for Galactic archaeology.

Stellar atmosphere models

Model atmospheres and synthetic spectra are used for deriving element abundances and other stellar parameters from medium and high resolution spectra by comparison with observed spectra. The most widely used synthetic spectra come either from the Kurucz models (e.g. Munari et al 2005) or the MARCS models (marcs.astro.uu.se). The models available at this time are one-dimensional local thermodynamic equilibrium (LTE) models, giving flux *vs* wavelength at various spectral resolutions for a wide range of stellar parameters. For example, the RAVE pipeline works by fitting the observed spectra ($R \sim 7000$, $\text{SNR} \sim 40$) in the Ca triplet region (840 – 880 nm) to a grid of such models. The internal accuracy of these fits is about 0.1 in $[M/H]$, 0.2 in $\log g$ and 135 K in T_{eff} .

Isochrone models

Isochrones are derived from stellar evolution models. Some of the widely used isochrone libraries come from the Padova, Dartmouth, Victoria-Regina, BaSTI, Geneva and Yale-Yonsei models. They can be found on the www and give isochrones of chosen age metallicity and sometimes $[\alpha/\text{Fe}]$, typically tabulating the stellar mass, bolometric luminosity, temperature, gravity, absolute magnitude in various photometric systems, and sometimes useful derived quantities such as the number of stars per solar mass of system at each step along the isochrone.

Galactic models

Galactic models are constructed to represent the stellar density distribution, kinematics and reddening of the various components of the Galaxy. They are used to simulate observations such as star counts and kinematics and are very useful for planning large surveys. Widely used examples include the **Besançon** model (Robin et al 2003), the **TRILEGAL** model (Girardi et al 2005) and the **Galaxia** model (Sharma et al 2011). Stellar evolution models are part of these Galactic models.

Generating a synthetic catalog of stars in accordance with a given model of galaxy formation has a number of uses. First, it helps to interpret the observational data. Secondly, it can be used to test the theories upon which the models are based. Moreover synthetic catalogs can be used to test the capabilities of different instruments, check for systematics and devise strategies to reduce measurement errors. This is well understood by the architects of galaxy redshift surveys who rely heavily on Λ CDM simulations to remove artefacts imposed by the observing strategy (e.g. Colless et al 2001).

Given the widespread use of synthetic catalogs, a need for faster and accurate methods to generate such synthetic catalogs has recently arisen due to the advent of large scale surveys in astronomy, e.g., future surveys like LSST and GAIA have plans to measure over 1 billion stars. In order to generate a synthetic catalog, one first needs to have a model of the Milky Way. While we are far from a dynamically consistent model, a working framework is fundamental to progress. Inevitably, this will require approximations or assumptions that may not be mutually consistent. Cosmologists already accept such compromises when they relate the observed galaxies to the dark-matter test particles that emerge from cosmological simulations.

There have been various attempts over the past few decades to create a Galaxy model that is constrained by observations. The earliest such attempt was by Bahcall & Soneira (1984) where they assumed an exponential disc with magnitude dependent scale heights. An evolutionary model using population synthesis techniques was presented by Robin & Creze (1986). Given a star formation rate (SFR) and an initial mass function (IMF), one calculates the

resulting stellar populations using theoretical evolutionary tracks. Local observations were then used to constrain the SFR and IMF. Bienayme et al (1987) later introduced dynamical self consistency to constrain the disc scale height (cf. Bertelli et al 1995).

The present state of the art is described in Robin et al (2003) and is known as the **Besançon** model. Here the disc is constructed from a set of isothermal populations that are assumed to be in equilibrium. Analytic functions for density distributions, the age/metallicity relation and the IMF are provided for each population. A similar scheme is also used by the photometric code **Trilegal** by Girardi et al (2005).

In spite of its popularity, the current **Besançon** model has important shortcomings. A web interface exists to generate synthetic catalogs from the model but it has limited applicability for generating wide area surveys and the output is not drawn correctly from a statistical distribution. Discrete step sizes for radial, and angular coordinates-ordinates need to be specified by the user and results might differ depending upon the chosen step size. The scale height and the velocity dispersion of the disc are in reality a function of age but, due to computational complexity, the disc is modeled as a finite set of isothermal discs of different ages. Increasing the number of discs enhances the smoothness of the model but at the price of computational cost (Girardi et al 2005).

In addition to the disc components, one also needs a model of the stellar halo. Under the hierarchical structure formation paradigm, the stellar halo is thought to have been produced by numerous accretion events and signatures for which should be visible as substructures in the stellar halo. Missions like GAIA, LSST and PanSTARRS are being planned which will enable us to detect substructures in the stellar halo.

A smooth analytic stellar halo as in the **Besançon** model is inadequate for testing schemes of substructure detection. Furthermore, such a halo does not accommodate known structures like the Sagittarius dwarf stream which may constitute a large fraction of the present halo (Ibata et al 1995; Chou et al 2010). Substructures have complex shapes and hence to model them we cannot use the approach of analytic density distributions as discussed earlier. However, N-body models are ideally suited for this task. Brown et al (2005) attempted to combine a smooth galaxy model with some simulated N-body models of disrupting satellites, but the stellar halo was not simulated in a proper cosmological context.

Recently, using hybrid N-body techniques, Bullock & Johnston (2005) have produced high resolution N-body models of the stellar halo which are simulated within a cosmological context; see also Cooper et al (2009) and De Lucia & Helmi (2008) for a similar approach. These can be used to make accurate predictions of the substructures in the stellar halo and also test the Λ CDM paradigm. However, as highlighted by Brown et al (2005) there are several unresolved issues related to sampling of an N-body model which has prevented their widespread use.

The new **Galaxia** code (Sharma et al 2011) allows for fast and accurate methods to convert analytic and N-body models of a galaxy into a synthetic catalog of stars. This relieves the burden of generating catalogs from modelers on one hand and on the other hand allow the testing of models generated by different groups. This is a new scheme for sampling the analytical models which enables the user to generate continuous values of the variables like position and age of stars. Instead of a set of discs at specified ages, our methodology allows us to generate a disc which is continuous in age.

As a concrete example, Sharma et al (2011) demonstrate the **Besançon** analytical model for the disc. To model the disc kinematics more accurately, **Galaxia** employs the Shu (1969) distribution function that describes the non-circular motion in the plane of the disc. This function has now been generalised for a range of rotation curves (Sharma & Bland-Hawthorn 2013). For the stellar halo, **Galaxia** uses the simulated N-body models of Bullock & Johnston (2005) which can reproduce the substructure in the halo. We show a scheme for sampling the N-body particles such that the sampled stars preserve the underlying phase space density of N-body particles.

Another powerful aspect of **Galaxia** is the use of Markov Chain Monte (MCMC) Carlo methods that allow the determination of many parameters from model fitting applied to a large stellar survey. For the first time, we find that two very different, large stellar surveys (GCS, RAVE) yield the same kinematic parameters for the local disk (Sharma et al 2013). Moreover, the more spatially extended RAVE data demand the use of a Shu distribution function to yield meaningful results. This has strong parallels with MCMC analysis of galaxy redshift surveys in the context of Λ CDM simulations. In time, it will be possible to apply MCMC model fitting of the most detailed Galaxy simulations to stellar surveys which of course reinforces the remarkable complementarity between near-field and far-field cosmology.

Finally, we would like to commend the ESA-ESO Working Group Report 4: Galactic Populations, Chemistry and Dynamics (Turón & Primas 2008). This is a very useful compendium of the major problems in Galactic astronomy, ways to attack them, and major surveys past, present and future.

“Discursive” is an interesting adjective that has two opposing meanings: either ‘proceeding logically’ or ‘jumping from topic to topic.’ Only the reader can decide which of these two interpretations best describes our review. The document has been compiled from lectures that were originally delivered in 2007 at Mürren in Switzerland as part of the Saas Fee series. But the lectures were further developed for seminar series delivered at the University of Wisconsin (2011), the University of Bologna (2011) and the University of Sydney (2012). Powerpoint slides for these lecture series are available upon request. Updated text from our 2002 *Annual Reviews of Astronomy & Astrophysics* article is also included.

We are indebted to Mark Krumholz, David Webster, Volker Bromm and Torgny Karlsson for important contributions to our published work, some aspects of which are repeated here. We have learnt a great deal from conversations with

Alice Quillen, Sanjib Sharma, James Binney, Jerry Sellwood, Jerry Ostriker, Simon White, Scott Tremaine and Paul Nulsen. JBH is supported by a Federation Fellowship from the Australian Research Council, and further acknowledges a Merton College Fellowship, a Leverhulme Fellowship (Oxford), and a Brittingham Scholarship (Wisconsin). Finally, we owe a huge debt of gratitude to Ben Moore and Francesca Matteucci for their patience and forbearance over the years.

