Sky & Telescope 2003

Tracing the Milky Way History from its Fossil Records

Cristina Chiappini

Like other spiral galaxies, our Galaxy, the Milky Way (MW), has several recognizable structural components that probably appeared at different stages in its galaxy formation process (Fig. 1). We believe that these different components hold clues about how the Milky Way formed and evolved. There are many so-called fossil records of this process, and these include not only kinematical properties, but also chemical ones. Many of the kinematical properties can be lost due to dissipation occurring in different phases of the Galaxy formation process (see "The New Galaxy" by K. Freeman & J. Blount-Hawthorn 2002 and "The evolution of our Galaxy" by James Binney, S&T March 1995), which makes it quite complex to try to use them alone for finding an answer to the puzzle of galaxy formation. In contrast, the chemical properties are probably able to provide a more clear picture of what happened in the past.

However, in order to understand how our Galaxy evolved into its present state, we would like to be able to combine both chemical and kinematical properties. Ideally, astronomers would like to compare their models for the formation of the Milky Way with a large sample of representative stars of a specific age, chemical composition, and a precise site of formation. This is not always an easy task given the fact that the best data available are confined to a small region around us, the so-called "solar vicinity." The GAIA mission, to be launched in 2009, will provide unprecedented positional and radial velocity measurements with the accuracy needed to produce a stereoscopic and kinematical census of about one billion stars in our Galaxy and throughout the Local Group (circa 1/6 of the Galactic stellar population). This data together with chemical information, now accessible with instruments like UVES on VLT, will represent a new era for the study of our Galaxy (see Petyman et al. 2001). But we are not sitting and waiting for new data! Over the last years, enormous progress has been made, and new discoveries have been reported almost every day! We would like to discuss some of these here and see how much astronomers already know about the formation of our Galaxy.

The study of chemical signatures is called "chemical evolution" (Tinsley 1980, Matteucci 2001). Its goal is to understand the processes involved in the enrichment of galaxies and to use this knowledge to reconstruct the chemical history of the interstellar gas in a galaxy since the Big Bang. Chemical evolution is especially interesting for studying our own Galaxy, since in this case it is possible to very precisely measure the abundances of stars and gas (i.e. the amount of a given element observed either in the photonized gas around hot stars or in the photosphere of stellar envelopes). As we will see in the next sections, the abundance of stars and gas in our Galaxy represents an important fossil record of its past evolution. By modeling this evolution, we can obtain a better understanding of how the Milky Way formed. Moreover, we can apply what we learned from the history of our own Galaxy to other galaxies and even try to get a big picture of how the entire universe evolved and was filled with chemical species.
Stars: Prodigious chemical factories

One of the most notable triumphs of theoretical astrophysics has been the understanding of the origin of chemical elements in the Universe. We now know that only a few elements such as hydrogen, helium, deuterium and lithium, were synthesized during the Big Bang, but almost everything else was manufactured inside stars. A direct consequence of this fact is that we, ourselves, are made of elements created inside stars. Thus, we could say that in some sense the appearance of life was possible thanks to the cosmic laboratory of our Universe (Tinsley 1997). But how could these elements end up in our planet or even being part of our bodies? To understand this, a further important fact is that stars are born and die. The death of a star is of fundamental importance, because most of the elements that were produced during its entire evolution are then ejected into the interstellar medium (the gas that exists in between stars, hereafter ISM – see “The Stuff between the Stars” by Gillian Knapp, S&T May 1995). These elements are then mixed with the gas already present, increasing its metallicity1. From this gas new stars (and planets) will form, and their chemical composition will be different from that of previous generations. These new stars will evolve and upon their death will again enrich the ISM with more chemical elements. In principle this process will continue until the exhaustion of the gas itself, which occurs, since some matter will remain locked up in low-mass stars and stellar remnants such as black holes or neutron stars (see also “Recycling in the Universe” by Alyssa A. Goddman, S&T November, 2000).

