Evolution of Stars and Stellar Populations

Maurizio Salaris
Astrophysics Research Institute,
Liverpool John Moores University, UK

Santi Cassisi
INAF-Astronomical Observatory of Collurania,
Teramo, Italy

John Wiley & Sons, Ltd

Evolution of Stars and Stellar Populations

Evolution of Stars and Stellar Populations

Maurizio Salaris
Astrophysics Research Institute,
Liverpool John Moores University, UK

Santi Cassisi
INAF-Astronomical Observatory of Collurania,
Teramo, Italy

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 33 Park Road, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 22 Worcester Road, Etobicoke, Ontario, Canada M9W 1L1

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Library of Congress Cataloging in Publication Data

Salaris, Maurizio.
Evolution of stars and stellar populations / Maurizio Salaris, Santi Cassisi.
p. cm.
Includes bibliographical references and index.
ISBN-13 978-0-470-09219-X (cloth : alk. paper) ISBN-10 0-470-09219-X (cloth : alk. paper)
ISBN-13 978-0-470-09220-3 (pbk. : alk. paper) ISBN-10 0-470-09220-3 (pbk. : alk. paper)
1. Stars—Evolution. 2. Stars—Populations. 3. Galaxies—Evolution. I. Cassisi, Santi. II. Title.
QB806.S25 2005
523.8′8—dc22

2005021402

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-09219-X (HB) 978-0-470-09220-3 (PB)
ISBN-10 0-470-09219-X (HB) 0-470-09220-3 (PB)

Typeset in 10.5/12.5pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire
This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

To Suzanne and my parents

To all the people who really loved me

Contents

Preface

The theory of stellar evolution is by now well established, after more than half a century of continuous development, and its main predictions confirmed by various empirical tests. As a consequence, we can now use its results with some confidence, and obtain vital information about the structure and evolution of the universe from the analysis of the stellar components of local and high redshift galaxies.

A wide range of techniques developed in the last decades make use of stellar evolution models, and are routinely used to estimate distances, ages, star formation histories and the chemical evolution of galaxies; obtaining this kind of information is, in turn, a necessary first step to address fundamental cosmological problems like the dynamical status and structure of the universe, the galaxy formation and evolution mechanisms. Due to their relevance, these methods rooted in stellar evolution should be part of the scientific background of any graduate and undergraduate astronomy and astrophysics student, as well as researchers interested not only in stellar modeling, but also in galaxy and cosmology studies.

In this respect, we believe there is a gap in the existing literature at the level of senior undergraduate and graduate textbooks that needs to be filled. A number of good books devoted to the theory of stellar evolution do exist, and a few discussions about the application of stellar models to cosmological problems are scattered in the literature (especially the methods to determine distances and ages of globular clusters). However, an organic and self-contained presentation of both topics, that is also able to highlight their intimate connections, is still lacking. As an example, the so-called 'stellar population synthesis techniques' – a fundamental tool for studying the properties of galaxies – are hardly discussed in any existing stellar evolution textbook.

The main aim of this book is to fill this gap. It is based on the experience of one of us (MS) in developing and teaching a third year undergraduate course in advanced stellar astrophysics and on our joint scientific research of the last 15 years. We present, in a homogeneous and self-contained way, first the theory of stellar evolution, and then the related techniques that are widely applied by researchers to estimate cosmological parameters and study the evolution of galaxies.

The first chapter introduces the standard Big Bang cosmology and highlights the role played by stars within the framework of our currently accepted cosmology. The two following chapters introduce the basic physics needed to understand how stars work, the set of differential equations that describes the structure and evolution of stars, and the numerical techniques to solve them.

Chapters 4 to 7 present both a qualitative and quantitative picture of the life cycle of single stars (although we give some basic information about the evolution of interacting binary systems when dealing with Type Ia supernovae progenitors) from their formation to the final stages. The emphasis in our presentation is placed on those properties that are needed to understand and apply the methods discussed in the rest of the book, that is, the evolution with time of the photometric and chemical properties (i.e. evolution of effective temperatures, luminosities, surface chemical abundances) of stars, as a function of their initial mass and chemical composition.

The next chapter describes the steps (often missing in stellar evolution books) necessary to transform the results from theoretical models into observable properties. Finally, Chapters 9 to 11 present an extended range of methods that can be applied to different types of stellar populations – both resolved and unresolved – to estimate their distances, ages, star formation histories and chemical evolution with time, building on the theory of stellar evolution we have presented in the previous chapters.

