How the Milky Way Formed

Its halo and disk suggest that the collapse of a gas cloud, stellar explosions and the capture of galactic fragments may have all played a role

by Sidney van den Bergh and James E. Hesser

Attempts to reconstruct how the Milky Way formed and began to evolve resemble an archaeological investigation of an ancient civilization buried below the bustling center of an ever changing modern city. From excavations of foundations, some pottery shards and a few bones, we must infer how our ancestors were born, how they grew old and died and how they may have helped create the living culture above. Like archaeologists, astronomers, too, look at small, disparate clues to determine how our galaxy and others like it were born about a billion years after the big bang and took on their current shapes. The clues consist of the ages of stars and stellar clusters, their distribution and their chemistry—all deduced by looking at such features as color and luminosity. The shapes and physical properties of other galaxies can also provide insight concerning the formation of our own.

The evidence suggests that our galaxy, the Milky Way, came into being as a consequence of the collapse of a vast gas cloud. Yet that cannot be the whole story. Recent observations have forced workers who support the hypothesis of a simple, rapid collapse to modify their idea in important ways. This new information has led other researchers to postulate that several gas cloud fragments merged to create the protogalactic Milky Way, which then collapsed. Other variations on these themes are vigorously maintained. Investigators of virtually all persuasions recognize that the births of stars and supernovae have helped shape the Milky Way. Indeed, the formation and explosion of stars are at this moment further altering the galaxy's structure and influencing its ultimate fate.

Much of the stellar archaeological information that astronomers rely on to decipher the evolution of our galaxy resides in two regions of the Milky Way: the halo and the disk. The halo is a slowly rotating, spherical region that surrounds all the other parts of the galaxy. The stars and star clusters in it are old. The rapidly rotating, equatorial region constitutes the disk, which consists of young stars and stars of intermediate age, as well as interstellar gas and dust. Embedded in the disk are the sweepingly curved arms that are characteristic of spiral galaxies such as the Milky Way. Among the middle-aged stars is our sun, which is located about 25,000 light-years from the galactic center. (When you view the night sky, the galactic center lies in the direction of Sagittarius.) The sun completes an orbit around the center in approximately 200 million years.

That the sun is part of the Milky Way was discovered less than 70 years ago. At the time, Bertil Lindblad of Sweden and the late Jan H. Oort of the Nether-

SIDNEY VAN DEN BERGH and JAMES E. HESSER both work at Dominion Astrophysical Observatory, National Research Council of Canada, in Victoria, British Columbia. Van den Bergh has a longtime interest in the classification and evolution of galaxies and in problems related to the age and size of the universe. He received his undergraduate degree from Princeton University and a doctorate in astronomy from the University of Göttingen. Hesser's current interests focus on the ages and compositions of globular star clusters, which are among the oldest constituents of the galaxy. He received his B.A. from the University of Kansas and his Ph.D. in atomic and molecular physics from Princeton.

MILKY WAY COMPONENTS include the tenuous halo, the central bulge and a highly flattened disk that contains the spiral arms. The nucleus is obscured by the stars and gas clouds of the central bulge. Stars in the bulge and halo tend to be old; disk stars such as the sun are young or middle-aged.