REFERENCES

- [1] Abadi MG, Navarro JF, Steinmetz M & Eke VR 2003, *Ap. J.* 597, 21
- [2] Abbasi et al 2012, *Phys. Rev. D* 85, 2002
- [3] Abe, F. et al 1999, *Astron. J.* 118, 261–72
- [4] Abel T, Bryan GL & Norman ML 2000, *Ap. J.* 540, 39
- [5] Abel T, Bryan GL & Norman ML 2002, *Science* 295, 93
- [6] Adams F et al 2006, *Ap. J.* 641, 504
- [7] Allen L et al 2007, In *Protostars and Planets V*, eds. Reipurth, Jewitt & Keil, 361–376
- [8] An FP et al 2012, *Phys. Rev. Lett.* 108, 1803
- [9] Ambartsumian VA 1958, *IAU Symp.* 5, 4
- [10] Anderson ME & Bregman JN 2011, *Ap. J.* 737, 22
- [11] Argast D, Samland M, Gerhard OE, Thielemann F-K 2000, *Astron. Astrophys.* 356, 873–87
- [12] Arimoto N & Yoshii Y 1987, *Astron. Astrophys.* 173, 23–38
- [13] Arnaboldi M, Freeman KC, Hui X, Capaccioli M, Ford H 1994, *ESO Messenger* 76, 40–44
- [14] Arnould M & Goriely S 2001, In *Astrophysical Ages and Time Scales*, eds T von Hippel, C Simpson & N Manset, *Astron. Soc. Pac.* 245, 252–261
- [15] Arp H 1962, *Ap. J.* 136, 66–74
- [16] Arras P & Wasserman I 1999, *MNRAS* 306, 257–78
- [17] Ashman KM & Zepf SE 1992, *Ap. J.* 384, 50–61
- [18] Athanassoula L 2007, *MNRAS* 377, 1569
- [19] Audouze J & Silk J 1995, *Ap. J.* 451, L49–52
- [20] Aumer M & Binney J 2009, *MNRAS* 397, 1286
- [21] Babcock HW 1939, *Lick Obs. Bull.* 498, 41
- [22] Babich D & Loeb A 2006, *Ap. J.* 640, 1
- [23] Babul A & Rees M 1992, *MNRAS* 255, 346–50
- [24] Bagchi J et al 2006, *Science* 314, 791
- [25] Bahcall JN & Soneira RM 1984, *Ap. J. Suppl.* 44, 73
- [26] Bakos J, Trujillo I & Pohlen M 2008, *Ap. J.* 683, L103
- [27] Baldry IK et al 2004, *Ap. J.* 600, 681
- [28] Baldry IK et al 2008, *MNRAS* 388, 945
- [29] Barkana R & Loeb A 1999, *Ap. J.* 523, 54
- [30] Barkana R & Loeb A 2007, *Rep. Prog. Phys.* 70, 627
- [31] Barnes S 2007, *Ap. J.* 669, 1167
- [32] Barnes JE & Hernquist LE 1992, *Nature* 360, 715
- [33] Bastian N et al 2012, *MNRAS* 419, 2606
- [34] Battaglia G et al 2005, *MNRAS* 364 433
- [35] Battaglia G et al 2012, *Ap. J.* 761, L31
- [36] Baugh C, Cole S, Frenk C 1996, *MNRAS* 283, 1361–78
- [37] Belenkiy A 2012, *Physics Today*, 65, 38
- [38] Bell E, de Jong R. 2000. *MNRAS* 312, 497–520
- [39] Bell EF et al 2008, *Ap. J.* 680, 295
- [40] Belokurov V et al 2009, *MNRAS* 397, 1748
- [41] Benjamin RA & Danly L 1997, *Ap. J.* 481, 764
- [42] Benjamin RA et al 2005, *Ap. J.* 630, L149
- [43] Bensby T et al 2005, *A&A* 433 185
- [44] Benson AJ 2010, *Physics Reports* 495, 33
- [45] Benson AJ & Bower R 2011, *MNRAS* 410, 2653
- [46] Bertelli G et al 1995, *Astron. Astrophys.* 301, 381
- [47] Bertelli G et al 1994, *Astron. Astrophys. Suppl.* 106, 275
- [48] Bertelli G & Nasi E 2001, *Astron. J.* 121, 1013–1023
- [49] Bertoldi F & McKee CF 1992, *Ap. J.* 395, 140
- [50] Bica E, Alloin D & Schmidt A, 1990, *MNRAS* 242, 241–49
- [51] Bienayme O, Robin A & Creze M 1987, *Astron. Astrophys.* 180, 94
- [52] Binggeli B, Sandage A & Tammann GA 1985, *Astron. J.* 90, 1681–771
- [53] Binney JJ 1977, *Ap. J.* 215, 483
- [54] Binney JJ & Lacey C 1988, *MNRAS* 230, 597–627
- [55] Binney JJ, Dehnen W & Bertelli G 2000, *MNRAS* 318, 658
- [56] Binney JJ & Tremaine SD 2008, *Galactic Dynamics: Second Edition*, (Princeton University Press)
- [57] Birnboim Y & Dekel A 2003, *MNRAS* 345, 349
- [58] Blake C et al 2011, *MNRAS* 418, 1707
- [59] Bland-Hawthorn J, Freeman KC & Quinn PJ 1997, *Ap. J.* 490, 143
- [60] Bland-Hawthorn J et al 1998, *MNRAS* 299, 611–624
- [61] Bland-Hawthorn J 1999, *Nature* 400, 220
- [62] Bland-Hawthorn J & Freeman KC 2000, *Science* 287, 79–84
- [63] Bland-Hawthorn J & Freeman KC 2004, *PASA* 21, 110
- [64] Bland-Hawthorn J & Freeman KC 2006, *Mem. Soc. della Societa Astron. Ital.*, 77, 1095
- [65] Bland-Hawthorn J & Maloney PR 2001, In *Extragalactic Gas at Low Redshift*, eds. J. Mulchaey & J. Stocke, *Astron. Soc. Pac.* 254, 267–82
- [66] Bland-Hawthorn J et al 1998, *MNRAS* 299, 611
- [67] Bland-Hawthorn J et al 2005, *Ap. J.* 629, 239
- [68] Bland-Hawthorn J et al 2007, *Ap. J.* 670, L109
- [69] Bland-Hawthorn J 2008, In *Galaxies in the Local Volume*, *Astrophys. Sp. Science* 1, 259
- [70] Bland-Hawthorn J 2009, In *The Galaxy Disk in Cosmological Context*, *IAU Symp.* 254, 241
- [71] Bland-Hawthorn J, Krumholz MR & Freeman KC 2010a, *Ap. J.* 713, 166
- [72] Bland-Hawthorn J et al 2010b, *Ap. J.* 721, 582
- [73] Blanton M & Moustakas J 2009, *Annu. Rev. Astron. Astrophys.* 47, 159
- [74] Blanton M et al 2003, *Ap. J.* 594, 186
- [75] Blitz L, Spergel DN, Teuben PJ, Hartmann D & Burton WB 1999, *Ap. J.* 514, 818–843
- [76] Blumenthal G, Faber S, Primack J & Rees M 1984, *Nature* 311, 517–525
- [77] Bondi H 1952, *MNRAS* 112, 195
- [78] Bondi H & Hoyle F 1944, *MNRAS* 104, 273
- [79] Bondi H, Hoyle F & Lyttleton RA 1947, *MNRAS* 107, 184
- [80] Bosma A 1978, PhD, University of Groningen, Netherlands
- [81] Boss L 1908, *Astron. J.* 26, 31–36
- [82] Bouwens R et al 2007, *Ap. J.* 670, 928
- [83] Bower R, Kodama T & Terlevich A 1998, *MNRAS* 299, 1193–1208
- [84] Bower R et al 2006, *MNRAS* 370, 645
- [85] Boylan-Kolchin M et al 2009, *MNRAS* 398, 1150
- [86] Braun R & Burton WB 1999, *Astron. Astrophys.* 341, 437–50
- [87] Bromm V & Yoshida N 2011, *Annu. Rev. Astron. Astrophys.* 49, 37
- [88] Bromm V, Coppi, PS & Larson RB 1999, *Ap. J.* 527, L5
- [89] Bromm V, Coppi, PS & Larson RB 2002, *Ap. J.* 564, 23
- [90] Bromm V, Yoshida N, Hernquist L & McKee CF 2009, *Nature* 459, 49
- [91] Brook C et al 2011, *MNRAS* 415, 1051
- [92] Brooks AM et al 2009, *Ap. J.* 694, 396
- [93] Brown RL & van den Bout PA 1991, *Astron. J.* 102, 1956
- [94] Brown A, Velazquez HM & Aguilar LA 2005, *MNRAS* 359, 1287
- [95] Bubar E & King JR 2010, *Astron. J.* 140, 293
- [96] Bullock JS, Kravtsov A & Weinberg D 2000, *Ap. J.* 539, 517–21
- [97] Bullock JS & Johnston KV 2005 *Ap. J.* 635, 931
- [98] Bullock JS et al 2001 *Ap. J.* 321, 559
- [99] Burbidge EM, Burbidge GR, Fowler WA & Hoyle F 1957, *Rev. Mod. Phys.* 29, 547–650
- [100] Bureau M, Freeman KC, Pfitzner DW & Meurer GR 1999, *Astron. J.* 118, 2158–71
- [101] Burris DL et al 2000, *Ap. J.* 544, 302–319
- [102] Burrows A et al 2006, *Ap. J.* 640, 878