Stars of different masses will enrich the primordial ISM with different chemical species. The different ways a star can finish its life are shown in Fig.2 together with their main contribution for the “chemical enrichment” of the universe. The more massive a star is, the higher will be the temperature attained in its inner core. This is why stars that have more than 8 times the mass of our Sun will be able to produce elements like O, while the less massive ones will never be able to produce this element. Massive stars have brief lives (on the order of a few million years) compared to lighter ones and are also less numerous. The lightest stars, some having merely one-tenth of the mass of our Sun, live longest and some of them have lifetimes superior to the age of the Galaxy. These ones in particular constitute important fossil records of the chemical properties of the gas from which they were born. The distinction is crucial, because it is primarily at the end of its life that a star makes its contribution of newly synthesized elements to the Galaxy.

Once the astronomers knew that most of the elements we see today where produced inside stars and that stars of very low mass have a lifetime which is longer than the age of the Galaxy, a natural question was salient: can we find the first generation of stars? Can we find stars that contain essentially no metals? These stars would represent an extremely important fossil record of the very early phases of the evolution of our Galaxy as they would have been born at a time where the gas of the Milky Way was essentially primordial and not yet enriched by subsequent stellar generations. In this sense, old Galactic stars (as it also happens for distant galaxies) take us to within the first steps of

1 The quantity of the elements heavier than He is often called “metallic”. It is expressed by Z, the ratio “metallicity” is sometimes used to indicate the abundance of this.
cosmic time. One thing is sure, life the way we know it today could not have existed in these early phases just because the several chemical elements were still not present in the universe in the quantities needed for life to exist!

Astronomers started melting the first stars a long time ago. They searched for stars that belong to the Galactic halo, as this is supposedly the oldest population in our Galaxy (see Fig. 1). In the last two decades, some metal-poor stars (as astronomers call stars with very few quantities of elements heavier than He) were found, with the most metal-poor one having an iron abundance less than 1/10 000 that of the Sun. Since no other star more metal poor than that had been found, astronomers started to think that either a population of the first stars never existed or that such a population of stars would have been made only by massive objects whose lifetimes were short and thus not observable anymore at the present time. A very recent result challenges all previous interpretations. In fact, in a recent article in Nature, Christlieb et al. (2002) report the discovery of a star with an iron abundance as low as 1/200 000 of the solar value. This is a halo star of 0.8 solar masses (so, almost the mass of our Sun), located 35,000 light years from us! This discovery suggests that the first generation of stars also contained long-lived low-mass objects and that previous failures in finding them are probably due to observational limitations. As more powerful telescopes are built, we should be able to find more of these objects.

How can we use the "fossil" information left by the chemistry to understand how did our Galaxy form? Or: The Formation of the Milky Way: collapse vs accretion or collapse + accretion? The open debate.

A direct consequence of the different timescales involved in the evolution of stars of different masses in binary systems that give rise to SNIa explosions (see Fig. 2), is that the ISM will be enriched faster in elements produced by short lived stars (massive stars) and slowly on those elements produced essentially by low and intermediate mass stars and SNIa systems. Hence, the behavior of the ratio of two elements which are restored into the ISM on different timescales, as a function of metallicity (Fe/H) (which essentially is a time-axis), can be used as a "clock" that measures how fast metal enrichment proceeded in a given region (Fig. 3). By analyzing the abundance ratios in stars belonging to the halo, bulge and disk, astronomers can see clear indications that the star formation in the halo and bulge had a peak in the past, being very inefficient afterwards. In other words, halo and bulge contain mainly "old stars". In contrast, the one in the disk was almost constant and continuous. This seems to be true for all spiral galaxies (although there are still controversies about the bulge formation) where the reddish older stars are located in the halo and bulge and the bluer younger ones are located in disks, especially inside their spinal arms.

The halo.

If we had been able to observe our Galaxy when the universe was young, we would have seen essentially the halo and of the bulge. What do we know about halo formation? The first authors to show that it is possible to combine stellar abundances and stellar
In order to study the Galaxy formation were Eggen, Lynden-Bell and Sandage (1962 - ELS). From a study of a kinematically selected sample of high velocity stars, they found a remarkable correlation between chemical properties (indicated essentially by [Fe/H] defined as log(Fe/H)-log(Fe/H☉)) and orbital eccentricity e. In the sense that stars with the lowest metallicity are invariably moving in highly elliptical orbits. Moreover, they show that the metal-poor stars reside in a halo suggesting that it was created during a rapid (around 10^10 years) collapse. Searle & Zinn (1978) proposed that, instead of a single-cloud collapse, the halo of the Milky Way formed by the aggregation of many cloud fragments, each of which may have already formed stars and globular clusters. Over the last 40 years, astronomers have been collecting more information and are now able to refine the ideas of ELS and Searle & Zinn, which, nevertheless, are certainly among the most influential papers in the subject of galaxy formation.

