Origins of Stellar Halos

Kathryn V. Johnston
Department of Astronomy, Columbia University
e-mail: kvj@astro.columbia.edu

Abstract. This contribution reviews ideas about the origins of stellar halos. It includes discussion of the theoretical understanding of and observational evidence for stellar populations formed “in situ” (meaning formed in orbits close to their current ones), “kicked-out” (meaning formed in the inner galaxy in orbits unlike their current ones) and “accreted” (meaning formed in a dark matter halo other than the one they currently occupy). At this point there is general agreement that a significant fraction of any stellar halo population is likely “accreted”. There is modest evidence for the presence of a “kicked-out” population around both the Milky Way and M31. Our theoretical understanding of and the observational evidence for an “in situ” population are less clear.

Keywords. galaxies: evolution — galaxies: formation — galaxies: halos — galaxies: structure

1. Introduction

Stellar halos contain only of order 1% of the stars in a typical disk galaxy, spread out over volumes ∼ 10^6 larger than the bulk of the other components. While the stellar halo of the Milky Way has been studied for decades using local samples or tracers at larger distances (e.g. Eggen et al. (1962), Searle & Zinn (1978)), the low density meant that it was not possible for a long time to detect these components around other galaxies. Interest in these diffuse galactic components really gained momentum with the advent of large-area, all sky surveys (SDSS and 2MASS in particular) which enabled star-count studies to push their sensitivity to sufficiently low surface-brightness (below 30 mag/arcsec^2) that the global structure of and substructure within the Milky Way could be studied (e.g. Newberg et al. (2002), Majewski et al. (2003)). At the same time, dedicated surveys have mapped Andromeda’s halo (e.g. Ibata et al. (2007)). These studies have revealed both halos to extend beyond 100kpc from their centers with rich substructure in the form of debris from past accretion events throughout (Newberg et al. (2002), Majewski et al. (2003), Belokurov et al. (2006), Ibata et al. (2007)).

It is easy to think of these diffuse components around galaxies as pretty (see surface brightness projections of simulations of accreted halos in left hand panels of Figure 1), but inconsequential to our broader interests of understanding structures in the Universe and how they form since they contain such a tiny fraction of the total mass. However, the existence of coherent debris-structures within halos has several interesting consequences. First, it provides dramatic confirmation of the widely accepted paradigm that galaxies form hierarchically. Second, the debris-structures themselves contain signatures (in their phase-space distributions and stellar populations) of the objects that have been destroyed and hence provide window into the nature of smaller galaxies in the past — in some cases to mass scales and epochs that cannot be directly observed at higher redshift. Third, the collective interpretation of all the debris-structures detectable within a halo can give some constraints on the accretion history of the parent galaxy. Fourth, the common origin of stars within an individual debris-structure can be exploited to provide very sensitive constraints on the distribution of mass in the dark matter halo within which...
Figure 1. Surface brightness, line-of-sight velocity and line-of-sight velocity dispersion projections of a simulation of a purely accreted stellar halo taken from Bullock & Johnston (2005). Each box is 300kpc on a side. The central galaxy, containing ∼99% of the light, is not shown. (Image credit: Sanjib Sharma)

it is orbiting — both the global dark matter distribution (Johnston et al. (1999)) and the dark matter subhalos that may also be present (Ibata et al. (2002), Johnston et al. (2002), Carlberg (2009)).

While the prospects in the previous paragraph are intriguing, the degree to which they can be applied to populations around a galaxy depends on what fraction of the stellar halo is truly accreted. The current observations suggest that accretion is responsible for the formation outskirts of the Milky Way’s and Andromeda’s stellar halos. However, it is unclear how dominant accretion is more generally in the formation of the inner parts of these halos and for other galaxies. This review will examine our understanding of the origins of stellar halos, concentrating on what is known from theoretical and observational perspectives about Milky Way like spiral galaxies.

2. Stellar Halo Formation Scenarios

For the purposes of this review, stellar halo formation scenarios will be split into three categories.

In situ populations are stars that were born in the stellar halo component and are still on the same (or at least, very similar) orbits that they were born on (as originally postulated by Eggen et al. (1962)).