We have included a number of references which are not meant to be a totally comprehensive list, but should be intended only as a first guide through the vast array of publications on the subject.

This book has greatly benefited from the help of a large number of friends and colleagues. First of all, special thanks go to David Hyder (Liverpool John Moores University) and Lucio Primo Pacinelli (Astronomical Observatory of Collurania) for their invaluable help in producing many of the figures for this book. Katrina Exter is warmly thanked for her careful editing of many chapters; Antonio Aparicio, Giuseppe Bono, Daniel Brown, Vittorio Castellani, Carme Gallart, Alan Irwin, Marco Limongi, Marcella Marconi and Adriano Pietrinferni are acknowledged for many discussions, for having read and commented on various chapters of this book and helped with some of the figures. We are also indebted to Leo Girardi, Phil James, Kevin Krisciunas, Bruno Leibundgut, Luciano Piersanti and Oscar Straniero for additional figures included in the book.

Sue Percival and Phil James are warmly acknowledged for their encouragement during the preparation of the manuscript, Suzanne Amin and Anna Piersimoni for their endless patience during all these months. We are also deeply indebted to Achim Weiss and the Max Planck Institut für Astrophysik (MS), Antonio Aparicio and the Instituto de Astrofisica de Canarias (SC) for their invitation and hospitality. During our stays at those institutes a substantial part of the manuscript was prepared. Finally, we wish to send a heartfelt thank-you to all colleagues with whom we have worked in the course of these wonderful years of fruitful scientific research.

Liverpool **Maurizio Salaris**
Teramo **Santi Cassisi**
March 2005

1 Stars and the Universe

1.1 Setting the stage

Stars are not distributed randomly in the universe, but are assembled through gravitational interactions into galaxies. Typical distances between stars in a given galaxy are of the order of 1 parsec (pc) whereas distances between galaxies are typically of the order of 100 kpc–1 Mpc (1 pc is the distance at which the semi-major axis of Earth orbit subtends an angle of 1 arcsecond; this corresponds to ~3.26 light years, where one light year is the distance travelled by light in one year, i.e. 9.4607×10^{17} cm).

There are three basic types of galaxies: spirals, ellipticals and irregulars (see Figure 1.1). Spiral galaxies (our galaxy, the Milky Way, is a spiral galaxy) constitute more than half of the bright galaxies that we observe within ~100 Mpc of the Sun. They generally comprise a faint spherical halo, a bright nucleus (or bulge) and a disk that contains luminous spiral arms; spirals have typical masses of the order of $10^{11} M_{\odot}$ ($1 M_{\odot}$ denotes one solar mass, i.e. 1.989×10^{33} g). Spirals are divided into normal and barred spirals, depending on whether the spiral arms emerge from the nucleus or start at the end of a bar springing symmetrically from the nucleus. Dust and young stars are contained in the disk whereas the nucleus and halo are populated by older stars. Elliptical galaxies account for ~10 per cent of the bright galaxies, have an elliptical shape, no sign of a spiral structure nor of dust and young stars, a mass range between $\sim 10^5$ and $\sim 10^{12} M_{\odot}$, and in general resemble the nuclei of spirals. There is no sign of significant rotational motions of the stars within ellipticals, whereas stars in the disks of spirals show ordered rotational motion.

These two broad types of galaxies are bridged morphologically by the so-called lenticular galaxies, which make up about 20 per cent of the galaxies, and look like elongated ellipticals without bars and spiral structure. The third broad group of galaxies are the irregulars, that show no regular structure, no rotational symmetry and are relatively rare and faint.