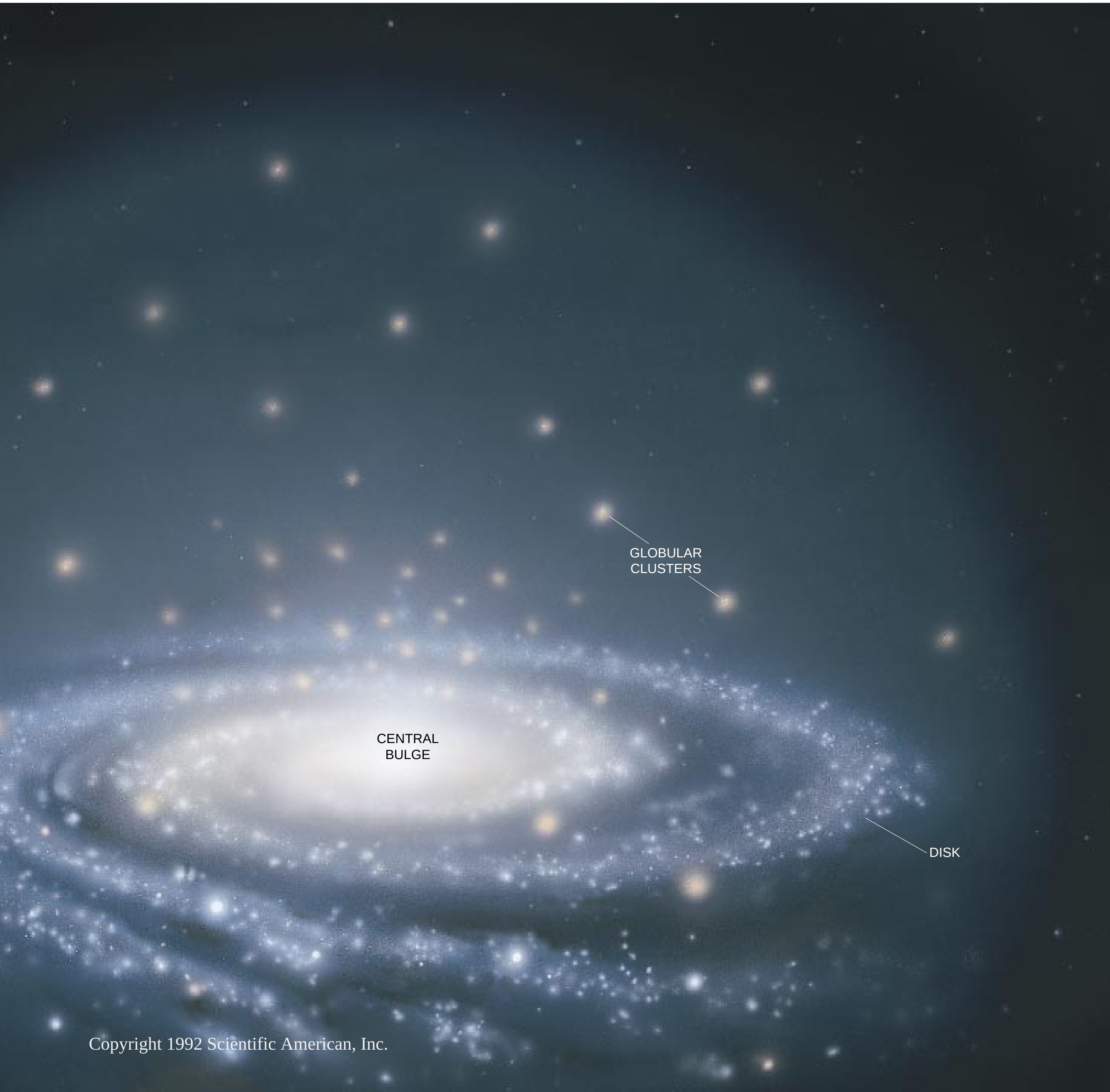
lands hypothesized that the Milky Way system is a flattened, differentially rotating galaxy. A few years later John S. Plaskett and Joseph A. Pearce of Dominion Astrophysical Observatory accumulated three decades' worth of data on stellar motions that confirmed the Lindblad-Oort picture.

In addition to a disk and a halo, the Milky Way contains two other subsystems: a central bulge, which consists primarily of old stars, and, within the bulge, a nucleus. Little is known about the nucleus because the dense gas clouds in the central bulge obscure it. The nuclei of some spiral galaxies, including the Milky Way, may contain a large black hole. A black hole in the nucleus of our galaxy, however, would not be as massive as those that seem to act as the powerful cores of quasars.

All four components of the Milky Way appear to be embedded in a large, dark corona of invisible material. In most spiral galaxies the mass of this invisible corona exceeds by an order of magnitude that of all the galaxy's visible gas and stars. Investigators are intensely debating what the constituents of this dark matter might be.

The clues to how the Milky Way developed lie in its components. Perhaps the only widely accepted idea is that the central bulge formed first, through the collapse of a gas cloud. The central bulge, after all, contains mostly massive, old stars. But determining when and how the disk and halo formed is more problematic.

In 1958 Oort proposed a model according to which the population of stars forming in the halo flattened into a