Kicked-out populations are stars that were born in the parent dark matter halo, but within a different stellar component (i.e. the central disk or bulge) than they are now currently: their orbits have evolved significantly (Purcell et al. (2010)).

Accreted populations are stars that were born in dark matter halos other than that of the parent galaxy which is now their host (as first put forward by Searle & Zinn (1978)).

Note the above scenarios were chosen in particular because they broadly separate what are likely to be different conditions under which star formation itself occurred, for example between different depths of dark matter potential and, as a consequence, influences on stellar enrichment (importance of feedback and/or accretion of gas). Hence the resulting populations may be currently distinguishable either by their orbits, their abundances or a combination of the two. However, a variety of categories have been used recently in the literature, so the reader is cautioned to be careful in comparing papers. For example: some authors broadly refer to both in situ and kicked-out stars as in situ...
since both populations are born in the same parent potential (e.g. Zolotov et al. (2009)); others have separated accreted populations into stars formed in a destroyed dwarf prior to and after its accretion by the parent galaxy (the “endo-debris” stars described in Tissera et al. (2013)), which is not a distinction made in this review; and still others have simply referred to stars born in satellites as ex-situ rather than accreted (Pillepich et al. (2015)).

There is little theoretical guidance as to which of the above formation scenarios is expected to be dominant. Many models in the literature simply explore the observable consequences for a stellar halo that is assumed to have formed entirely by accretion using collisionless N-body simulations (Bullock & Johnston (2005), De Lucia & Helmi (2008), Cooper et al. (2010)). The advantage of these models is that they are computationally inexpensive and physically simple so it is possible to design them to resolve even small infalling dwarf galaxies. However, no stars in these models are formed in a self-consistent manner, and the contributions of the stellar components of the parent galaxy to the potential are often not considered.

More rigorous explorations of the full range of formation scenarios requires the use of hydrodynamical simulations, which are much more computationally expensive and are only now reaching resolutions necessary to look at the global structure and, in particular, substructures within stellar halos. From the first such studies that attempted to look at these components (Abadi et al. (2006)), most have agreed that the simulated stellar halos were formed through multiple mechanisms, in particular both via kicking-out populations from the inner galaxy and via accretion (Zolotov et al. (2009), McCarthy et al. (2012), Tissera et al. (2012), Pillepich et al. (2015)). The simulated halos are dominated by accretion in their outskirts (beyond 20-30kpc) but may have significant (or even dominant) kicked-out disk populations within that radius. There is disagreement among the models of whether a true in situ population (as defined in this proceedings) is produced, with only a couple of studies finding any evidence for their presence (Samland & Gerhard (2003), Cooper et al. (2015)). This disagreement highlights the fundamental limit of these models as they all rely on (differing) prescriptions for star formation and feedback below their respective resolutions. Hence, these results should be considered indicative rather than predictive. They nevertheless provide powerful data sets to understand the distinctive signatures that the different formation mechanisms leave in the simulated stellar distributions, and to suggest how the true contribution of stars formed in situ, kicked-out or accreted might be actually assessed from observations (e.g. separating formation mechanisms, as suggested by Zolotov et al. (2010), in the chemical abundance planes).

3. Indicators of Origins

In any discussion of observational signatures of formation mechanism it is important to be clear on the nature of the samples being used, especially when comparing conclusions across different studies.

First, global statements cannot be made without global samples. For example, Solar Neighborhood studies (i.e. within ~ 1 kpc) may not be representative even of the populations in the inner halo due to intrinsic variations around the Galaxy. Moreover, the inner halo (within ~ 10-30 kpc) may have a different dominant formation mechanism than the outer halo.

Second, different stellar populations are expected to be sensitive to different formation mechanisms. For example, studies of M giants (the reddest, most metal-rich giant stars) have found the outer stellar halo to be dominated by a single large accretion event (i.e. the Sagittarius stream, see Majewski et al. (2003)) because only the most massive dwarf
galaxies are metal rich enough to host this population. The bulk of the stellar halo is actually more metal poor than this, so the contribution of large as compared to small accreted dwarfs should really be assessed in other populations as well before a definitive statement can be made. Indeed, stellar population variations for different substructures have been detected in both the Milky Way (Bell et al. (2010)) and Andromeda (Ibata et al. (2014)).