- [103]Burstein D, Faber SM, Gaskell CM & Krumm N 1984, *Ap. J.* 287, 586–609
- [104]Butcher H 1987, *Nature* 328, 127–31
- [105]Cabrera-Lavers A et al 2008 *Astron. Astrophys.* 491, 781
- [106]Calcaneo-Roldan C, Moore B, Bland-Hawthorn J & Sadler EM 2000, *MNRAS* 314, 324–33
- [107]Caldwell N & Rose J 1998, *Astron. J.* 115, 1423–1432
- [108]Caloi V & d'Antona F 2011, *MNRAS* 417, 228
- [109]Carlberg RG & Sellwood JA 1985, *Ap. J.* 292, 79–89
- [110]Carlberg RG 1986, *Ap. J.* 310, 593–596
- [111]Carlberg et al 1985, *Ap. J.* 294, 674
- [112]Carney BW, Aguilar L, Latham DW & Laird JB 1990, *Astron. J.* 99, 201–20
- [113]Carney BW, Laird JB, Latham DW & Aguilar LA 1996, *Astron. J.* 112, 668–92
- [114]Carney BW 1993, In *The Globular Cluster – Galaxy Connection* eds. GH Smith & JP Brodie, *Astron. Soc. Pac.* 48, 234–45
- [115]Carollo D et al 2007, *Nature* 450, 1020
- [116]Carpenter JM 2000, *Astron. J.* 120, 3139–61
- [117]Carraro G, Ng Y & Portinari L 1998, *MNRAS* 296, 1045–56
- [118]Carretta E et al 2002, *Astron. J.* 124, 481
- [119]Castro S et al 1997, *Astron. J.* 114, 376–87
- [120]Castro S, Porto de Mello GF & da Silva L 1999, *MNRAS* 305, 693
- [121]Cayrel R et al 2001, *Nature* 409, 691–92
- [122]Cayrel R et al 2004, *Astron. Astrophys.* 416, 1117
- [123]Ceccarelli C et al 2005, In *Protostars and Planets V*, eds. Reipurth, Jewitt & Keil, 47–62
- [124]Chaboyer B, Demarque P, Kernan P & Krauss L 1998, *Ap. J.* 494, 96–110
- [125]Chaboyer B, Demarque P & Sarajedini A 1996, *Ap. J.* 459, 558–69
- [126]Chaboyer B 1998, *Physics Reports*, 307, 23–30
- [127]Chandrasekhar S 1961, *Hydrodynamic & Hydromagnetic Stability*, Dover Books
- [128]Charlot S, Silk J 1994, *Ap. J.* 432, 453–463
- [129]Chiappini C, Matteucci F & Romano D 2001, *Ap. J.* 554, 1044–58
- [130]Chiappini C et al 1999, *Ap. J.* 515, 226
- [131]Chiba M 2001, *Ap. J.* 565, 17–23
- [132]Chiba M & Beers T 2000, *Astron. J.* 119, 2843–65
- [133]Chou M-Y et al 2010, *Ap. J.* 708, 1290
- [134]Christensen-Dalsgaard J 1986, *IAU Symp.* 123, 295
- [135]Christlieb N et al 2001, In *Astrophysical Ages and Time Scales*, eds T von Hippel, C Simpson & N Manset, *Astron. Soc. Pac.* 245, 298–300
- [136]Christlieb N et al 2002, *Nature* 419, 904
- [137]Clark PC, Glover SCO & Klessen RS 2008, *Ap. J.* 672, 757
- [138]Clark PC, Glover SCO, Klessen RS & Bromm V 2011, *Ap. J.* 727, 110
- [139]Clarke CJ, Bonnell IA, Hillenbrand LA 2000, In *Protostars and Planets IV*, eds. V Mannings, AP Boss, SS Russell, 151–177 (University of Arizona Press)
- [140]Clayton D 1987, *Ap. J.* 315, 451–459
- [141]Clayton D 1988, *MNRAS* 234, 1–36
- [142]Codis S et al 2012, *MNRAS* 427, 3320
- [143]Cohen JG, Christlieb N, Beers TC, Gratton R & Carretta E 2002, *Astron. J.* 124, 470
- [144]Cohen JG et al 2007, *Ap. J.* 659, L161
- [145]Cole S, Aragon-Salamanca A, Frenk CS, Navarro JF, Zepf SE 1994, *MNRAS* 271, 781–806
- [146]Cole S et al 2005, *MNRAS* 362, 505
- [147]Coleman MG et al 2004 *Astron. J.* 127, 832
- [148]Colin P, Avila-Reese V & Valenzuela O 2000, *Ap. J.* 542, 622–30
- [149]Colless MM et al 2001, *MNRAS* 328, 1039
- [150]Combes F et al 1990, *Astron. Astrophys.* 233, 82
- [151]Concannon KD, Rose JA & Caldwell N 2000, *Ap. J.* 536, L19–22
- [152]Connors TW, Kawata D & Gibson BK 2006, *MNRAS* 371, 108
- [153]Conroy C 2012, *Ap. J.* 758, 21
- [154]Conroy C & Spergel D 2011, *Ap. J.* 726, 36
- [155]Cooke R et al 2011, *MNRAS* 412, 1047
- [156]Cooper AP et al 2009, *MNRAS* 406, 744
- [157]Courteau S, de Jong RS & Broeils AH 1996, *Ap. J.* 457, L73–76
- [158]Cowan JJ, McWilliam A, Sneden C & Burris DL 1997, *Ap. J.* 480, 246–54
- [159]Cox TJ & Loeb A 2008, *MNRAS* 386, 461
- [160]Creze M, Chereul E, Bienayme O & Pichon C 1998, *Astron. Astrophys.* 329, 920–36
- [161]Croton D et al 2012, *MNRAS* 365, 11
- [162]Da Costa G 1999, In *The Third Stromlo Symposium, The Galactic Halo*, eds. B Gibson, T Axelrod & M Putman, *Astron. Soc. Pac.* 165:153–166
- [163]Dalcanton JJ, Spergel DN & Summers FJ 1997, *Ap. J.* 482, 659–76
- [164]Danforth CW & Shull MJ 2008, *Ap. J.* 679, 194
- [165]Davis M, Efstathiou G, Frenk CS & White SDM 1985, *Ap. J.* 292, 371
- [166]Davies RL et al 2001, *Ap. J.* 548, L33-L36
- [167]de Bruijne J 1999, *MNRAS* 306, 381–93
- [168]de Grijs R, Kregel M & Wesson KH 2001, *MNRAS* 324, 1074–86
- [169]de Jong R & Lacey C 2000, *Ap. J.* 545, 781–97
- [170]de Jong R et al 2012, *SPIE Conf.* 8446, 0
- [171]De Lucia & Helmi 2008, *MNRAS* 391, 14
- [172]De Silva GM et al 2006, *Astron. J.* 131, 455
- [173]De Silva GM et al 2007a, *Astron. J.* 133, 694
- [174]De Silva GM et al 2007b, *Astron. J.* 133, 1161
- [175]De Silva GM, Freeman KC & Bland-Hawthorn J 2009, *PASA* 26, 11
- [176]De Silva GM et al 2013, *MNRAS* in press
- [177]de Vaucouleurs G 1961, *Ap. J. Suppl.* 6, 213
- [178]de Vaucouleurs G 1970a, *Science* 167, 1203
- [179]de Vaucouleurs G 1970b, *IAU Symp.* 38, 18
- [180]De Young DS 2003, *New Astron. Rev.* 47, 545
- [181]de Zeeuw PT, Hoogerwerf R, de Bruijne JHJ, Brown AG & Blaauw A 1999, *Astron. J.* 117, 354–99
- [182]Dehnen WA 2000, *Astron. J.* 119, 800–12
- [183]Dekel A & Silk J 1986, *Ap. J.* 303, 39–55
- [184]Dekel A et al 2009 *Nature* 457, 451
- [185]Dijkstra M et al 2004, *Ap. J.* 601, 666
- [186]Dodd KN 1953 *MNRAS* 113, 484
- [187]Doroshkevich et al 1967, *Soviet Astron.* 11, 233
- [188]Dressler A 1980, *Ap. J.* 236, 351–365
- [189]Driver SP et al 2007, *Ap. J.* 657, L85
- [190]Driver SP et al 2013, *MNRAS* in press
- [191]Dubois Y et al 2012, *MNRAS* 420, 2662
- [192]Durrell PR, Harris WE & Pritchett CJ 2001, *Astron. J.* 121, 2557–71
- [193]Edmunds MG 1990, *Nature* 348, 395–396
- [194]Edvardsson B et al 1993, *Astron. Astrophys.* 275, 101–52
- [195]Efremov YN 2011, *Astron. Reports* 55, 108–122
- [196]Efstathiou GP 1992, *MNRAS* 256, 43P
- [197]Efstathiou GP et al 1990, *Nature* 348, 705
- [198]Efstathiou GP et al 2002, *MNRAS* 330, L29–L35
- [199]Eggen OJ, Lynden-Bell D & Sandage AR 1962, *Ap. J.* 136, 748–66
- [200]Eggen OJ & Sandage AR 1959, *MNRAS* 119, 255–77
- [201]Eggen OJ & Sandage AR 1969, *Ap. J.* 158, 669–84
- [202]Eggen OJ 1977, *Ap. J.* 215, 812–26
- [203]Einasto J, Saar E, Kaasik A & Chernin AD 1974, *Nature* 252, 111
- [204]Eisenstein DJ et al 2005, *Ap. J.* 633, 560
- [205]Ekholm T & Teerikorpi P 1994, *Astron. Astrophys.* 284, 369
- [206]Ellis RS, Abraham RG, Brinchmann J & Menanteau F 2000, *Astron. Geophys.* 41, 10–16
- [207]Elmegreen BG 2007, *Ap. J.* 668, 1064
- [208]Elmegreen BG 2010, *Ap. J.* 712, L184
- [209]Elmegreen BG, Efremov Y, Pudritz RE & Zinnecker H 2000, In *Protostars and Planets IV*, eds V Mannings, AP Boss, SS Russell, 179–215 (University of Arizona Press)
- [210]Elmegreen D et al 2005, *Ap. J.* 631, 85
- [211]Englmaier P & Gerhard OE 2006, *Celest. Mech. & Dyn. Astron.*, 94, 369–379
- [212]Eskridge PB et al 2000, *Astron. J.* 119, 536
- [213]Exter K, Barlow MJ, Walton NA & Clegg RES 2001, *Ap&SS*277, 199–99
- [214]Faber S 1973, *Ap. J.* 179, 731–54
- [215]Fabian AC 1994, *Annu. Rev. Astron. Astrophys.* 32, 277
- [216]Fabian AC & Nulsen PE 1977 *MNRAS* 180, 479
- [217]Falcon-Barroso J et al 2006, *MNRAS* 369, 529
- [218]Fall SM & Efstathiou GP 1980, *MNRAS* 193, 189–206
- [219]Fall SM 1983, In *Internal Kinematics and Dynamics of Galaxies*, ed. E Athanassoula, 391–98 (Reidel)
- [220]Fall SM 2006, *Ap. J.* 652, 1129