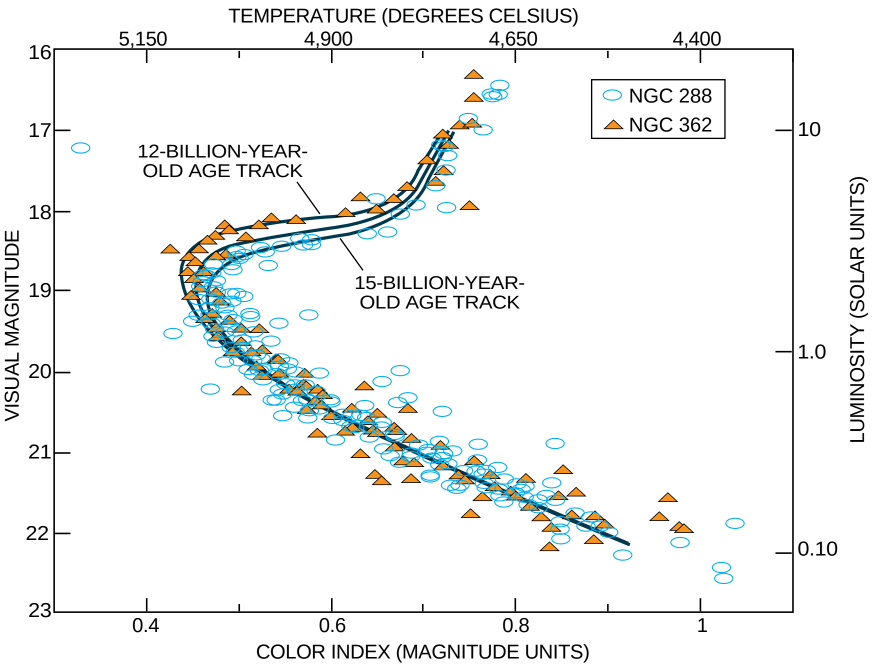


thick disk, which then evolved into a thin one. Meanwhile further condensation of stars from the hydrogen left over in the halo replenished that structure. Other astronomers prefer a picture in which these populations are discrete and do not fade into one another. In particular, V. G. Berman and A. A. Suchkov of the Rostov State University in Russia have indicated how the disk and halo could have developed as separate entities.

These workers suggest a hiatus between star formation in the halo and that in the disk. According to their model, a strong wind propelled by supernova explosions interrupted star formation in the disk for a few billion years. In doing so, the wind would have ejected a significant fraction of the mass of the protogalaxy into intergalactic space. Such a process seems to have prevailed in the Large Magellanic Cloud, one of the Milky Way's small satellite galaxies. There an almost 10-billion-year interlude appears to separate the initial burst of creation of conglomerations of old stars called globular clusters and the more recent epoch of star formation in the disk. Other findings lend additional weight to the notion of distinct galactic components. The nearby spiral M33 contains a halo but no nuclear bulge. This characteristic indicates that a halo is not just an extension of the interior feature, as many thought until recently.

In 1962 a model emerged that served as a paradigm for most investigators. According to its developers—Olin J. Eggen, now at the National Optical Astronomical Observatories, Donald Lynden-Bell of the University of Cambridge and Allan R. Sandage of the Carnegie Institution—the Milky Way formed when a large, rotating gas cloud collapsed rapidly, in about a few hundred million years. As the cloud fell inward on itself, the protogalaxy began to rotate more quickly; the rotation created the spiral arms we see today. At first, the cloud consisted entirely of hydrogen and helium atoms, which were forged during the hot, dense initial stages of the big bang. Over time the protogalaxy started to form massive, short-lived stars. These stars modified the composition of galactic matter, so that the subsequent generations of stars, including our sun, contain significant amounts of elements heavier than helium.

Although the model gained wide acceptance, observations made during the past three decades have uncovered a number of problems with it. In the first place, investigators found that many of the oldest stars and star clusters in the galactic halo move in retrograde orbits—that is, they revolve around the galactic center in a direction opposite to that of most other stars. Such orbits suggest that the protogalaxy was quite clumpy and turbulent or that it captured sizable gaseous fragments whose matter was moving in different directions. Second, more refined dynamic models show that the protogalaxy would not have collapsed as smoothly as predicted by the simple model; instead the densest parts would have fallen inward much faster than more rarefied regions.