Evolution of Stars and Stellar Populations Maurizio Salaris and Santi Cassisi

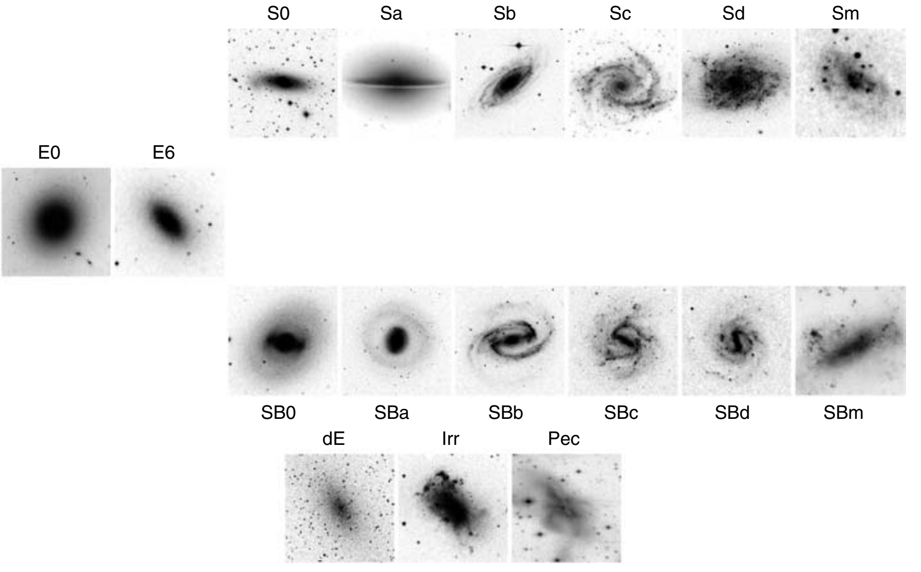

Figure 1.1 The so-called *tuning fork diagram*, i.e. the galaxy morphological classification. Elliptical galaxies are denoted by E (the various subclasses are denoted by the approximate value of the ellipticity) spirals by S, barred spirals by SB; examples of Dwarf elliptical (dE), Irregular (Irr) and Peculiar (Pec) galaxies are also displayed (courtesy of P. James)

Many galaxies show various types of non-thermal emission over a large wavelength range, from radio to X-ray, and are called active galaxies. These active galaxies display a large range of properties that can probably be explained invoking one single mechanism (possibly related to accretion of matter onto a black hole); the difference in their properties is most likely due to the fact that we are observing the same kind of object at different angles, and therefore we see radiation from different regions within the galaxy. Examples of active galaxies are the Seyfert galaxies, radio galaxies, BL Lac objects and quasars. There are also so-called starburst galaxies, e.g. galaxies displaying a mild form of activity, and showing a strong burst of star formation.

For many years it was believed that galaxies extend as far as they are visible. However, starting from the 1970s, the orbits of neutral hydrogen clouds circling around individual spiral galaxies provided rotation curves (e.g. rotational velocity as a function of the distance d from the galactic centre) that, instead of dropping as $\sqrt{d}$ beyond the edge of the visible matter distribution (as expected from Keplerian orbits after the limit of the mass distribution is reached) show a flat profile over large distances well beyond this limit. This can be explained only by a steady increase with distance of the galaxy mass, beyond the edge of the visible mass distribution. This dark matter reveals its presence only through its gravitational pull, since it does not produce any kind of detectable radiation.

Even galaxies are not distributed randomly in the universe, but are aggregated in pairs or groups, which in turn are often gathered into larger clusters of galaxies. Our galaxy (often referred to as the Galaxy) belongs to the so-called Local Group of galaxies, that includes about 20 objects (mainly small) among them the Large Magellanic Cloud (LMC) the Small Magellanic Cloud (SMC) and Andromeda (M31). The nearest cluster of galaxies is the Virgo cluster (at a distance of about 20 Mpc). Further away are other galaxy clusters, among them the Coma cluster, located at a distance of about 100 Mpc, that contains thousands of objects. Deep galaxy surveys (e.g. the APM, COSMOS, 2dF and SDSS surveys) have studied and are still probing the distribution of galaxies in the universe, and have revealed even more complex structures, like filaments, sheets and superclusters, that are groupings of clusters of galaxies.

Dark matter is also found within clusters of galaxies. This can be inferred studying the X-ray emission of the hot ionized intracluster gas that is accelerated by the gravitational field of the cluster. A rough comparison of visible and dark matter contribution to the total matter density of the universe tells us that about 90 per cent of the matter contained in the universe is dark.

It is evident from this brief description that overall the universe appears to be clumpy, but the averaged properties in volumes of space of the order of 100 Mpc are smoother, and the local inhomogeneities can be treated as perturbations to the general homogeneity of the universe.

The dynamical status of the universe is revealed by spectroscopic observations of galaxies. The observed redshift of their spectral lines shows that overall galaxies are receding from us (in the generally accepted assumption that the observed redshift is due to the Doppler effect) with a velocity v that increases linearly with their distance D, so

$$v = H_0 \times D,$$

as first discovered by Hubble and Humason during the 1920s (hence the name of Hubble law for this relationship). The constant H_0 is called the Hubble constant. Taken at face value this relationship seems to locate us in a privileged point, from where all galaxies are escaping. However, if one considers the overall homogeneity of the universe, the same Hubble law has to apply to any other location and the phenomenon of the recession of the galaxies might be looked upon as an expansion of the universe as a whole; a useful and widely used analogy is that of the two-dimensional surface of a balloon that is being inflated. If the galaxies are points drawn on the surface of the balloon, they will appear to be receding from each other in the same way as the Hubble law, irrespective of their location.