- [221] Fall SM, Chandar R & Whitmore BC 2005, *ApJ* 631, L133
- [222] Fall SM, Chandar R & Whitmore BC 2009, *Ap. J.* 704, 453
- [223] Fall SM, Krumholz MR & Matzner CD 2010, *Ap. J.* 710, L142
- [224] Fan X et al 2002, *Astron. J.* 123, 1247
- [225] Fardal MA et al 2001, *Ap. J.* 562, 605
- [226] Faúndez S et al 2004, *Astron. Astrophys.* 426, 97
- [227] Feltzing S, Holmberg J & Hurley JR 2001, *Astron. Astrophys.* 377, 911–24
- [228] Feltzing S & Holmberg J 2000, *Astron. Astrophys.* 357, 153–63
- [229] Ferraro F et al 2009, *Nature* 462, 483
- [230] Flynn C et al 2006, *MNRAS* 372, 1149
- [231] Font A, Navarro J, Stadel J & Quinn T 2001, *Ap. J.* 563, L1–L4
- [232] Fontani F et al 2005, *Astron. Astrophys.* 443, 921
- [233] Fontanot F et al 2009, *MNRAS* 397, 1776
- [234] Förster-Schreiber NM et al 2009, *Ap. J.* 706, 1364
- [235] Fowler WA & Hoyle F 1960, *Astron. J.* 65, 345–45
- [236] Frayer DT & Brown RL 1997, *Ap. J. Suppl.* 113, 221–43
- [237] Grebel A et al 2005, *Nature* 434, 871
- [238] Grebel A et al 2010a, *Ap. J.* 708, 560
- [239] Grebel A et al 2010b, *Nature* 464, 72
- [240] Freedman W et al 2001, *Ap. J.* 553, 47
- [241] Freeman KC, Illingworth G & Oemler A 1983, *Ap. J.* 272, 488–508
- [242] Freeman KC 1970, *Ap. J.* 160, 811–30
- [243] Freeman KC 1987, *Annu. Rev. Astron. Astrophys.* 25, 603–32
- [244] Freeman KC 1991, In *Dynamics of Disk Galaxies*, ed. B Sundelius (University of Goteborg), 15
- [245] Freeman KC 1993, In *The Globular Cluster – Galaxy Connection*, eds. GH Smith & JP Brodie, *Astron. Soc. Pac.* 48, 608–14
- [246] Freeman KC 1997, In *The Nature of Elliptical Galaxies*, eds. Arnaboldi, Da Costa & Saha, *Astron. Soc. Pac.* 116, 1
- [247] Freeman KC 2008, In *Formation and Evolution of Galaxy Disks*, eds. Funes & Corsini, *Astron. Soc. Pac.* 396, 3
- [248] Freeman KC & Bland-Hawthorn J 2002, *Annu. Rev. Astron. Astrophys.* 40, 487
- [249] Freeman KC & Bland-Hawthorn J 2008, In *Panoramic Views of Galaxy Formation & Evolution*, *Astron. Soc. Pac.* 399, 439
- [250] Freeman KC et al 2013, *MNRAS* 428, 3660
- [251] Friel ED 1995, *Annu. Rev. Astron. Astrophys.* 33, 381–414
- [252] Friel ED & Janes KA 1993, *Astron. Astrophys. TA* 77, 75–91
- [253] Fry AM, Morrison HL, Harding P & Boroson TA 1999, *Astron. J.* 118, 1209–19
- [254] Fryer CL 1999, *Ap. J.* 522, 413
- [255] Fryer CL & Heger A 2000, *Ap. J.* 541, 1033
- [256] Fryer CL, Woosley S & Heger A 2001, *Ap. J.* 550, 372
- [257] Fulbright JP et al 2007, *Ap. J.* 661, 1152
- [258] Fumagalli M et al 2011, *Science* 334, 1245
- [259] Furlanetto S & Loeb A 2004, *Ap. J.* 611, 642
- [260] Gaensler BM et al 2008, *PASA* 25, 184
- [261] Gallagher JS et al 2003, *Ap. J.* 588, 326
- [262] Gao L & Theuns T 2007, *Science* 317, 1527
- [263] Garcia-Berro E & Iben I 1994, *Ap. J.* 434, 306
- [264] Gardner E & Flynn C 2010, *MNRAS* 405, 545
- [265] Genzel R et al 2006, *Nature* 442, 786
- [266] Gilmore G & Reid IN 1983, *MNRAS* 202, 1025–47
- [267] Gilmore G, Wyse RFG & Jones BJ 1995, *Astron. J.* 109, 1095–1111
- [268] Gilmore G, Wyse RFG & Kuijken K 1989, *Annu. Rev. Astron. Astrophys.* 27, 555–627
- [269] Gilroy K, Sneden C, Pilachowski CA & Cowan JJ 1988, *Ap. J.* 327, 298–320
- [270] Gimenez A & Favata F 2001, In *Astrophysical Ages & Time Scales*, eds. T von Hippel, C Simpson & N Manset, *Astron. Soc. Pac.* 245, 304–306
- [271] Girardi L et al 2005, *Astron. Astrophys.* 436, 895
- [272] Glazebrook K 2012, *astro-ph/1212.3065*
- [273] Gnedin NY 2000, *Ap. J.* 542, 535
- [274] Gnedin NY & Kravtsov A 2006, *Ap. J.* 645, 1054
- [275] Gnedin OY, Lee HM & Ostriker JP 1999, *Ap. J.* 522, 935–49
- [276] Götz M & Köppen J 1992, *Astron. Astrophys.* 262, 455–67
- [277] Gomez A et al 1997, In *Hipparcos - Venice '97*, 621–24 (ESA Publ.)
- [278] Gonzalez V et al 2011, *Ap. J.* 735, L34
- [279] Goriely S & Arnould M 1996, *Astron. Astrophys.* 312, 327–337
- [280] Goriely S & Arnould M 2001, *Astron. Astrophys.* 379, 1113–22
- [281] Gottlöber S, Yepes G, Wager C & Sevilla R 2006, *astro-ph/0608289*
- [282] Gottlöber S, Hoffman Y & Yepes G 2010, *astro-ph/1005.2687*
- [283] Gough DO 1987, *Nature* 326, 257–259
- [284] Gough DO 2001, In *Astrophysical Ages & Time Scales*, eds. T von Hippel, C Simpson & N Manset, *Astron. Soc. Pac.* 245, 31–43
- [285] Governato F et al 2012, *MNRAS* 422, 1231
- [286] Gratton RG et al 1997, *Ap. J.* 491, 49–71
- [287] Gratton R, Sneden C & Carretta E 2004, *Annu. Rev. Astron. Astrophys.* 42, 385
- [288] Grevecich J & Putman ME 2009, *Ap. J.* 696, 385
- [289] Grebel EK 2001, *Ap&SS* 277, 231–39
- [290] Grebel EK & Gallagher JS 2004, *Ap. J.* 610, L89
- [291] Gregg MD, West MJ. 1998. *Nature* 396, 549–52
- [292] Grenon M 1999, *Ap&SS* 265, 331–36
- [293] Greif T et al 2011, *Ap. J.* 736, 147
- [294] Griffin R 1998, *Observatory*, 118, 223–25
- [295] Guenther DB 1989, *Ap. J.* 339, 1156–59
- [296] Guenther DB & Demarque P 1997, *Ap. J.* 484, 937–59
- [297] Gunn JE & Gott JR 1972, *Ap. J.* 176, 1
- [298] Gunn JE 1977, *Ap. J.* 218, 592
- [299] Guth A 1997, *The Inflationary Universe* (Addison-Wesley)
- [300] Haiman Z, Thoul AA & Loeb A 1996, *Ap. J.* 464, 523
- [301] Hamann F & Ferland G 1999, *Annu. Rev. Astron. Astrophys.* 37, 487–531
- [302] Harding P et al 2001, *Astron. J.* 122, 1397–1419
- [303] Harris WE 1991, *Annu. Rev. Astron. Astrophys.* 29, 543–79
- [304] Harris WE & Pudritz RE 1994, *Ap. J.* 429, 177–91
- [305] Harris WE 2002, *IAU Symp.* 207, 755
- [306] Hartmann L, Ballesteros-Paredes J & Bergin EA 2001, *Ap. J.* 562, 852
- [307] Hartkopf WI & Yoss KM 1982, *Astron. J.* 87, 1679–709
- [308] Heckman TM et al 1989, *Ap. J.* 342, 735
- [309] Heger A & Woosley S 2002, *Ap. J.* 567, 532
- [310] Heitsch F & Putman ME 2009 *Ap. J.* 698, 1485
- [311] Helmi A 2008, *Astron. Astrophys. Rev.* 15, 145
- [312] Helmi A, Springel V & White SDM 2001, *Phys. Rev. D* 66:6, 063502
- [313] Helmi A, White SDM, de Zeeuw PT & Zhao H-S 1999, *Nature* 402, 53–55
- [314] Helmi A & White SDM 1999, *MNRAS* 307, 495–517
- [315] Helmi A, Zhao H-S & de Zeeuw PT 1999, In *The Third Stromlo Symposium: The Galactic Halo*, eds. BK Gibson, TS Axelrod, ME Putman, *Astron. Soc. Pac.* 165:125–29
- [316] Henry RBC 1998, In *Abundance Profiles, Diagnostic Tools for Galaxy History*, eds. D Friedli, M Edmunds, C Robert & L Drissen, 47–59
- [317] Hermit S et al 1996, *MNRAS* 283, 709
- [318] Hernquist L & Quinn PJ 1988, *Ap. J.* 331, 682–698
- [319] Heyer MH & Brunt CM 2004, *Ap. J.* 615, L45
- [320] Hibbard JE & van Gorkom JH 1996, *Astron. J.* 111, 655
- [321] Hill V et al 2002, *Astron. Astrophys.* 387, 560
- [322] Hill A et al 2008, *Ap. J.* 686, 363
- [323] Hirshfeld A, McClure R & Twarog BA 1978, In *The HR Diagram – the 100th Anniversary of Henry Norris Russell*, eds. AG Davis Philip & DS Hayes, 163 (Reidel)
- [324] Hoffman Y & Ribak E 1991, *Ap. J.* 380, L5
- [325] Hogan CJ & Dalcanton JJ 2001 *Ap. J.* 561, 35–45
- [326] Hohl F 1970, *IAU Symp.* 38, 368
- [327] Holtzman JA et al 1992, *AJ*, 103, 691–702
- [328] Hopkins AM, Irwin MJ & Connolly AJ 2001, *Ap. J.* 558, L31–33
- [329] Hopkins A & Beacom J 2006, *Ap. J.* 651, 142
- [330] Hopkins A, Irwin MJ & Connolly AJ 2001, *Ap. J.* 558, L31
- [331] Hosokawa T, Omukai K, Yoshida N & Yorke HW 2011, *Science* 334, 1250
- [332] Howard CD et al 2009, *Ap. J.* 702, L153
- [333] Hoyle F 1966, *Galaxies, Nuclei & Quasars* (London: Heinemann)
- [334] Hoyle F & Lyttleton RA 1939, *Proc. Cam. Phil. Soc.* 35, 405
- [335] Hubble E & Humason M 1931, *Ap. J.* 74, 43–80
- [336] Huff EM & Stahler SW 2006, *Ap. J.* 644, 355
- [337] Hunt R 1971, *MNRAS* 154, 141
- [338] Hunt R 1975, *MNRAS* 173, 465
- [339] Ibata R, Gilmore G & Irwin MJ 1994, *Nature* 370, 194–196
- [340] Ibata R, Gilmore G & Irwin MJ 1995, *MNRAS* 277, 781–800
- [341] Ibata R, Irwin M, Lewis G, Ferguson AMN & Tanvir N 2001a, *Nature* 412, 49–52