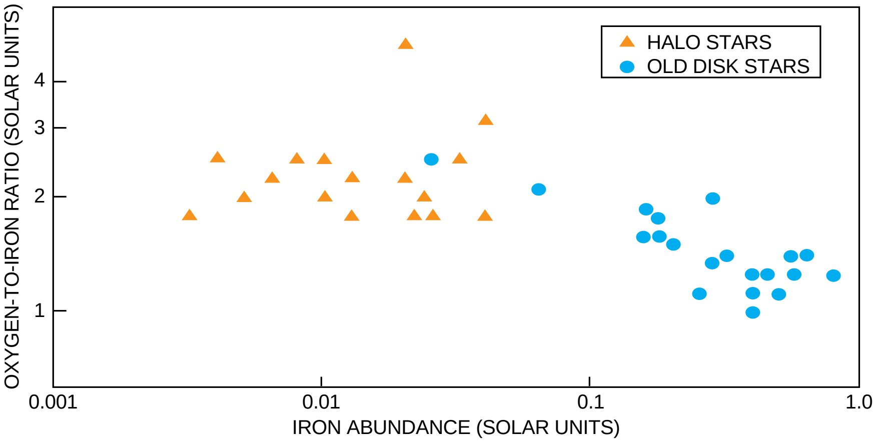


COLOR-LUMINOSITY DIAGRAMS can be used to determine stellar ages. The one above compares the plots of stars in globular clusters NGC 288 and NGC 362 with age tracks (*black lines*) generated by stellar evolution models. The color index, expressed in magnitude units, is a measure of the intensity of blue wavelengths minus visual ones. In general, the brighter the star, the lower the color index; the trend reverses for stars brighter than about visual magnitude 19. The plots suggest the clusters differ in age by about three billion years. The temperature (inversely related to the color index) and luminosity have been set to equal those of NGC 288.

Third, the time scale of galaxy formation may have been longer than that deduced by Eggen and his colleagues. Exploding supernovae, plasma winds pouring from massive, short-lived stars and energy from an active galactic nucleus are all possible factors. The galaxy may also have subsequently rejuvenated itself by absorbing large inflows of pristine intergalactic gas and by capturing small, gas-rich satellite galaxies.

Several investigators have attempted to develop scenarios consistent with the findings. In 1977 Alar Toomre of the Massachusetts Institute of Technology postulated that most galaxies form from the merger of several large pieces rather than from the collapse of a single gas cloud. Once merged in this way, according to Toomre, the gas cloud collapsed and evolved into the Milky Way now seen. Leonard Searle of the Carnegie Institution and Robert J. Zinn of Yale University have suggested a somewhat different picture, in which many small bits and pieces coalesced. In the scenarios proposed by Toomre and by Searle and Zinn, the ancestral fragments may have evolved in chemically unique ways. If stars began to shine and supernovae started to explode in different fragments at different times, then each ancestral fragment would have its own chemical signature. Recent work by one of us (van den Bergh) indicates that such differences do indeed appear among the halo populations.

Discussion of the history of galactic evolution did not advance significantly beyond this point until the 1980s. At that time, workers became able to record more precisely than ever before extremely faint images. This ability is critically important because the physical theories of stellar energy production—and hence the lifetimes and ages of stars—are most secure for so-called main-sequence stars. Such stars burn hydrogen in their cores; in general, the more massive the star, the more quickly it completes its main-sequence life. Unfortunately, this fact



OXYGEN-TO-IRON RATIOS as a function of metallicity (abundance of iron) for halo and old disk stars indicate different formation histories. The high ratios in metal-deficient halo stars suggest that those stars incorporated the oxygen synthesized in supernovae of types Ib, Ic and II. Type Ia supernovae seem to have contributed material only to the disk stars. Beatriz Barbuy and Marcia Erdelyi-Mendes of the University of São Paulo made the measurements.

means that within the halo the only remaining main-sequence stars are the extremely faint ones. The largest, most luminous ones, which have burned past their main-sequence stage, became invisible long ago. Clusters are generally used to determine age. They are crucial because their distances from the earth can be determined much more accurately than can those of individual stars.

The technology responsible for opening the study of extremely faint halo stars is the charge-coupled device (CCD). This highly sensitive detector produces images electronically by converting light intensity into current. CCDs are far superior in most respects to photographic emulsions, although extremely sophisticated software, such as that developed by Peter B. Stetson of Dominion Astrophysical Observatory, is required to take full advantage of them. So used, the charge-coupled device has yielded a tenfold increase in the precision of measurement of color and luminosity of the faint stars in globular clusters.

Among the most important results of the CCD work done so far are more precise age estimates. Relative age data based on these new techniques have revealed that clusters whose chemistries suggest they were the first to be created after the big bang have the same age to within 500 million years of one another. The ages of other clusters, however, exhibit a greater spread.

The ages measured have helped researchers determine how long it took for the galactic halo to form. For instance, Michael J. Bolte, now at Lick Observatory, carefully measured the colors and luminosities of individual stars in the globular clusters NGC 288 and NGC 362 [*see illustration on opposite page*]. Comparison between these data and stellar evolutionary calculations shows that NGC 288 is approximately 15 billion years old and that NGC 362 is only about 12 billion years in age. This difference is greater than the uncertainties in the measurements. The observed age range indicates that the collapse of the outer halo is likely to have taken an order of magnitude longer than the amount of time first envisaged in the simple, rapid collapse model of Eggen, Lynden-Bell and Sandage.