Superimposed on the general recession of galaxies are local peculiar velocities due to the gravitational pull generated by the local clumpiness of the universe. For example, the Milky Way and M31 are moving towards each other at a speed of about $120\,\mathrm{km\,s^{-1}}$, and the Local Group, is approaching the Virgo Cluster at a speed of $\sim 170\,\mathrm{km\,s^{-1}}$. On a larger scale, the Local Group, Virgo Cluster and thousands of other galaxies are streaming at a speed of about 600 km s^{-1} towards the so called Great Attractor, a concentration of mass in the Centaurus constellation, located at a distance of the order of 70 Mpc. These peculiar velocities become negligible with respect to the general recession of the galaxies (Hubble flow) when considering increasingly distant objects, for which the recession velocity predicted by the Hubble law is increasingly high.

Another discovery of fundamental importance for our understanding of the universe was made serendipitously in 1965 by Penzias and Wilson. Observations of electromagnetic radiation in a generic frequency interval reveal peaks associated with discrete sources – i.e. stars or galaxies – located at specific directions; when these peaks are eliminated there remains a dominant residual radiation in the microwave frequency range. The spectrum of this cosmic microwave background (CMB) radiation is extremely well approximated by that of a black body with a temperature of 2.725 K. After removing the effect of the local motion of the Sun and of our galaxy, the CMB temperature is to a first approximation constant when looking at different points in the sky, suggesting a remarkable isotropy which is hard to explain in terms of residual emission by discrete sources. From the CMB temperature one easily obtains the energy density associated to the CMB, $\epsilon_{\rm CMB}$, given that $\rho_{\rm CMB} = \epsilon_{\rm CMB}/c^2 \sim 4.64 \times 10^{-34}$ g cm^3

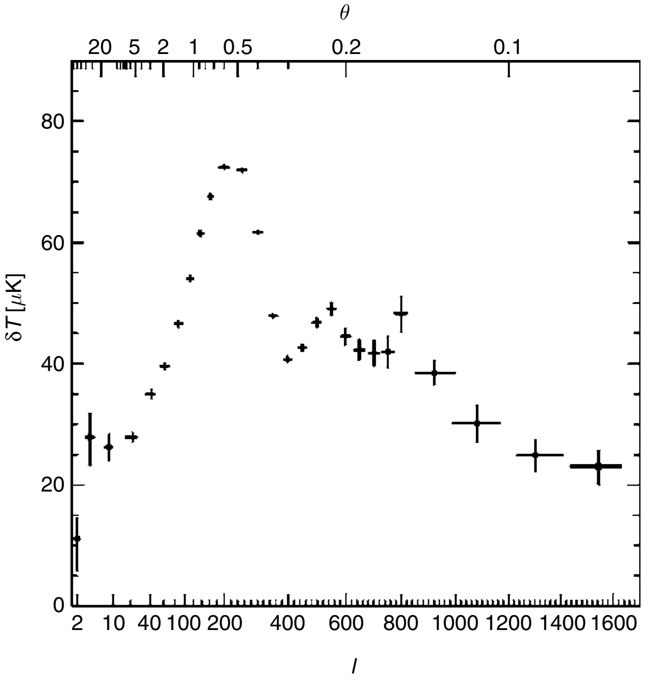

Figure 1.2 Plot of the CMB temperature fluctuations (in units of 10^{-6} K) as a function of the angular scale in degrees (upper horizontal axis) and the so-called wave number $l \sim \pi/\theta$ (lower horizontal axis); this is also called the power spectrum of the CMB fluctuations

(c denotes the speed of light, $2.998 \times 10^{10}\,\mathrm{cm\,s^{-1}}$). This CMB photon density is the dominant component of the present radiation density in the universe; a rough comparison of ρ_{CMB} with the present matter density ρ shows that at the present time the density associated with the photons is about three orders of magnitude lower than the matter density, including the dark matter contribution.

In 1992 the COBE satellite first discovered tiny variations δT of the CMB temperature, of the order of $\delta T/T \sim 10^{-5}$ (where T is the global mean of the CMB temperature) when looking at different points in the sky. By computing the average over the sky of the ratio $\delta T/T$ (temperature fluctuation) measured from any two points separated by an angle θ, one obtains what is called the angular power spectrum of the CMB temperature anisotropies, displayed in Figure 1.2. This power spectrum shows the existence of a series of peaks located at specific angular scales.