There are in fact many signatures that our Galaxy did not evolve as a “closed box” model (this model assumes that the Galaxy is a closed-box system and that its total mass does not change with time). The most important example is the discovery by Ibata et al. (1994) of the Sagittarius dwarf spheroidal galaxy, which is currently merging with the MW. Moreover part of the halo of the Milky Way seems to have been built by the accretion of smaller systems (for example some of the halo globular clusters have chemical properties which indicate that they are probably younger than most of the globular clusters found in the halo). But how much of the MW halo was formed by a fast collapse and how much was formed by accretion of smaller systems is still an open question. Certainly, the low scatter in some of the abundance ratios at low metallicities (which define the “plateau” seen in Fig. 3) suggests that most of the field halo stars (i.e., stars that not belong to globular clusters) should have been formed when the halo was already assembled, as originally proposed by ELS. In fact, if the halo were the result of random evolution of independent clouds, we should expect a larger spread in the [OIII] metallicity in halo stars, especially because of the different metallicities at which the change in the [OIII] metallicity would occur in the different clouds (Fig. 3). Moreover, very recently, some authors reported abundance ratio measurements in stars belonging to nearby dwarf galaxies (now possible with UVES in the VLT). They concluded that their chemical signatures do not match the ones of the halo stars suggesting that, if indeed the halo was formed by the accretion of dwarf galaxies, those should have been different from the ones we observe today. There is probably some truth in both pictures: the inner halo and the bulge have properties consistent with a dissipative collapse, while the outer halo is likely to have had a more chaotic origin (see also Pagel 1997).

The thin disk:

The general belief today is that star formation did start in the galactic thin disk before it had reached its final mass. In fact, infall of gas may have been very important throughout past galactic history. Such a matter is assumed to be unprocessed and hence to have a pristine metal content (given by the Big Bang nucleosynthesis). The infall has often been used as a way to explain the shortage of stars with low heavy element content in the galactic disk (a problem referred to in the literature as the G-dwarf Problem – see Pagel 1997 for a clear discussion). The latter is evident when comparing observations with
predictions from the so-called simple model (see Fig. 4). This is easy to understand because if star formation in the disk started at a time when only a small fraction of its final mass had collapsed to the plane, there would be only a relatively small number of stars with very low heavy element content, since the heavy elements produced in this first generation of stars would be available to be incorporated in further stars formed as the galactic disk increased in mass.

Slow infalling gas onto the galactic disk can also explain why our Galaxy is still in its active phase, i.e. it is still forming new stars. But where did this infalling gas come from? Is this gas left from the formation of the halo that is then used to form the disk as often suggested by many authors? In particular, models such as ELS suggest that the formation of the disk involved a smooth dissipational collapse of the halo and assume a continuous evolutionary transition in the formation of the thick disk and the thin disk. We now believe that this is not what happened. The first argument against the idea of a continuous collapse formation of our Galaxy is the fact that the angular momentum distribution of halo stars suggests that the halo is quite disconnected from the evolution of at least the local part of the disk (Page 1997) and that the residual halo gas would have been used to form the bulge and not the disk. But a second argument is in fact given by the metallicity distribution of the G-dwarfs. Chappell et al. (1997) show that in order to be able to reproduce this distribution we should assume that the galactic disk at the Sun's position formed quite slowly (much more slow than the formation of the halo which probably took around 1 billion years as suggested by the abundance ratios of halo stars discussed before). The so-called "Two-Infall Model" proposed by Chappell et al. (1997) was based on the following idea: a first infall episode gave rise to the halo and maybe part of the thick-disk and the bulge (see below), then a second infall episode, completely independent from the first one, formed the thin-disk in a longer timescale. In this way it was possible to reproduce the completely different metallicity distributions of halo and thin disk. A third strong chemical argument suggesting a discontinuity between the halo and thin-disk will be discussed below as it is probably related with the origin of the thick-disk.