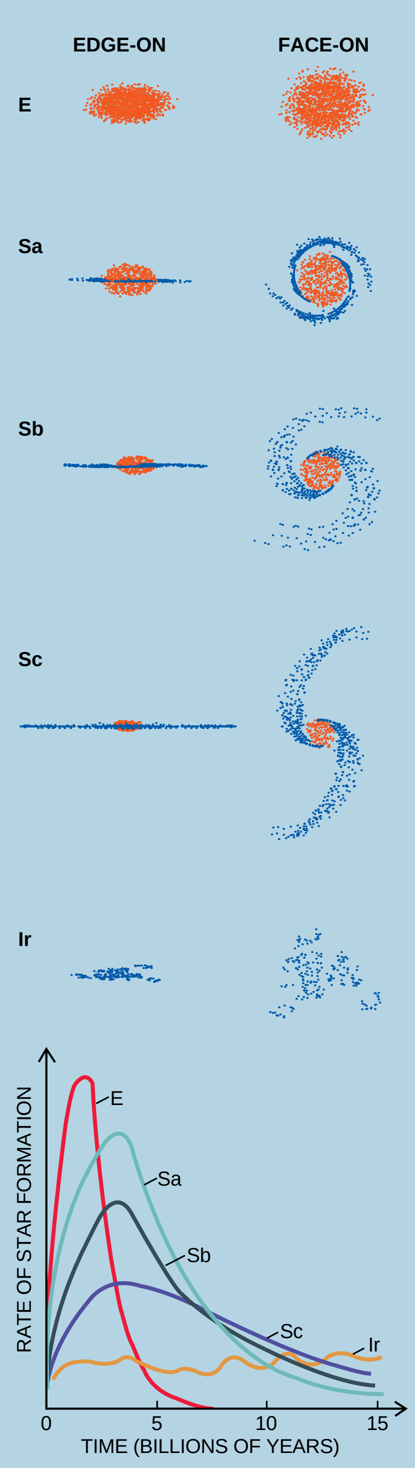
Of course, it is possible that more than one model for the formation of the galaxy is correct. The Eggen–Lynden-Bell–Sandage scenario may apply to the dense bulge and inner halo. The more rarefied outer parts of the galaxy may have developed by the merger of fragments, along the lines theorized by Toomre or by Searle and Zinn. If so, then the clusters in the inner halo would have formed before those in the more tenuous outer regions. The process would account for some of the age differences found for the globular clusters. More precise modeling may have to await the improved image quality that modifications to the *Hubble Space Telescope* cameras will afford.

Knowing the age of the halo is, however, insufficient to ascertain a detailed formation scenario. Investigators need to know the age of the disk as well and then to compare that age with the halo's age. Whereas globular clusters are useful in determining the age of the halo, another type of celestial body—very faint white dwarf stars—can be used to determine the age of the disk. The absence of white dwarfs in the galactic disk near the sun sets a lower limit on the disk's age. White dwarfs, which are no longer producing radiant energy, take a long time to cool, so their

GLOBULAR CLUSTERS, such as Messier 5 above, appear to be among the oldest objects known. They offer invaluable insight into the halo's formation some 15 billion years ago. The 100,000 or so stars exhibit similar abundances of heavy elements, implying that the gas cloud from which each arose was chemically homogeneous.

absence means that the population in the disk is fairly young—less than about 10 billion years. This value is significantly less than the ages of clusters in the halo and is thus consistent with the notion that the bulk of the galactic disk developed after the halo.

It is, however, not yet clear if there is a real gap between the time when formation of the galactic halo ended and when creation of the old thick disk began. To estimate the duration of such a transitional period between halo and disk, investigators have compared the ages of the oldest stars in the disk with those of the youngest ones in the halo. The oldest known star clusters in the galactic disk, NGC 188 and NGC 6791, have ages of nearly eight billion years, according to Pierre Demarque and David B. Guenther of Yale and Elizabeth M. Green of the University of Arizona. Stetson and his colleagues and Roberto Buonanno of the Astronomical Observatory in Rome and his co-workers examined globular clusters in the halo population. They found the youngest globulars—Palomar 12 and Ruprecht 106—to be about 11 billion years old. If the few billion years' difference between the disk objects and the young globulars is real, then young globulars may be the missing links between the disk and halo populations of the galaxy.