- [342] Ibata R, Lewis GF, Irwin MJ, Totten E & Quinn T 2001b, *Ap. J.* 551, 294–311
- [343] Ibata R, Irwin MJ, Lewis GF & Stolte A 2001c, *Ap. J.* 547, L133–36
- [344] Ibata R & Lewis G 1998, *Ap. J.* 500, 575–90
- [345] Ibata R, Wyse RFG, Gilmore G, Irwin MJ & Suntzeff NB 1997, *Astron. J.* 113, 634–55
- [346] Ikeuchi S 1986, *Ap&SS*118, 509–14
- [347] Irwin M et al 2000, *Ap. J.* 628, L105
- [348] Ishimaru Y & Wanajo S 2000, In *The First Stars*, eds. A Weiss, TG Abel & V Hill, 189–193 (Berlin, Springer)
- [349] Ivezić Z et al 2008, *Ap. J.* 684, 287
- [350] Jablonka P, Martin P & Arimoto N 1996, *Astron. J.* 112, 1415–22
- [351] Janes KA & Phelps RL 1994, *Astron. J.* 108, 1773–85
- [352] Janka H et al 2007, *Phys. Rep.*, 442, 38
- [353] Jeffries RD 2007, *MNRAS* 381, 1169
- [354] Jenkins A & Binney J 1990, *MNRAS* 245, 305–17
- [355] Jenkins A et al 2001, *MNRAS* 321, 372–84
- [356] Johnson JA, Bolte M 2001, *Ap. J.* 554, 888–902
- [357] Johnston KV, Hernquist L & Bolte M 1996, *Ap. J.* 465, 278–287
- [358] Johnston KV, Sackett PD & Bullock JS 2001, *Ap. J.* 557, 137–149
- [359] Johnston KV, Spergel DN & Haydn C 2002, *Ap. J.* 570, 656
- [360] Johnston KV 1998, *Ap. J.* 495, 297–308
- [361] Jones L, Smail I & Couch W 2000, *Ap. J.* 528, 118–122
- [362] Jones L & Worthey G 1995 *Ap. J.* 446, L31–L35
- [363] Joung MR, Bryan GL & Putman ME 2012, *Ap. J.* 745, 148
- [364] Just A 2001, In *Disks of Galaxies, Kinematics, Dynamics and Perturbations*, eds. E Athanassoula, A Bosma, I Puerari, *Astron. Soc. Pac.* 275, 117
- [365] Käppeler F, Thielemann FK & Wiescher M 1998, *Annu. Rev. Nucl. Part. Sci.* 48, 175–251
- [366] Kafle PR, Sharma S, Lewis GF & Bland-Hawthorn J 2012, *Ap. J.* 761, 98
- [367] Kafle PR, Sharma S, Lewis GF & Bland-Hawthorn J 2013, *MNRAS* press
- [368] Kahn FD & Woltjer L 1959, *Ap. J.* 130, 705
- [369] Kaiser N 2002, *Elements of Astrophysics* (University of Hawaii Press)
- [370] Kalirai J 2012, *Nature* 486, 90
- [371] Kalnajs A 1983, *IAU Symp.* 100, 109–115 (+comment)
- [372] Kamionkowski M & Liddle AR 2000, *Phys. Rev. Lett.* 84, 4525–8
- [373] Karlsson T 2005, *Astron. Astrophys.* 439, 93
- [374] Karlsson T & Gustafsson B 2001, *Astron. Astrophys.* 379, 461–81
- [375] Karlsson T & Gustafsson B 2005, *Astron. Astrophys.* 436, 879
- [376] Karlsson T, Bland-Hawthorn J, Freeman KC & Silk J 2012, *Ap. J.* 759, 111
- [377] Karlsson T, Bromm V & Bland-Hawthorn J 2013, *Rev. Mod. Phys.* in press
- [378] Katz N, Keres D, Dave R & Weinberg DH 2003, In *The IGM/Galaxy Connection*, 185 (Kluwer)
- [379] Kauffmann G 1996, *MNRAS* 281, 487–492
- [380] Kauffmann G & Charlot S 1998, *MNRAS* 294, 705–17
- [381] Kauffmann G, White SDM & Guiderdoni B 1993, *MNRAS* 264, 201–18
- [382] Kauffmann G et al 2004, *MNRAS* 353, 713
- [383] Kauffmann G, Li C & Heckman T 2010, *MNRAS* 409, 491
- [384] Kennicutt RC 1989, *Ap. J.* 344, 685
- [385] Keres D et al 2005, *MNRAS* 363, 2
- [386] Keres D & Hernquist LE 2009, *Ap. J.* 700, L1
- [387] Keres D et al 2009, *MNRAS* 395, 160
- [388] Kimm T et al 2012, *MNRAS* 425, L96
- [389] King IR 1958, *Astron. J.* 63, 109–13
- [390] Kinman TD 1959, *MNRAS* 119, 538–58
- [391] Kirby E et al 2008, *Ap. J.* 685, L43
- [392] Kitayama T & Ikeuchi S 2000, *Ap. J.* 529, 615
- [393] Klement RJ 2010, *Astron. Astrophys. Rev.* 18, 567–594
- [394] Klypin A, Kravtsov AV, Valenzuela O & Prada F 1999, *Ap. J.* 522, 82–92
- [395] Klypin A, Hoffman Y, Kravtsov AV & Gottlöber S 2003, *Ap. J.* 596, 19
- [396] Klypin A, Trujillo-Gomez S & Primack J 2011, *Ap. J.* 740, 102
- [397] Kochanek C 1996, *Ap. J.* 457, 228–43
- [398] Komatsu E et al 2011, *Ap. J. Suppl.* 192, 18
- [399] Kormendy J 1993, *IAU Symp.* 153, 209
- [400] Kormendy J & Kennicutt RC 2004, *Annu. Rev. Astron. Astrophys.* 42, 603
- [401] Kormendy J, Drory N, Bender R & Cornell ME 2010, *Ap. J.* 723, 54
- [402] Kravtsov AV, Klypin A & Hoffman Y 2002 *Ap. J.* 571, 563
- [403] Kroupa P 2001, *MNRAS* 322, 231
- [404] Kroupa P 2002, *MNRAS* 330, 707–18
- [405] Kroupa P, Pawlowski M & Milgrom M 2012, *Int. J. Mod. Phys.*, 21:14
- [406] Krumholz MR & Tan JC 2007, *Ap. J.* 654, 304
- [407] Kuijken K & Gilmore G 1991, *Ap. J.* 367, L9–L13
- [408] Kuiper E et al 2011, *MNRAS* 417, 1088
- [409] Kuntschner H 2000, *MNRAS* 315, 184–208
- [410] Kuntschner J & Davies RL 1998, *MNRAS* 295, L29–L33
- [411] Kurucz RL 1991, In *The Solar Interior and Atmosphere*, eds. AN Cox, WC Livingston, M Matthews, 663 (Tucson, U. Arizona Press)
- [412] Kurucz RL 1995, In *Laboratory and Astronomical High Resolution Spectra*, eds. AJ Sauval, R Blomme, N Grevesse, 17
- [413] Lacey C & Cole S 1993, *MNRAS* 262, 627
- [414] Lacey C & Fall M 1983, *MNRAS* 204, 791–810
- [415] Lacey C & Fall M 1985, *Ap. J.* 290, 154–70
- [416] Lada CJ & Lada EA 2003, *Annu. Rev. Astron. Astrophys.* 41, 57
- [417] Larsen SS 2009, *Astron. Astrophys.* 494, 539
- [418] Larson R 1974, *MNRAS* 169, 229–246
- [419] Law R et al 2007, *Ap. J.* 669, 929
- [420] Law DR, Makewski S & Johnston KV 2009, *Ap. J.* 703, L67
- [421] Law DR et al 2012, *Ap. J.* 745, 85
- [422] Lee Y-W 1999, *Nature* 402, 55–57
- [423] Leggett S et al 1998, *Ap. J.* 497, 294
- [424] Lemaître G 1927, *Annales de la Société Scientifique de Bruxelles, Serie A* 47:49, 1927
- [425] Liddle A & Lyth D 2000, *Cosmological Inflation and Large-Scale Structure* (Cambridge University Press)
- [426] Lilly SJ et al 1996, *Ap. J.* 406, L1
- [427] Lin D & Pringle J 1987, *Ap. J.* 320, L87–91
- [428] Lineweaver CH 1999, *Science* 284, 1503–1507
- [429] Liu W & Chaboyer B 2000, *Ap. J.* 544, 818–29
- [430] Lockman FJ et al 2008, *Ap. J.* 679, L21
- [431] Loeb A 2010, *How Did the First Stars and Galaxies Form?* (Princeton University Press, Princeton)
- [432] López-Corredoira M, Cabrera-Lavers A & Gerhard OE 2005, *Astron. Astrophys.* 439, 107
- [433] López-Corredoira M et al 2007, *Astron. J.* 133, 154
- [434] Luck RE & Bond HE 1985, *Ap. J.* 292, 559–77
- [435] Lynden-Bell D & Kalnajs A 1972, *MNRAS* 157, 1–30
- [436] Lynden-Bell D & Lynden-Bell RM 1995, *MNRAS* 275, 429–42
- [437] Mac Low M & Ferrara A 1999, *Ap. J.* 513, 142–55
- [438] Madau P, Pozetti L & Dickinson M 1998, *Ap. J.* 498, 106
- [439] Maeda K et al 2002, *Ap. J.* 565, 405
- [440] Majewski SR, Hawley SL & Munn JA 1996, In *Formation of the Halo ... Inside & Out* eds. H Morrison, A Sarajedini, *Astron. Soc. Pac.* 92, 119–129
- [441] Majewski SR, Ostheimer JC, Kunkel WE & Patterson RJ 2000, *Astron. J.* 120, 2550–68
- [442] Majewski SR 1993, *Annu. Rev. Astron. Astrophys.* 31, 575–638
- [443] Malin DF & Carter D 1980, *Nature* 285, 643–45
- [444] Malin DF & Hadley B 1997, *PASA* 14, 52–58
- [445] Maller AH & Bullock JS 2004, *MNRAS* 355, 694
- [446] Maloney PR 1993, *Ap. J.* 414, 41–56
- [447] Maloney PR & Bland-Hawthorn J 2001, *Ap. J.* 553, L129
- [448] Mandelbaum R et al 2006, *MNRAS* 368, 715
- [449] Manuel O 2000, *Origins of Elements in the Solar System*, (New York: Kluwer)
- [450] Marquez A & Schuster WJ 1994, *Astron. Astrophys. Suppl.* 108, 341–58
- [451] Marinacci F et al 2011, *MNRAS* 415, 1534
- [452] Martinez-Delgado D et al 2010, *Astron. J.* 140, 962–967
- [453] Masiropietro C et al 2005, *MNRAS* 363, 509
- [454] Mateo M. 1998, *Annu. Rev. Astron. Astrophys.* 36, 435–506
- [455] Mathews GJ, Bazan G & Cowan JJ 1992, *Ap. J.* 391, 719–735
- [456] Matsuda Y et al 2005, *Ap. J.* 634, L125
- [457] Matteucci F & Francois P 1989, *MNRAS* 239, 885–904
- [458] Matthews LD, Gallagher JS & van Driel W 1999, *Astron. J.* 118, 2751
- [459] Matzner CD 2007, *Ap. J.* 659, 1394