Finally, the ability to distinguish the different formation mechanisms in observations depends on the quality, quantity and dimensions of information available in the sample. The importance of these considerations for stellar halo studies is amply illustrated by considering the variety of samples in a few recent studies. Carollo et al. (2007) studied the kinematics and metallicities of SDSS calibration stars, probing a wide range of stellar metallicities and ages, and a volume up to 20 kpc from the Sun. Deason et al. (2011) analyzed brighter, blue horizontal branch stars in SDSS which could probe the halo to 40 kpc, but with a sample biased towards metal poor populations. Nissen & Schuster (2010) (and subsequently Nissen & Schuster (2011), Schuster et al. (2012), Nissen & Schuster (2012)) used a sample of 94, relatively metal rich (-1.6 < [Fe/H] < -0.4) stars and studied their abundance patterns and full orbital elements in great detail, but their sample was limited to within ~335 pc of the Sun. Sheffield et al. (2012) took high-resolution spectra of M-giant stars selected from the Two Micron All Sky Survey which had significant velocities relative to the disk (hence likely to members of the stellar halo), a tracer only of the highest metallicity halo population and possibly biased towards populations that may have been recently kicked-out from the Galactic disk.

The following subsections briefly outline a few examples of the observational evidence for various formation scenarios, broadly divided by the types and numbers of observed dimensions available, working from the easiest to observe (requiring, at a minimum, projected spatial distributions on the sky) to the most difficult (requiring proper motions and/or high-resolution spectroscopy).

3.1. Spatial Distributions

As noted in Section 1 (and apparent in the left-hand panel of Figure 1), abundant spatial substructure in the outskirts (beyond ~ 10-30kpc) of both the Milky Way and Andromeda galaxies indicate the importance of accretion in forming these galaxies, even suggesting it is dominant in these regions. Several groups are pushing for analogous star count studies around other nearby galaxies using the Hubble Space Telescope (see discussions in these proceedings and Monachesi et al. (2013), Monachesi et al. (2015)), but are necessarily limited in their scope by the finite field-of-view.

Another approach could be to look at the global structure of these halos, averaging over (projected) azimuth to produce radial surface density profiles. In the case of the Milky Way, a break in the power law that best-represents the density profile has been reported (Deason et al. (2014)), while the stellar halo of M31 seems to be well-represented by a single power law (Gilbert et al. (2012), Ibata et al. (2014)). It is tempting to attribute breaks as a transition in the physical formation mechanism, e.g. from dominance by a kicked-out to an accreted population. However, the robustness of this interpretation is not clear. Deason et al. (2013) showed that several of the 11 stellar halo models of Bullock & Johnston (2005) — all of which were built by simply superposing accretion events — showed breaks in their radial density profiles.

3.2. Velocity and Metallicity Distributions

Low- or medium-resolution spectroscopy can add the additional dimensions of line-of-sight velocity and overall metallicity to maps of stellar populations. The right hand
panels of Figure 1 show some expectations for a purely accreted halo. For example, around the Milky Way, Schlaufman et al. (2009) used SEGUE spectroscopy of a sample of metal-poor main-sequence turnoff stars to find substructures in velocity and use the frequency of their occurrence as an indicator of the importance of accretion in forming the stellar halo. In subsequent work they found these substructures to be chemically distinct from the kinematically-smooth halo component (Schlaufman et al. (2011)). They also showed that an increase in spatial autocorrelation of [Fe/H] for stars in their smooth component with increasing distance from the Galactic center (Schlaufman et al. (2012)). These observations were consistent with models of purely accreted stellar halos beyond 15 kpc, but not within.