MORPHOLOGICAL CLASSIFICATION of galaxies (*top*) ranges from ellipticals (E) to spirals (subdivided into categories Sa, Sb and Sc) and irregulars (Ir). The history of star formation varies according to morphology (*bottom*). In elliptical galaxies, stars developed in an initial burst. Star formation in spirals was less vigorous but continues today. In most irregular galaxies the birthrate of stars has probably remained constant.

At present, unfortunately, the relative ages of only a few globular clusters have been precisely estimated. As long as this is the case, one can argue that the Milky Way could have tidally captured Palomar 12 and Ruprecht 106 from the Magellanic Clouds. This scenario, proposed by Douglas N. C. Lin of the University of California at Santa Cruz and Harvey B. Richer of the University of British Columbia, would obviate the need for a long collapse time. Furthermore, the apparent age gap between disk and halo might be illusory. Undetected systematic errors may lurk in the age-dating processes. Moreover, gravitational interactions with massive interstellar clouds may have disrupted the oldest disk clusters, leaving behind only younger ones.

Determining the relative ages of the halo and disk reveals much about the sequence of the formation of the galaxy. On the other hand, it leaves open the question of how old the entire galaxy actually is. The answer would provide some absolute framework by which the sequence of formation events can be discerned. Most astronomers who study star clusters favor an age of some 15 to 17 billion years for the oldest clusters (and hence the galaxy).

Confidence that those absolute age values are realistic comes from the measured abundance of radioactive isotopes in meteorites. The ratios of thorium 232 to uranium 235, of uranium 235 to uranium 238 or of uranium 238 to plutonium 244 act as chronometers. According to these isotopes, the galaxy is between 10 and 20 billion years old. Although ages determined by such isotope ratios are believed to be less accurate than those achieved by comparing stellar observations and models, the consistency of the numbers is encouraging.

Looking at the shapes of other galaxies alleviates to some extent the uncertainty of interpreting the galaxy's evolution. Specifically, the study of other galaxies presents a perspective that is unavailable to us as residents of the Milky Way—an external view. We can also compare information from other galaxies to see if the processes that created the Milky Way are unique.

The most immediate observation one can make is that galaxies come in several shapes. In 1925 Edwin P. Hubble found that luminous galaxies could be arranged in a linear sequence according to whether they are elliptical, spiral or irregular [*see top illustration on this page*]. From an evolutionary point of view, elliptical galaxies are the most advanced. They have used up all (or almost all) of their gas to generate stars, which probably range in age from 10 to 15 billion years. Unlike spiral galaxies, ellipticals lack disk structures. The main differences between spiral and irregular galaxies is that irregulars have neither spiral arms nor compact nuclei.

The morphological types of galaxies can be understood in terms of the speed with which gas was used to create stars. Determining the rate of gas depletion would corroborate estimates of the Milky Way's age and history. Star formation in elliptical galaxies appears to have started off rapidly and efficiently some 15 billion years ago and then declined sharply. In most irregular galaxies the birth of stars has taken place much more slowly and at a more nearly constant rate. Thus, a significant fraction of their primordial gas still remains.

The rate of star formation in spirals seems to represent a compromise between that in ellipticals and that in irregulars. Star formation in spirals began less rapidly than it did in ellipticals but continues to the present day.