A comprehensive theory for the structure and evolution of the universe must be able to explain the basic observations outlined above in terms of evolutionary processes rooted in accepted physics theories. The following sections introduce briefly the Hot Big Bang theory, which is the presently widely accepted cosmological theory. Detailed presentations of cosmology at various levels of complexity can be found in [11], [57], [118] and [142].

1.2 Cosmic kinematics

A cornerstone of the Big Bang theory is the so-called cosmological principle: it states that the large-scale structure of the universe is homogeneous and isotropic. Homogeneity means that the physical properties of the universe are invariant by translation; isotropy means that they are also rotationally invariant. Both these properties can be applied only considering average properties of large volumes of space, where the local structures (galaxies, clusters of galaxies) are smeared out over the averaging volumes.

As discussed before, the adequacy of the cosmological principle can be empirically verified by studying the distribution of clusters of galaxies on scales of the order of 100 Mpc and by the isotropy of the CMB. Locally the universe is clumpy, but this clumpiness disappears when averaging the matter density over large enough volumes. In this way the local clumpiness is treated as a perturbation to the general smoothness of the universe. The universe is then treated as a fluid whose particles are galaxies, moving according to the Hubble law; within this picture of a cosmic fluid the cosmological principle implies that every co-moving observer (i.e. moving with the Hubble flow) in the cosmic fluid has the same history.

A first step when discussing events happening in the universe is to set up an appropriate coordinate system. For the time coordinate a natural choice is to use standard clocks co-moving with the cosmic fluid, that will define a cosmic time t; an operational way to synchronize t for co-moving observers at different locations is to set t to the same value when each observer sees that a property of the cosmic fluid, i.e. the average local density of matter ρ, has reached a certain agreed value. After synchronization, by virtue of the cosmological principle, the two observers must be able to measure exactly the same value (possibly different from the one at the time of synchronization) of that property whenever their clocks show the same time.

As for the three spatial coordinates, the cosmological principle greatly restricts the possible geometries. The assumption of homogeneity and isotropy requires that the tridimensional space has a single curvature, i.e. it must have the same value at all positions, but can in principle depend on time. The space–time interval ds between two events in an homogeneous and isotropic static space can be written as follows

$$ds^2 = c^2 dt^2 - \left(\frac{dr^2}{1 - Kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right)$$

where K is the spatial curvature, dt the cosmic time separation, r the radial coordinate and θ and ϕ the polar and azimuthal angles in spherical coordinates, respectively. The expansion (or contraction) of the universe can be accounted for by redefining the radial coordinate r as $r \equiv R(t)\chi - \chi$ being dimensionless – and the curvature K as $K(t) \equiv k/R(t)^2$. The constant k and coordinate χ are defined in a way that $k = +1$ for a positive spatial curvature, $k = 0$ for a flat space and $k = -1$ for a negative curvature. $R(t)$ is the so-called cosmic scale factor, that has the dimension of a distance and is dependent on the cosmic time t. With these substitutions one obtains the so-called Friedmann–Robertson–Walker (FRW) metric:

$$ds^2 = c^2 dt^2 - R(t)^2 \left(\frac{d\chi^2}{1 - k\chi^2} + \chi^2 d\theta^2 + \chi^2 \sin^2 \theta d\phi^2 \right) \tag{1.1}$$

The values of the three spatial coordinates χ, θ and ϕ are constant for an observer at rest with respect to the expansion of the cosmic fluid. One can easily see that the factor $R(t)$ in Equation (1.1) allows a scaling of the spatial surfaces that depends only on time, thus preserving the homogeneity and isotropy dictated by the cosmological principle. It is important to stress that it is only by virtue of the cosmological principle that we can uniquely define a four-dimensional coordinate system co-moving with the cosmic flow. As an example, the definition of cosmic time would be impossible in a universe without homogeneity and isotropy, because we could not synchronize the various clocks using mean properties (that would not be the same everywhere at a given time t) of the cosmic flow.