But before that, another important property of our thin-disk musts attention. Apart from the differences in the amount of chemical elements in the halo, disk and bulge, differences are also seen along the disk itself. In fact, essentially all spiral galaxies show abundance gradients along their disks such that the amount of "rare" elements is larger in the inner regions compared to the outer ones. As the chemical enrichment is linked to the stellar activity, this is telling us that more stellar generations occurred in the inner disk than in the outer parts. This is what leads to the so-called "inside-out" picture for the formation of the Milky Way. In this view, the inner thin disk formed faster than the outer parts that could be still forming now.

One of the best ways to test the prediction that the disk formed from "inside-out" is to measure the abundance gradient of deuterium. The fact that deuterium is only destroyed and never created by stars makes this element a good indicator of past star formation. Regions with a high rate of stellar activity should be depleted in deuterium, unless they had been re-supplied with primordial gas. In fact, chemical evolution models can
contain the primordial value of deuterium (very important to test the Big Bang Nucleosynthesis theory! — see a clear discussion in Pagel 1997) by trying to reproduce the observed abundances of deuterium in the Solar System and local interstellar medium. Current models suggest a primordial (D/H) < 4 x 10^-5, which implies a baryonic density value in close agreement with the recent results obtained from the Cosmic Microwave background (Chapline, Renda & Matteucci 2002 A&A). On the other hand, the abundance of deuterium in the inner parts of the galactic disk represents a test for the models, since in the absence of inflow of primordial gas deuterium would be almost completely depleted. Moreover, if the idea that the thin disk of the Milky Way is still in process of formation in the outer parts is true we would expect the deuterium value there to be close to the primordial one (see Fig. 5). These model predictions still await measurements of the abundance gradient of the deuterium, which presently are not available. But new results are expected either with FUSE or from radio studies of far out molecular clouds (see Lubowitch et al. 2000, Hoopes et al. 2003).

The thick disk: its controversial origin

The thick disk was first identified by Yoshii (1982) and soon confirmed by Gilmore and Reid (1983) who named it the "thick disk" of the Milky Way. The stars belonging to the thick disk look unlike the halo stars, but not as fast as the thin disk stars. Moreover, the metallicity distribution of thick disk stars is intermediate between that of the thin disk and that of the halo. The mean metallicity found for thick disk stars is [Fe/H] = -0.6dex, while that of the halo is around -1.5dex and that of the thin disk is around -0.1dex.

The next piece of evidence suggesting that something happened between the formation of the halo and that of the thin disk was found by Gratton and collaborators (Gratton et al. 1997, 2000). These authors presented a compilation of stars for which both the kinematical properties and the [Fe/O] abundance ratio were known from which they could identify three kinematically distinct populations (a nice summary of this discussion was presented by Matteucci 2001): a) a population made of halo and thick disk stars originating from an early fast dissipative collapse (all of them located in the [Fe/O] plane), reflecting the fact that they were formed from a gas which was essentially enriched by massive stars), b) a population of thin-disk stars originating from a subsequent even more dissipative collapse which formed the thin disk (on a long timescale as discussed before) and c) an intriguing population of thick-disk stars the origin of which should be different from the others. In other words, they seem to have been formed in satellite galaxies and then accreted by the MW given their unusual chemical and kinematical properties. Moreover, they also observed a discontinuity in the [Fe/O] vs. [O/H] plot which was interpreted as a break in the formation of stars between the halo and the thin-disk phases (see Fig. 6). In particular, what they observed was that around [O/H] = -0.2dex the oxygen abundance remained constant whereas the [Fe/O] ratio kept increasing. This is interpreted as a break in the star formation since in this case oxygen would not be produced whereas Fe would continue to originate from the long-living systems giving rise to type Ia SNe that were born before this break.
On the basis of these data, we included another important new element in the "Two Infall model" which was the "threshold" in the star formation rate. Essentially we stop forming stars every time that the density of gas available is below a certain critical value (this mechanism is in fact observed in many galaxies, see Kennicutt 1996). This model aimed at reproducing mostly the halo (and the dissipational part of the thick disk) and the thin disk components and not explicitly the thick disk which seems to have a mixed origin: part of it made of stars born during a dissipational collapse and part of it made of stars accreted from smaller satellites. The existence of the threshold in the "Two-Infall model" leads naturally to a halt in the star formation rate at the end of the first infall episode, thus creating the halo in the star formation suggested not only by the observations discussed above but also by dynamical models (Larson 1976).