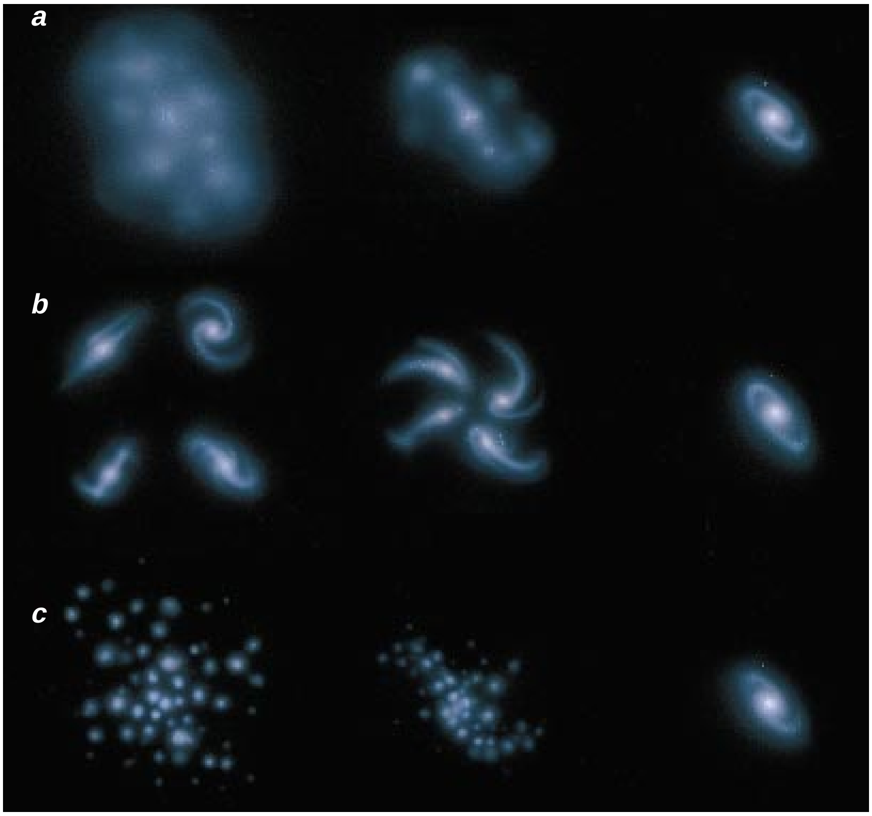
Spirals are further subdivided into categories Sa, Sb and Sc. The subdivisions refer to the relative size of the nuclear bulges and the degree to which the spiral arms coil. Objects of type Sa have the largest nuclear bulges and the most tightly coiled arms. Such spirals also contain some neutral hydrogen gas and a sprinkling of young blue stars. Sb spirals have relatively large populations of young blue stars in their spiral arms. The central bulge, containing old red stars, is less prominent than is the central bulge in spirals of type Sa. Finally, in Sc spirals the light comes mainly from the young blue stars in the spiral arms; the bulge population is inconspicuous or absent. The Milky Way is probably intermediate between types Sb and Sc.

Information from other spirals seems consistent with the data obtained for the Milky Way. Like those in our galaxy, the stars in the central bulges of other spirals arose early. The dense inner regions of gas must have collapsed first. As a result, most of the primordial gas initially present near the centers has turned into stars or has been ejected by supernova-driven winds.

There is an additional kind of evidence on which to build our understanding of how the Milky Way came into existence: the chemical composition of stars. This information helps to pinpoint the relative ages of stellar populations. According to stellar models, the chemistry of a star depends on when it formed. The chemical differences exist because first-generation stars began to “pollute” the protogalaxy with elements heavier than helium. Such so-called heavy elements, or “metals,” as astronomers refer to them, were created in the interiors of stars or during supernova explosions. Examining the makeup of stars can provide stellar evolutionary histories that corroborate or challenge age estimates.

Different types of stars and supernovae produce different relative abundances of these metals. Researchers believe that most “iron-peak” elements (those closest to iron in the periodic table) in the galaxy were made in supernovae of type Ia. The progenitors of such supernovae are thought to be pairs of stars, each of which has a mass a few times that of the sun. Other heavy elements—the bulk of oxygen, neon, magnesium, silicon and calcium, among others—originated in supernovae that evolved from single or binaries of massive, short-lived stars. Such stars have initial masses of 10 to 100 solar masses and violently end their lives as supernovae of type Ib, Ic or II.

Stars that subsequently formed incorporated some of these heavy elements. For instance, approximately 1 to 2 percent of the mass of the sun consists of elements other than hydrogen or helium. Stars in nuclear bulges generally harbor proportionally more heavy elements than do stars in the outer disks and halos. The abundance



MODELS OF GALAXY FORMATION fall into three general categories. In the Eggen–Lynden-Bell–Sandage model, the Milky Way formed by the rapid collapse of a single gaseous protogalaxy (*a*). In the Toomre model, several large aggregates of gas merged (*b*). The Searle-Zinn picture is similar to the Toomre model except that the ancestral fragments consisted of much smaller but more numerous pieces (*c*).

MESSIER 83 is a typical type Sc spiral galaxy. The Milky Way probably has a similar appearance, although its arms may be somewhat more tightly coiled.

LARGE MAGELLANIC CLOUD is one of the Milky Way's two largest satellite galaxies. Slowly spiraling into the Milky Way, the cloud will briefly rejuvenate our galaxy at some time in the distant future.

of heavy elements decreases gradually by a factor of 0.8 for every kiloparsec (3,300 light-years) from the center to the edge of the Milky Way disk. Some 70 percent of the 150 or so known globular clusters in the Milky Way exhibit an average metal content of about one twentieth that of the sun. The remainder shows a mean of about one third that of the sun.