In an analogous spectroscopic survey of giant stars in selected fields around M31 (the SPLASH survey), Gilbert et al. (2009) found that stars in kinematically cold structures tended to be more metal rich than those assigned to the kinematically hot components. They also showed that higher surface brightness features are generally more metal rich than lower surface brightness features — a trend that is seen in simulations and can be attributed to more massive (and hence more metal rich) accretion events leading to the most dominant substructures (see illustration in Figure 2).

Dorman et al. (2013) extracted HST photometry of M31’s disk from the PHAT survey and spectroscopy of 5000 giants from the SPLASH survey and used the combined information on the luminosity function and velocity distribution of stars to show an excess of stars with disk-like luminosity functions moving at high speeds relative to the disk. They concluded this was evidence for a kicked-out population in M31’s stellar halo.

3.3. Full Orbit and Abundance Distributions

The Milky Way is the one place where the dimensions to be studied can be further augmented with detailed abundances from high-resolution spectroscopy and/or astrometric measurements of proper motions and distances to allow the calculation of orbital elements. Indeed, the original Eggen et al. (1962) work that started this field looked at the orbits of stars of different metallicities.

A recent series of papers (Nissen & Schuster (2010), Nissen & Schuster (2011), Schuster et al. (2012), Nissen & Schuster (2012)) has used a sample of 94, metal rich ($-1.6 < [\text{Fe/H}] < -0.4$) dwarf stars limited to within $\sim335$ pc of the Sun to investigate halo formation.
scenarios. The authors found that the quality of their data allowed them to see distinct high- and low-\(\alpha\) sequences (separated by less than 0.2 dex in \([\alpha/\text{Fe}]\)). (Note that this difference in abundances is far smaller than the almost 1 dex spread in \(\alpha\) elements that is seen in dwarf spheroidal satellites of the Milky Way and clearly distinguishes the satellite populations from the field halo stars — see summary in Venn et al. (2004).) They find the high-\(\alpha\) stars have abundance patterns that suggest an origin in regions with a high star formation rate (with contributions from only SN II and massive stars), are on orbits with apocenters out to \(\sim 16\) kpc and a range of eccentricities, and have ages 2-3 Gyrs older than the low-\(\alpha\) stars. The low-\(\alpha\) stars have abundance patterns betraying additional pollution by low-mass AGB stars and are on more eccentric orbits with larger apocenters. The authors conclude that the high-\(\alpha\) arise from either in situ or kicked-out primordial populations while the low-\(\alpha\) stars are likely accreted.

The above work shows the power of such a high-dimensional data set. Subsequent work with APOGEE measurements of stellar abundances has allowed these dual populations to be traced over a much larger volume to study the disk/halo interface across a significant fraction of the Galaxy (Hawkins et al. (2015), Hayden et al. (2015)).

4. Aside: a Case-Study of the Triangulum-Andromeda Stellar Clouds

This section discusses just one group of stellar substructures in our own Milky Way halo as an illustration of how additional dimensions of information can introduce revised interpretations on the contributions of the different possible formation mechanisms. The Triangulum-Andromeda clouds are examples of stellar over-densities in the halo of the Milky Way thought to have formed through the disruption of a satellite and hence contributing to the accreted component of the stellar halo. A single over-density was originally identified in M giant stars selected from the Two Micron All Sky Survey (2MASS) covering the region \(100^\circ < l < 160^\circ\) and \(-35^\circ < b < -15^\circ\) at approximately 20 kpc from the Sun (Rocha-Pinto et al. (2004)). Spectroscopic follow-up of these M giants showed them to have a velocity distribution with a small dispersion (\(\approx 25\) km s\(^{-1}\)) and a shallow gradient with mean velocities increasing with decreasing \(l\) (in the Galactic-Standard-of-Rest – GSR – frame). A double sequence of main-sequence stars in the foreground of the PAndAS survey was later found in the same region (called TriAnd1 and TriAnd2 by Martin et al. (2007)). Sheffield et al. (2014) found M giant counterparts to both of these sequences in the 2MASS data, at distance of \(\approx 18\) kpc and \(\approx 28\) kpc respectively (i.e. Galactocentric \((R, Z) \approx (24, -8)\) kpc and \((33, -12)\) kpc for the central pointing at \((l, b) = (125^\circ, -25^\circ)\). M giants in both TriAnd1 and TriAnd2 followed the same velocity distribution found in the original study. Sheffield et al. (2014) assumed that the large Galactocentric radius and distance below the mid-plane combined with the small dispersion was suggestive of neither a disk nor random halo population and used simulations of satellite disruption to show that both TriAnd1 and TriAnd2 could plausibly be due to single accretion event of a satellite on a retrograde orbit.