- [460]McCarthy IG et al 2008, *MNRAS* 383, 593
- [461]McConnachie A 2012, *Astron. J.* 144:4, 1
- [462]McConnachie A et al 2009, *Nature* 461, 66
- [463]McGaugh SS, Schombert JM, Bothun GD & de Blok WJG 2000, *Ap. J.* 533, L99–L102
- [464]McGaugh SS, Schombert JM, de Blok WJG & Zagursky MJ 2010, *Ap. J.* 708, L14–L17
- [465]McKee CF & Ostriker EC 2007, *Annu. Rev. Astron. Astrophys.* 45, 565
- [466]McKee CF & Tan JC 2002, *Nature* 416, 59–61
- [467]McKellar A 1940, *PASP* 52, 187
- [468]McWilliam A, Preston GW, Sneden C & Searle L 1995, *Astron. J.* 109, 2736–2756
- [469]McWilliam A & Rich RM 1994, *Ap. J. Suppl.* 91, 749–91
- [470]Melendez L et al 2008, *Astron. Astrophys.* 484, L21
- [471]Metcalfe RB 2001, *astro-ph/0109347*
- [472]Meusinger H, Stecklum B & Reimann H-G 1991, *Astron. Astrophys.* 245, 57–74
- [473]Mitschang AW et al 2013, *MNRAS* 428, 2321
- [474]Miyamoto M & Nagai R 1975, *PASJ* 27, 533
- [475]Mo H, van den Bosch FC & White SDM 2010, *Galaxy Formation & Evolution* (Cambridge University Press)
- [476]Moore CE, Minnaert MGJ & Houtgast J 1966, *The Solar Spectrum 2935 Å to 8770 Å* (NBS Monogr. 61) (Washington: NBS)
- [477]Molla M, Ferrini F & Diaz A 1996, *Ap. J.* 466, 668–85
- [478]Moore B, Katz N & Lake G 1996, *Ap. J.* 457, 455–59
- [479]Moore B et al 1999, *Ap. J.* 524, L19–22
- [480]Moore B et al 2001, *Phys. Rev. D* 64, 063508–19
- [481]Moore B et al 2006, *Ap. J.* 368, 563
- [482]Morgan WW 1959, *Astron. J.* 64, 432
- [483]Morrison HL 1993, *Astron. J.* 106, 578–90
- [484]Morrison HL et al 2009, *Ap. J.* 694, 130
- [485>Mortlock D et al 2011, *Nature* 474, 616
- [486]Mueller KE et al 2002 *Ap. J. Suppl.* 143, 469
- [487>Muldrew S et al 2012, *MNRAS* 419, 2670
- [488]Munari U et al 2005, *Astron. Astrophys.* 442, 1127
- [489]Munari U et al 2008, *Astron. Astrophys.* 488, 969
- [490]Murray SD & Lin DNC 1990, *Ap. J.* 363, 50
- [491]Murray SD & Lin DNC 2004, *Ap. J.* 615, 586
- [492]Mushotzky R 1999, In *The Hy-Redshift Universe*, eds. AJ Bunker & WJM van Breugel, *Astron. Soc. Pac.* 193, 323–335
- [493]Nakamura F & Li Z-Y 2007, *Ap. J.* 662, 395
- [494]Navarro JF, Frenk CS & White SDM 1997 *Ap. J.* 490, 493
- [495>Ness M et al 2012, *Ap. J.* 756, 22
- [496>Ness M et al 2013, *MNRAS* 430, 836
- [497>Newberg HJ et al 2001, *Ap. J.* 569, 245
- [498]Neyman J, Page T & Scott E 1961, *Astron. J.* 66, 633
- [499>Nichols M & Bland-Hawthorn J 2009, *Ap. J.* 707, 1642
- [500>Nichols M & Bland-Hawthorn J 2011, *Ap. J.* 732, 17
- [501>Nichols M & Bland-Hawthorn J 2013, *Ap. J.* in press
- [502>Nissen PE & Schuster WJ 2012, *Astron. Astrophys.* 543, 28
- [503>Noguchi M 1998, *Nature* 392, 253–55
- [504>Nomoto K 1987, *Ap. J.* 322, 206
- [505>Nomoto K et al 2005, *Nucl. Phys. A* 758, 263
- [506>Norberg P et al 2001, *MNRAS* 328, 64–70
- [507>Nordström B et al 2004, *A&A*, 418 919
- [508>Norman ML 2011, *IAU Symp.* 270, 7
- [509>Norris JE, Ryan SG, Beers TC 1996, *Ap. J. Suppl.* 107, 391–421
- [510>Norris JE, Ryan SG, Beers TC 1997, *Ap. J.* 488, 350–63
- [511>Norris JE, Ryan SG 1989, *Ap. J.* 336, L17–19
- [512>Norris JE et al 2007, *Ap. J.* 670, 774
- [513>Norris JE, Yong D, Gilmore G & Wyse RFG 2010a, *Ap. J.* 711, 150
- [514>Norris JE et al 2010b, *Ap. J.* 723, 1632
- [515>OShea BW & Norman ML 2007, *Ap. J.* 654, 66
- [516>OShea BW & Norman ML 2008, *Ap. J.* 673, 14
- [517>Odenkirchen M et al 2001, *Ap. J.* 548, L165–69
- [518>Okrochov M & Tumlinson J 2010, *Ap. J.* , 716, L41
- [519>Oort J 1965, In *Galactic Structure*, eds. A. Blaauw, M. Schmidt (University of Chicago Press), 455–511
- [520>Oort J 1966, *Bull. Astron. Inst. Neth.* 18, 421
- [521>Ortolani S et al 1995, *Nature* 377, 701–703
- [522>Ostriker JP & Peebles PJE, 1973 *Ap. J.* 186, 467
- [523>Ostriker JP, Peebles PJE & Yahil A 1974, *Ap. J.* 193, L1
- [524>Oswald TD, Smith JA, Wood MA & Hintzen P 1996, *Nature* 382, 692
- [525>Page BEJ 1965, *Roy. Obs. Bull.* 104, 127–51
- [526>Page BEJ 1989, In *Evolutionary Phenomena in Galaxies* (Cambridge University Press), 201–223
- [527>Page BEJ & Patchett BE 1975, *MNRAS* 172, 13–40
- [528>Page BEJ & Tautvaisiene G 1997, *MNRAS* 288, 108–116
- [529>Panter B, Jimenez R, Heavens AF & Charlot S 2007, *MNRAS* 378, 1550
- [530>Peacock JA 1999, *Cosmological Physics* (Cambridge University Press)
- [531>Peacock JA et al 2001, *Nature* 410, 169
- [532>Peacock JA 2013, *astro-ph/1301.7286*
- [533>Peebles PJE & Dicke RH 1968, *Ap. J.* 154, 891–908
- [534>Peebles PJE 1970, *Astron. J.* 75, 13–20
- [535>Peebles PJE 1971, *Physical Cosmology* (Princeton University Press)
- [536>Peebles PJE 1974, *Ap. J.* 189, L51–53
- [537>Peebles PJE 1993, *Principles of Physical Cosmology* (Princeton University Press)
- [538>Peebles PJE, Seager S & Hu W 2000, *Ap. J.* 539, L1–4
- [539>Peletier R & de Grijs R 1998, *MNRAS* 300, L3–6
- [540>Penton S, Stocke JT & Shull JM 2004, *Ap. J. Suppl.* 152, 29
- [541>Penzias A & Wilson R 1965, *Ap. J.* 142, 419
- [542>Percival W et al 2001, *MNRAS* 327, 1297
- [543>Perlmutter S et al 1998, *Nature* 391, 51
- [544>Peterson JR et al 2003, *Ap. J.* 590, 207
- [545>Pettini M et al 2003, *Ap. J.* 594, 695
- [546>Pfenniger D, Combes F & Martinet L 1994, *Astron. Astrophys.* 285, 79–93
- [547>Pflamm-Altenburg J & Kroupa P 2007, *MNRAS* 375, 855
- [548>Phelps R & Janes K 1996, *Astron. J.* 111, 1604–08
- [549>Pichon C et al 2011, *MNRAS* 418, 2493
- [550>Pitts E & Tayler RJ 1989, *MNRAS* 240, 373–395
- [551>Plume R et al 1997, *Ap. J.* 476, 730
- [552>Podsiadlowski P et al 2002, *Ap. J.* 567, 491
- [553>Pohlen M, Dettmar R-J & Lüticke R 2000, *Astron. Astrophys.* 357, L1–L4
- [554>Pohlen M et al 2008, *Astron. Soc. Pac.* 390, 247
- [555>Pont F & Eyer L 2004, *MNRAS* 351, 487
- [556>Portnoy D, Pistinner S & Shaviv G 1993, *Ap. J. Suppl.* 86, 95
- [557>Prada F et al 2012, *MNRAS* 423, 3018
- [558>Prescott MKM et al 2013, *Ap. J.* in press
- [559>Press WH & Schechter P 1974, *Ap. J.* 187, 425–438
- [560>Preston GW, Beers TC & Shectman SA 1994, *Astron. J.* 108, 538–554
- [561>Pritchett C & van den Bergh S 1994, *Astron. J.* 107, 1730–1736
- [562>Prochaska JX, Naumov SO, Carney BW, McWilliam A & Wolfe AM 2000, *Astron. J.* 120, 2513–2549
- [563>Purcell S, Bullock JS & Kaplinghat M 2009, *Ap. J.* 703, 2275
- [564>Putman ME et al 2003, *Ap. J.* 597, 948
- [565>Putman ME, Peek JEG & Jounge MR 2012, *Annu. Rev. Astron. Astrophys.* 50, 491
- [566>Quillen AC & Garnett D. 2001 In *Galaxy Disks and Disk Galaxies*, eds. G Jose, SJ Funes and EM Corsini, *Astron. Soc. Pac.* 230, 87–88
- [567>Quillen AC 2002, *astro-ph/0202253*
- [568>Quinn PJ 1984, *Ap. J.* 279, 596–609
- [569>Quinn PJ & Goodman J 1986, *Ap. J.* 309, 472–495
- [570>Quinn PJ, Hernquist L & Fullagar D 1993, *Ap. J.* 403, 74–93
- [571>Quinn PJ & Zurek W 1988, *Ap. J.* 331, 1–18
- [572>Quinn T, Katz N & Efstathiou GP 1996, *MNRAS* 278, 49–54
- [573>Rafelski M et al 2012, *Ap. J.* 755, 89
- [574>Ramirez I et al 2010, *Astron. Astrophys.* 521, 33
- [575>Rauch M et al 2005, *Ap. J.* 632, 58
- [576>Rauch M et al 2011, *MNRAS* 418, 1115
- [577>Rauscher T, Heger A, Hoffman RD & Woosley SE 2002, *Ap. J.* 576, 323
- [578>Recio-Blanco, A et al 2006, *MNRAS* 370, 141
- [579>Rees MJ 1986, *MNRAS* 218, 25–30
- [580>Rees MJ & Ostriker JP 1977, *MNRAS* 179, 541
- [581>Reid IN 1998, In *Highlights of Astronomy*, ed. J Andersen, 11A, 562 (Kluwer)
- [582>Reid IN et al 2007, *Ap. J.* 665, 767
- [583>Renzini A 2000, In *From Extrasolar Planets to Cosmology*, eds. J Bergeron, A Renzini (Berlin: Springer)