Detailed studies of stellar abundances reveal that the ratio of oxygen to iron-peak elements is larger in halo stars than it is in metal-rich disk stars [*see upper illustration on page 75*]. This difference suggests the production of heavy elements during the halo phase of galactic evolution was dominated by supernovae of types Ib, Ic and II. It is puzzling that iron-producing type Ia supernovae, some of which are believed to have resulted from progenitor stars with lifetimes as short as a few hundred million years, did not contribute more to the chemical mixture from which halo stars and some globular clusters formed. This failure would seem to imply that the halo collapsed very rapidly—before supernovae of type Ia could contribute their iron to the halo gas.

That idea, however, conflicts with the four-billion-year age spread observed among galactic globular clusters, which implies that the halo collapsed slowly. Perhaps supernova-driven galactic winds swept the iron-rich ejecta from type Ia supernovae into intergalactic space. Such preferential removal of the ejecta of type Ia supernovae might have occurred if supernovae of types Ib, Ic and II exploded primarily in dense gas clouds. Most of type Ia supernovae then must have detonated in less dense regions, which are more easily swept out by the galactic wind.

Despite the quantity of data, information about metal content has proved insufficient to settle the controversy concerning the time scale of disk and halo formation. Sandage and his colleague Gary A. Fouts of Santa Monica College find evidence for a rather monolithic collapse. On the other hand, John E. Norris and his collaborators at the Australian National Observatory, among others, argue for a significant decoupling between the formation of halo and disk. They also posit a more chaotic creation of the galaxy, similar to that envisaged by Searle and Zinn.

Such differences in interpretation often reflect nearly unavoidable effects arising from the way in which particular samples of stars are selected for study. For example, some stars exhibit chemical compositions similar to those of "genuine" halo stars, yet they have kinematics that would associate them with one of the subcomponents of the disk. As vital as it is, chemical information alone does not resolve ambiguities about the formation of the galactic halo and disk. "Cats and dogs may have the same age and metallicity, but they are still cats and dogs" is the way Bernard Pagel of the Nordic Institute for Theoretical Physics in Copenhagen puts it.

As well as telling us about the past history of our galaxy, the disk and halo also provide insight into the Milky Way's probable future evolution. One can easily calculate that almost all of the existing gas will be consumed in a few billion years. This estimate is based on the rate of star formation in the disks of other spirals and on the assumption that the birth of stars will continue at its present speed. Once the gas has been depleted, no more stars will form, and the disks of spirals will then fade. Eventually the galaxy will consist of nothing more than white dwarfs and black holes encapsulated by the hypothesized dark matter corona.

Several sources of evidence exist for such an evolutionary scenario. In 1978 Harvey R. Butcher of the Kapteyn Laboratory in the Netherlands and Augustus Oemler, Jr., of Yale found that dense clusters of galaxies located about six billion light-years away still contained numerous spiral galaxies. Such spirals are, however, rare or absent in nearby clusters of galaxies. This observation shows that the disks of most spirals in dense clusters must have faded to invisibility during the past six billion years. Even more direct evidence for the swift evolution of galaxies comes from the observation of so-called blue galaxies. These galaxies are rapidly generating large stars. Such blue galaxies seem to be less common now than they were only a few billion years ago.

Of course, the life of spiral galaxies can be extended. Copious infall of hydrogen from intergalactic space might replenish the gas supply. Such infall can occur if a large gas cloud or another galaxy with a substantial gas reservoir is nearby. Indeed, the Magellanic Clouds will eventually plummet into the Milky Way, briefly rejuvenating our galaxy. Yet the Milky Way will not escape its ultimate fate. Like people and civilizations, stars and galaxies leave behind only artifacts in an evolving, ever dynamic universe.

FURTHER READING

Galactic Astronomy: Structure and Kinematics. Dimitri Mihalas and James Binney. W. H. Freeman and Company, 1981.

The Milky Way as a Galaxy. Gerard Gilmore, Ivan R. King and Pieter C. van der Kruit. University Science Books, Mill Valley, Calif., 1990.

The Formation and Evolution of Star Clusters. Edited by Kenneth Janes. Astronomical Society of the Pacific, 1991.

The Stellar Populations of Galaxies. Edited by B. Barbuy and A. Renzini. Kluwer Academic Publishers, Dordrecht, Holland, 1992.