In subsequent work, Price-Whelan et al. (2015) took spectra of RR Lyrae stars in the same region as the TriAnd clouds with the aim of using those stars that had line-of-sight velocities co-incident with the clouds to provide better indicators of their distances. This experiment failed because no RR Lyrae could be found with co-incident velocities. More quantitatively, the number ratio of RR Lyrae to M giant stars associated with the TriAnd clouds was estimated to be \(f_{RR, MG} < 0.38\) (at 95% confidence). This ratio is quite unlike any of the known satellites of the Milky Way (\(f_{RR, MG} \approx 0.5\) for the very largest and \(f_{RR, MG} >> 1\) for the smaller satellites) and more like the population of stars born in the much deeper potential well inhabited by the Galactic disk (\(f_{RR, MG} < 0.01\)).
These results provide dramatic support for the recent proposal by Xu et al. (2015) that, rather than stars accreted from other galaxies, the TriAnd clouds could represent stars kicked-out from our own disk. N-body simulations of a Milky-Way-like galaxy perturbed by the impact of a dwarf galaxy demonstrate that, in the right circumstances, concentric rings propagating outwards from that Galactic disk might possibly produce similar overdensities (Purcell et al. (2011)). If this interpretation is correct, the TriAnd clouds would be the first populations of disk stars to be found in our Galactic halo and a clear signature of the importance of this formation mechanism for stellar halos more generally.

5. Summary of Current Status and Future Prospects

This review has outlined three plausible formation scenarios of stellar halo populations (in situ, kicked out or accreted). While hierarchical structure formation suggests that accretion must play a role in the formation of all galaxies, theoretical studies do not yet provide clear predictions for the importance of the other populations: there is still debate about whether a pure in situ population, formed in the halo itself and still orbiting there, should exist at all. Observationally, abundant substructure around the Milky Way and Andromeda galaxies confirm the dominance of accretion in forming their outskirts. Within 10’s of kpc of the centers of these galaxies the situation is less clear, as orbital times in these regions are much smaller and spatial substructure from accretion has time to phase-mix away. First studies adding velocity and abundance dimensions in these regions have pointed to possible kicked-out populations (Dorman et al. (2013), Price-Whelan et al. (2015)) as well as significant accreted populations (Nissen & Schuster (2011)). The clear separation of the local stellar halo into “high-α” and “low-α” sequences at low-metallicity in abundance-space suggests there might also be a true in situ component (Nissen & Schuster (2012), Hawkins et al. (2015)).

Future data sets promise breakthroughs in this field. Currently, the GALAH survey is collecting high-resolution spectra for millions of Milky Way stars with the aim of “chemically tagging” both components (as suggested by Freeman & Bland-Hawthorn (2002)). In the near future, data from ESA’s Gaia satellite (Perryman (2002)) should allow the calculation of orbital elements for the nearby stellar halo, so substructures from accretion not apparent in space-alone can be detected as associations in orbit-space. Coincidentally, the samples of external galaxies surveyed on large scales to very low surface brightness will become much larger. For example, the wide field-of-view and spatial resolution of the proposed WFIRST satellite will allow it to look at 50-100 galaxies within 10Mpc and create maps of coverage and depth analogous to those available for Andromeda today. On a slightly longer timescale, LSST promises to create catalogues of millions of galaxies surveyed to surface brightness of ~29 mag/arcsec² (Ivezic et al. (2008)) — with 1-10% of these expected to host a detectable remnant of an accretion event (Johnston et al. (2008)). We will then be able to see if our results for the Milky Way and Andromeda (an outer stellar halo dominated by accretion) are more generally true.

References