- [584]Renzini A 2002 In *Chemical Enrichment of Intracluster and Intergalactic Medium*, eds. F Matteucci and R Fusco-Femiano, *Astron. Soc. Pac.* 253, 331
- [585]Ribaudo J et al 2011 *Ap. J.* 743, 207
- [586]Rich RM 2001, In *Astrophysical Ages and Time Scales*, eds. T von Hippel, C Simpson & N Manset, *Astron. Soc. Pac.* 245, 216–25
- [587]Richardson J et al 2011, *Ap. J.* 732, 76
- [588]Riess A et al 1998, *Astron. J.* 116, 1009
- [589]Rix H-W & Bovy J 2012, *Astron. Astrophys. Rev.* in press
- [590]Rix H-W & Zaritsky D 1995, *Ap. J.* 447, 82–102
- [591]Roberts MS 1967, *IAU Symp.* 31, 189
- [592]Roberts MS 1975, *Ap. J.* 201, 327
- [593]Roberts MS 1976, *Comm. Astrophys.* 6, 105
- [594]Roberts MS & Rots AH 1973, *Astron. Astrophys.* 26, 483
- [595]Robertson B et al 2010, *Nature* 468, 49
- [596]Robin A & Creze M 1986, *Astron. Astrophys.* 157, 71
- [597]Robin A et al 2003, *A&A* 409 523
- [598]Robotham A et al 2012, *MNRAS* 424, 1448
- [599]Rocha-Pinto HJ et al 2000, *Ap. J.* 531, L115
- [600]Rocha-Pinto HJ, Scalo J, Maciel WJ & Flynn C 2000, *Astron. Astrophys.* 358, 869–85
- [601]Rodgers AW, Harding P & Sadler EM 1981, *Ap. J.* 244, 912–18
- [602]Rood HJ, Page TL, Kintner EC & King IR 1972, *Ap. J.* 175, 627
- [603]Rosdahl J & Blaizot J 2012 *MNRAS* 423, 344
- [604]Rose J 1994, *Astron. J.* 107, 206–29
- [605]Roskar R et al 2008a, *Ap. J.* 675, L65
- [606]Roskar R et al 2008b, *Ap. J.* 684, L79
- [607]Roth K et al 1993, *Ap. J.* 413, 67
- [608]Rubin V & Ford K 1970, *Ap. J.* 159, 379
- [609]Rubin V, Thonnard N & Ford WK 1978, *Ap. J.* 225, L107
- [610]Rubin KHR et al 2012, *Ap. J.* 747, L26
- [611]Ryan SG, Norris JE & Beers TC 1996, *Ap. J.* 471, 254–78
- [612]Ryan Weber E, Pettini M & Madau P 2006, *MNRAS* 371, L78
- [613]Ryan Weber E, Pettini M, Madau P & Zych BJ 2009, *MNRAS* 395, 1476
- [614]Salvadori S & Ferrara A 2009, *MNRAS* 395, L6
- [615]Sanchez-Gil, C et al 2011, *MNRAS* 415, 753
- [616]Sancisi R et al 2008, *Astron. Astrophys. Rev.* 15, 189
- [617]Sandage A 1990, *JRASC* 84, 70–88
- [618]Sandage A & Cacciari C 1990, *Ap. J.* 350, 645–661
- [619]Sandage A & Visvanathan N 1978, *Ap. J.* 225, 742–750
- [620]Sarajedini A et al 2007, *Astron. J.* 133, 1658
- [621]Schaye J 2004, *Ap. J.* 609, 667
- [622]Schlegel DJ et al 1998, *Ap. J.* 500, 525
- [623]Schmidt M 1963, *Ap. J.* 137, 758–69
- [624]Schneider R & Omukai K 2010, *MNRAS* 402, 429
- [625]Schönmakers RHM, Franx M & de Zeeuw PT 1997, *MNRAS* 292, 349–64
- [626]Schommer RA, Suntzeff NB, Olszewski EW & Harris HC 1992, *Astron. J.* 103, 447–59
- [627]Schönrich RA, Binney JJ & Asplund MA 2012, *MNRAS* 420, 1281
- [628]Schuster WJ et al 2012, *Astron. Astrophys.* 538, 21
- [629]Schweizer F 1987, In *Nearly Normal Galaxies – From the Planck Time to the Present*, ed. S Faber, 18–25 (New York: Springer-Verlag)
- [630]Scott D & Rees MJ 1990, *MNRAS* 247, 510
- [631]Searle L & Zinn R 1978, *Ap. J.* 225, 357–79
- [632]Searle L 1977, In *The Evolution of Galaxies & Stellar Populations*, eds. BM Tinsley, RB Larson, 219 (New Haven, Yale)
- [633]Sedgewick R 1992, *Algorithms in C++* (Addison-Wesley)
- [634]Sellwood JA 1999, In *Astrophysical Disks*, eds. J Sellwood & J Goodman, 327
- [635]Sellwood JA & Binney JJ 2002, *MNRAS* 336, 785
- [636]Sembach KR et al 2003, *Ap. J. Suppl.* 146, 165
- [637]Shang Z et al 1998, *Ap. J.* 504, L23–26
- [638]Shapiro PR & Kang H 1987, *Ap. J.* 318 32
- [639]Shapiro PR et al 2006, *Ap. J.* 646, 681
- [640]Sharma S, Bland-Hawthorn J, Johnston K & Binney JJ 2011, *Ap. J.* 730, 3
- [641]Sharma S, Steinmetz M & Bland-Hawthorn J 2012, *Ap. J.* 750, 107
- [642]Sharma S & Bland-Hawthorn J 2013, *Ap. J.* in press
- [643]Sharma S et al 2013, *Ap. J.* in press
- [644]Sheth RK & Tormen G 1999, *MNRAS* 308, 119–126
- [645]Sheth RK & Tormen G 2001, *MNRAS* 323, 1
- [646]Shigeyama T & Tsujimoto T 1998, *Ap. J.* , 507, L135–39
- [647]Shima E et al 1985, *MNRAS* 217, 367
- [648]Shirley YL et al 2003, *Ap. J. Suppl.* 149, 375
- [649]Shull JM, Smith BD & Danforth CW 2012, *Ap. J.* 759:23
- [650]Siebert A et al 2011, *Astron. J.* 141, 187
- [651]Silk J 1977, *Ap. J.* 211, 638
- [652]Silk J & Mamon GA 2012, *Res. Astron. Astrophys.* 12, 917–946
- [653]Simon JD & Geha M 2007, *Ap. J.* 670, 313
- [654]Slipher VM 1913, *Low. Obs. Bull.*, 2, 56
- [655]Slipher VM 1914, *Low. Obs. Bull.*, 2, 66
- [656]Smith S 1936, *Ap. J.* 83, 23
- [657]Smith MC et al 2007, *MNRAS* 379, 755
- [658]Snedden C et al 2000, *Ap. J.* 533, L139–L142
- [659]Soderblom D 2010, *Annu. Rev. Astron. Astrophys.* 48, 581
- [660]Soubiran C & Girard P 2005, *Astron. Astrophys.* 438, 139
- [661]Spergel D & Steinhardt P 2000, *Phys. Rev. Lett.* 84, 3760–63
- [662]Spiegel EA 1970, *IAU Symp.* 39, 201
- [663]Spite M & Spite F 1978, *Astron. Astrophys.* 67, 23–31
- [664]Spitzer L & Schwarzschild M 1953, *Ap. J.* 118, 106–12
- [665]Springel V et al 2005, *Nature* 435, 629
- [666]Springel V et al 2008, *MNRAS* 391, 1685
- [667]Stacy A, Greif TH & Bromm V 2010, *MNRAS* 403, 45
- [668]Stacy A, Greif TH & Bromm V 2012, *MNRAS* 422, 290
- [669]Stadel J et al 2009, *MNRAS* 398, L21
- [670]Stahler SW, Palla F & Ho PTP 2000, In *Protostars and Planets IV*, eds. V Mannings, AP Boss & SS Russell, 327–351 (University of Arizona Press)
- [671]Starkenburger E et al 2010, *Astron. Astrophys.* 513, 34
- [672]Steidel CC et al 2011, *Ap. J.* 736, 160
- [673]Sternberg A, McKee CF & Wolfire MG 2002, *Ap. J. Suppl.* 143, 419
- [674]Sternberg A & Soker N 2008, *MNRAS* 389, L13
- [675]Strateva I et al 2001, *Astron. J.* 122, 1861
- [676]Strömgren B 1987, In *The Galaxy: Proceedings of the NATO ASI*, eds. G Gilmore, R Carswell, 229–246 (Reidel)
- [677]Sun XH et al 2008 *Astron. Astrophys.* 477, 573
- [678]Sutherland RS & Dopita MA 1993, *Ap. J. Suppl.* 88, 253
- [679]Tabor G & Binney JJ 1993, *MNRAS* 263, 323
- [680]Talbot RJ & Arnett WD 1971, *Ap. J.* 170, 409–422
- [681]Tamburro D et al 2008, *Astron. J.* 136, 2872
- [682]Tamm A et al 2012, *Astron. Astrophys.* 546, 4
- [683]Tan JC, Krumholz MR & McKee CF 2006, *Ap. J.* 641, L121
- [684]Tanvir N et al 2009, *Nature* 461, 1254
- [685]Taylor BJ 2000, *Astron. Astrophys.* 362, 563–579
- [686]Thoul A & Weinberg D 1996, *Ap. J.* 465, 608–16
- [687]Ting YS et al 2012, *MNRAS* 421, 1231
- [688]Tinsley B 1980, *Fund. Cos. Phys.* 5, 287–388
- [689]Tolstoy E, Hill V & Tosi M 2009, *Annu. Rev. Astron. Astrophys.* 47, 371
- [690]Trager S, Faber SM, Worthey G & Gonzalez J 2000, *Astron. J.* 120, 1645–76
- [691]Trager S, Worthey G, Faber SM, Bustein D & Gonzales J 1998, *Ap. J. Suppl.* 116 1–28
- [692]Travaglio C, Gallino R, Zinner E, Amari S & Woosley S 1998, *Meteoritics & Planetary Science* 33, A155–156
- [693]Travaglio C et al 1999, *Ap. J.* 521, 691–702
- [694]Tremaine SD 1993, In *Back to the Galaxy*, eds. SS Holt, F. Verter (New York: AIP), 599–609
- [695]Tremaine SD 1999, *MNRAS* 307, 877–83
- [696]Tremaine SD 1999, *Ap. J.* 525, 1223 (Comment)
- [697]Tremaine SD & Gunn JE 1979, *Phys. Rev. Lett.* 42, 407–410
- [698]Tripicco MJ & Bell RA 1995, *Astron. J.* 110, 3035–3049
- [699]Truran JW, Burles S, Cowan J & Sneden C 2001, In *Astrophysical Ages and Time Scales*, eds. T von Hippel, C Simpson & N Manset, *Astron. Soc. Pac.* 245, 226–234
- [700]Truran JW 1981, *Astron. Astrophys.* 97, 391–93
- [701]Tsikoudi V 1980, *Ap. J. Suppl.* 43, 365–77
- [702]Tsujimoto T & Shigeyama T 1998, *Ap. J.* 508, L151
- [703]Tsujimoto T, Shigeyama T & Yoshii Y 2000, *Ap. J.* 531, L33–36
- [704]Tsuribe T & Omukai K 2006, *Ap. J.* 642, L61
- [705]Tsuribe T & Omukai K 2008, *Ap. J.* 676, L45
- [706]Tully RB 1987, *Catalogue of Nearby Galaxies* (University of Hawaii Press)
- [707]Tully RB & Fisher JR 1977, *Astron. Astrophys.* 54, 661–673
- [708]Tully RB, Verheijen MA, Pierce MJ, Huang J-S & Wainscoat RJ 1996, *Astron. J.* 112, 2471–2499