Even though the previous ideas seem to explain many of the observed properties in our Galaxy one important question remains open: where did the gas which forms the thin disk came from? The suggestions range from leftovers of the first flash collapse that formed the halo and that, due to a larger angular momentum, took a long time to settle into a disk or gas brought to our Galaxy during a major impact with another galaxy. Now we enter into a completely open field that is at the present the reason of many hot debates among astronomers.

A Major Impact with our Galaxy around 10 Gyr ago?

Unfortunately, I don't have the space left to discuss here all the pieces of evidence that point to a major merger with our Galaxy in the past. I would like just to mention that there are some discontinuities seen also in the kinematical properties (see for example Wyse and Gilmore 1995) of stars that have a typical thick-disk metallicity. But to be consistent with the smooth properties (both chemical and kinematical) among the thin-disk stars this major impact should have occurred 9 or 10 Gyr ago.

One might think that this is the reason for the gap observed in the star formation rate suggested by the abundance discontinuity discussed above. Moreover, most of the major mergers with the Milky Way must have happened before the galaxy formed its thin disk of stars and gas. In fact, the disk is too fragile and so its existence implies that mergers cannot have contributed more than a few percent of its mass in the last 5 billion years.

The gas around the MW unveiled by FUSE: are we seen the "infall"?

Do we see the infalling gas required by the chemical evolution models to explain the observed properties of our Galaxy? In fact, if the outer disk is indeed still forming as predicted by the chemical evolution models, we might even be able to observe the infall of gas clouds right now. Astronomers using the FUSE Ultraviolet Spectroscopic Explorer (FUSE — http://fuse.gsfc.nasa.gov) have shown that the Milky Way is surrounded by a huge, extended "corona" of hot gas but the origin of it is still controversial (see Stenback et al 2005). As discussed by these authors this gas traces a variety of phenomena, including tidal interactions with the Magellanic Clouds, accretion of gas, outflow from the Galactic disk, warm/hot gas interactions in a highly extended Galactic corona, and
Figure Captions:

Figure 1 (this figure could be a combination of Figs 2.3 and 3.4 of my MScTh article) - Our Galaxy is a spiral galaxy of morphological type Sbc in the Hubble classification and it can be described as a central flattened disk of gas and stars with a stellar bulge in the center. The disk (both thin and thick disks) and bulge are surrounded by a spherical halo made of mass-poor old stars and globular clusters. The entire system is embedded in a massive halo of dark matter whose nature is still unknown. The stellar populations of these main components have quite different chemical compositions, kinematics and dynamical properties reflecting different evolutionary histories. The stellar halo extends up to around 300 kpc from the center, the bulge has an extension of around 2 kpc and the disk is thin (around 200 pc thick) and extends up to around 20 kpc from the center. The Sun is located at a distance of 8.5 kpc from the center.

Figure 2 - The final stages of stars of different masses and their contribution to the chemical enrichment of the Galaxy upon their death. Low and intermediate mass stars die as planetary nebulae (HST images) thus ejecting their envelope into the ISM and enriching it mainly on He, C and N. Massive stars end their lives in a more violent way, exploding as supernovae of type II (SN 1987 - HST/Chandra image and/or Crab Nebula – a supernova remnant HST). In the case of massive stars the carbon core contracts reaching high enough temperatures leading to C, Ne, O and Si, burning. Silicon burning leads finally to iron. Iron is the most stable form of nuclear matter and there is no energy to be gained by burning it into heavier elements. Without any further source of energy to balance the gravity, the iron core collapses, reaching nuclear densities. The falling upper layers of matter will then bounce on the core leading to what astronomers call a SN type II. These stars enrich the ISM in several elements, but mainly in oxygen and other alpha-elements (like Mg, Si, Ca) and produce a small quantity of iron. A third way of seeding the ISM with elements heavier than helium is the explosion of an intermediate mass star in a binary system: the so-called Type Ia Supernova event (show a schematic figure?).

The importance of these objects is that in this case the products of the stellar evolution that are ejected into the ISM are mainly iron and small quantities of S, Si and Ca. This explosion happens when in a binary system, a C-O white dwarf accretes mass from its companion star (a red Giant) eventually overzealous the Chandrasekhar limit thus generating an explosion. The explosion "clock" in this case is given by the mass of the

---

1 kpc corresponds to 3.76 light years. For the Galaxy we often quote distances in kpc (for 3000 light years). In kpc we our Sun is located at a distance of around 27000 light years from the Galactic center.
Intragalactic gas in the Local Group. Distinguishing between the various phenomena will require continuing studies of the distances, kinematics, elemental abundances, and physical states of the different types of high velocity features found in this study. This is a very recent new area that certainly reserves still many surprises for the coming years!

Another idea is that the so-called High Velocity Clouds, patches of hydrogen gas falling into the galactic disk, could also be an observational proof of the slow build-up of our Gal...
star that leaves its envelope and which is an interstellar-mass star. This means that the enrichment in gas, due to type Ia SNe will happen with a delay relative to the enrichment of oxygen or other elements mainly produced during SN11 explosions.

Figure 3 - Abundance ratios in “cosmic clocks” (from a sketch by Jeff). Panel a: Shows a schematic indication of how the [O/Fe] ratio depends on the star formation history. If star formation has taken place rapidly relatively to the solar neighborhood, as may apply to the Galactic halo and for most of the elliptical galaxies, the [O/Fe] plateau extends to larger [Fe/H] values. This happens because in a system that forms all its stars very fast, essentially the SNIa did not have time to enrich the gas before the process of star formation ended. This is why such systems do not show the “knee” in the [O/Fe] plot as the one we observe in the solar vicinity. Conversely, if it has been relatively slow or has happened in bursts separated by long intervals, as may apply to the Magellanic Clouds (SMC, LMC), the [O/Fe] ratio “knee” happens at lower metallicities compared to the solar vicinity, before the type Ia SNe managed to raise too much the [O/Fe] content. Panel b: If the disk of our Galaxy really formed “inside-out” as current believed, the models predict that also the [O/Fe] should be a function of the galactocentric distance as shown in the figure. The inner regions, which evolved faster would present an extended plateau compared to the one obtained in the outer parts of the galactic disk. This is a prediction that will have to wait to be confirmed or disproved until astronomers will be able to measure precise stellar abundances outside the solar vicinity. We note that in the Milky Way, the stars that belong to the halo and most of the thick disk stars are located at the [O/Fe] plateau suggesting that both, the halo and thick disk are old, and formed during the first billion year of the Galaxy evolution. The solar vicinity disk stars show a low [O/Fe] ratio as this component is much younger than the other two.

Figure 4 - The G-dwarf metallicity distribution — models predicting a fast formation of the solar vicinity do not explain the metallicity distribution of G and K dwarf stars. The best fit to these particular important observational constraint is obtained by assuming an e-folding time for the gas to settle into a disk at the solar vicinity, was of the order of 7 billion years. Such a long timescale for the formation of the solar vicinity was first suggested in Chiappini et al. (1997). According this chemical evolution model, the MW formed by “two-infall” episodes: the first one formed the halo-thick disk components quite fast (in less than a billion years) and a second one that formed the thin disk. The latter was a slow continuous infall whose e-folding time is probably a function of the galactocentric distance thus forming the disk “inside-out”. Infall has been known as a solution for the so-called “G-dwarf problem”. More recently, the same was shown using K-dwarfs for which the hypothesis that all the stars have lifetimes shorter than the age of the Galaxy is still more correct and again, a timescale of around 7 billion years for the formation of the solar vicinity was found (Kornavé et al. 2002).

Figure 5 - The Deuterium abundance gradient — The abundance gradient of Deuterium predicted by a chemical evolution model is able to fit many of the available constraints in the Milky Way. The models predict that the deuterium abundance

1 [O/Fe] in brackets refers to the normalized value to the solar abundance on logarithmic scale so that [O/Fe] = log(O/Fe) - log(Sun/Fe)
Towards the outer parts of the galactic disk should approach the primordial value as far from the Galactic Center no much stellar activity occurred and primordial infall of gas is probably still operative. In the inner parts we predict that deuterium was depleted.

Figure 6 – Similar to figure 8 of the American Scientist article