- [709] Tully RB et al 2008, *Ap. J.* 676, 184
 [710] Tumlinson J 2010, *Ap. J.* 708, 1398
 [711] Turk MJ, Abel T & OShea B 2009, *Science* 325, 601
 [712] Turner E 1976a, *Ap. J.* 208, 20
 [713] Turner E 1976b, *Ap. J.* 208, 304
 [714] Turoc C & Primas F 2008, ESA-ESO WG Report No. 4: *Galactic Populations, Chemistry & Dynamics* (ESO: Garching)
 [715] Twarog BA, Ashman KM & Anthony-Twarog BJ 1997, *Astron. J.* 114, 2556–2585
 [716] Twarog BA 1980, *Ap. J.* 242, 242–259
 [717] Tyson JA et al 1998, *Astron. J.* 116, 102–110
 [718] Ulrich RK 1986, *Ap. J.* 306, L37–40
 [719] Umeda H & Nomoto K 2003, *Nature* 422, 871
 [720] Unavane M, Wyse RFG & Gilmore G 1996, *MNRAS* 278, 727–36
 [721] van Albada TS, Bahcall JN, Begeman K & Sancisi R 1985, *Ap. J.* 295, 305
 [722] van den Bergh S 1962, *Astron. J.* 67, 486–90
 [723] van den Bergh S 2000, *The Galaxies of the Local Group* (Cambridge University Press)
 [724] van den Bergh S & McClure RD 1980, *Astron. Astrophys.* 88, 360–62
 [725] van der Kruit P 1979, *Astron. Astrophys. Suppl.* 38, 15
 [726] van der Kruit P 2002, In *The Dynamics, Structure & History of Galaxies – Workshop in Honour of KC Freeman*, eds. G Da Costa & H Jerjen, 7
 [727] Vanzella E et al 2011, *Ap. J.* 730, 35
 [728] Vazdekis A 1999, *Ap. J.* 513, 224–41
 [729] Veilleux S, Cecil G & Bland-Hawthorn J 2005, *Annu. Rev. Astron. Astrophys.* 43, 769
 [730] Venn K et al 2004 *Astron. J.* 128, 1177
 [731] Vivas AK et al 2001, *Ap. J.* 554, L33–36
 [732] Wakker BP 2001, *Ap. J. Suppl.* 136, 463
 [733] Walker IR, Mihos JC & Hernquist L 1996, *Ap. J.* 460, 121–35
 [734] Wallace L, Hinkle K & Livingston W 1998, *An atlas of the spectrum of the solar photosphere (3570–7405Å)* (Tucson: NOAO)
 [735] Wallerstein G, Greenstein JL, Parker R, Helfer HL & Aller LH 1963, *Ap. J.* 137, 280–300
 [736] Wallerstein G et al 1997, *Rev. Mod. Phys.* 69, 995–1084
 [737] Walsh SM, Jerjen H & Willman B 2007, *Ap. J.* 662, L83
 [738] Weil M, Bland-Hawthorn J & Malin DF 1997, *Ap. J.* 490, 664–81
 [739] Weinberg S 1977, *The First Three Minutes: A Modern View of the Origin of the Universe* (London: Andre Deutsch)
 [740] Weiner BJ, Vogel SN & Williams TB 2001, In *Extragalactic Gas at Low Redshift*, eds. J. Mulchaey & J. Stocke, *Astron. Soc. Pac.* 254, 256–266
 [741] Weinmann SM et al 2013, *MNRAS* in press (1204.4184)
 [742] Westin J, Sneden C, Gustafsson B & Cowan J 2000, *Ap. J.* 530, 783–99
 [743] White M & Croft RAC 2000, *Ap. J.* 539, 497–504
 [744] White SDM 1976, *MNRAS* 177, 717–33
 [745] White SDM 1981, *MNRAS* 195, 1037
 [746] White SDM 1983, *Ap. J.* 274, 53
 [747] White SDM 1984, *Ap. J.* 286, 38
 [748] White SDM 1996, In *Cosmology & Large Scale Structure*, eds. R Schaeffer, J Silk, M Spiro & J Zinn-Justin, 349 (Amsterdam: Elsevier)
 [749] White SDM & Rees MJ 1978, *MNRAS* 183, 341–58
 [750] White SDM et al 1993, *Nature* 366, 429
 [751] Whitmore BC, Schweizer F, Leitherer C, Borne K & Robert C 1993, *Astron. J.* 106, 1354–70
 [752] Whitmore BC & Schweizer F 1995, *Astron. J.* 109, 960–80
 [753] Whitmore B, Chandar R & Fall SM 2007, *Astron. J.* 133, 1067
 [754] Widrow L & Dubinski J 1998, *Ap. J.* 504, 12–26
 [755] Williams JP & McKee CF 1997, *Ap. J.* 476, 166
 [756] Williams MEK et al 2011, *Ap. J.* 728, 102
 [757] Williams BF et al 2013, *Ap. J.* in press
 [758] Wolfe AM, Gawiser E & Prochaska JX 2006, *Annu. Rev. Astron. Astrophys.* 43, 861
 [759] Woosley SE & Weaver TA 1995, *Ap. J. Suppl.* 101, 181
 [760] Worthey G, Dorman B & Jones L 1996, *Astron. J.* 112, 948–53
 [761] Worthey G, Faber SM, Gonzales J & Burstein D 1994, *Ap. J. Suppl.* 94, 687–722
 [762] Worthey G 1994, *Ap. J. Suppl.* 95, 107–149
 [763] Wright EL 2003, *New Astron. Rev.* 47, 877–881
 [764] Wright EL 2006, *PASP* 118, 1711
 [765] Wylie de Boer L et al 2012, *Ap. J.* 755, 35
 [766] Wyse RFG 2001, In *Galaxy Disks and Disk Galaxies*, eds. JG Funes, SJ & EM Corsini, *Astron. Soc. Pac.* 230, 71
 [767] Wyse RFG & Silk J 1989, *Ap. J.* 339, 700–711
 [768] Xie T, Allen M & Langer WD 1995, *Ap. J.* 440, 674
 [769] Yepes G, Gottlöber S, Martínez-Vaquero LA & Hoffman Y 2009, *AIP Conf.* 1115, 80–91
 [770] Yokoi K, Takahashi K & Arnould M 1983, *Astron. Astrophys.* 117, 65–82
 [771] Yoshida N, Bromm V & Hernquist L 2004, *Ap. J.* 605, 579
 [772] Yoshida N, Oh SP, Kitayama T & Hernquist L 2007, *Ap. J.* 663, 687
 [773] Yoshida N, Omukai K & Hernquist L 2008, *Science* 321, 669
 [774] Yoshii Y & Peterson BA 1995, *Ap. J.* 444, 15
 [775] Younger JD et al 2007, *Ap. J.* 671, 1531
 [776] Yun MS, Carilli CL, Kawabe R, Tutui Y, Kohno K & Ohta K 2000, *Ap. J.* 528, 171–178
 [777] Zheng Z et al 1999, *Astron. J.* 117, 2757–2780
 [778] Ziegler BL et al 2005, 4433, 519
 [779] Zinn R 1985, *Ap. J.* 293, 424–444
 [780] Zinnecker H & Cannon RD 1986, In *Star Forming Galaxies and Related Objects*, eds. D Kunth, TX Thuan, J Thanh Van, 155 (Paris: Editions Frontières)
 [781] Zwicky F 1933, *Helvetica Physica Acta*, 6, 110–127
 [782] Zucker DB et al 2004, *Ap. J.* 612, L117
 [783] Zucker DB et al 2006, *Ap. J.* 650, L41
 [784] Zucker DB et al 2007, *Ap. J.* 659, L21
 [785] Zurek W, Quinn PJ & Salmon J 1988, *Ap. J.* 330, 519–34