The geometrical properties of the three-dimensional space determined by the value of k can be briefly illustrated as follows. Let us consider at a cosmic instant t a sphere with centre at an arbitrary origin where $\chi = 0$, and surface located at a fixed value χ; the difference between the coordinates of the centre of the sphere and the surface is equal to $r = R(t)\chi$. The area A of the spherical surface of coordinate radius $r = R(t)\chi$ is, by definition, $A = 4\pi r^2 = 4\pi R(t)^2 \chi^2$. The physical radius $R_{\rm p}$ of the spherical surface is the distance between the centre and surface of the sphere measured with a standard rod at the same cosmic time t. This means that one has to determine the interval Δs^2 between the two events assuming $dt=0$, so that $R_p=\sqrt{-\Delta s^2}$. From the FRW metric one obtains

$$R_p = R(t)\int_0^{\chi} \frac{d\chi}{\sqrt{1-k\chi^2}} \tag{1.2}$$

R_p is equal to $R(t)\sin^{-1}\chi$, $R(t)\chi$ and $R(t)\sinh^{-1}\chi$ when $k=1, 0$ and -1, respectively. When $k=0$ one has $\chi=R_p/R(t)$, and $A=4\pi R_p^2$, i.e. r is equal to R_p and the area A increases as R_p^2, as in Euclidean geometry. When $k=+1$ one has $r=R(t)\sin(R_p/R(t))$ and $A=4\pi R(t)^2\sin^2(R_p/R(t))$, which reaches a maximum value $A=4\pi R(t)^2$ when $R_p=(\pi/2)R(t)$, then becomes zero when $R_p=\pi R(t)$ and has in general a periodic behaviour. This means that in the case of $k=+1$ space is closed and the periodicity corresponds to different circumnavigations. In the case of $k=-1$ then $A=4\pi R(t)^2\sinh^2(R_p/R(t))$, which increases with R_p faster than in the case of a Euclidean space.

It is easy to see how simply $R(t)$ describes the observed expansion of the universe. Let us set $\chi=0$ at the location of our own galaxy, that is approximately co-moving with the local cosmic fluid (hence its spatial coordinate does not change with time) and consider another galaxy – also at rest with respect to the expansion of the universe – whose position is specified by a value χ of the radial coordinate (the angles θ and ϕ are assumed to be equal to zero for both galaxies). Its proper distance (defined in the same way as for the proper radius R_p discussed before) D at a given cosmic time t is given by:

$$D = R(t)\int_0^{\chi} \frac{d\chi}{\sqrt{1-k\chi^2}}$$

As in the case of Equation (1.2) D is equal to $R(t)\sin^{-1}\chi$, $R(t)\chi$ and $R(t)\sinh^{-1}\chi$ when $k=1, 0$ and -1, respectively. The velocity v of the recession of the galaxy due to the expansion of the universe is

$$v = \frac{dD}{dt} = \frac{dR(t)}{dt}\int_0^{\chi} \frac{d\chi}{\sqrt{1-k\chi^2}} = \frac{dR(t)}{dt}\frac{1}{R(t)}D$$

This looks exactly like the Hubble law; in fact, by writing

$$H(t) = \frac{dR(t)}{dt}\frac{1}{R(t)} \tag{1.3}$$

we obtain

$$v = H(t)\times D$$

$H(t)$ corresponds to the Hubble constant and one can notice that its value can change with cosmic time. The value of $H(t)$ determined at the present time is denoted as H_0.

This result is clearly independent of the location of the origin for the radial coordinate χ, since any position in the universe is equivalent according to the cosmological principle. It is important to notice that locally, e.g. within the solar system or within a given galaxy, one cannot see any effect of the cosmic expansion, since the local gravitational effects dominate. For distances large enough $(D > c/H(t))$ the last equation predicts recession velocities larger than the speed of light, an occurrence that seems to go against special relativity. The contradiction is, however, only apparent, given that galaxies recede from us faster than the speed of light (superluminal recession) because of the expansion of space; locally, they are at rest or moving in their local inertial reference frame with peculiar velocities $\ll c$.

In the following section we will briefly describe the observational counterpart of $v = H(t) \times D$ and show how it probes the evolution of the kinematic status of the universe.

1.2.1 Cosmological redshifts and distances

What we measure to estimate the recession velocity of galaxies is a redshift z, that can be related to the change of $R(t)$ with time. Consider light reaching us (located at $\chi = 0$) from a galaxy at a radial coordinate χ. Two consecutive maxima of the electromagnetic wave are emitted at times t_e and $t_e + \delta t_e$ and received at times t_0 and $t_0 + \delta t_0$; if $\delta t_e = \delta t_0$ we would not observe any redshift since the wavelength of the electromagnetic wave is given by the spatial distance between the two consecutive maxima, i.e. the observed wavelength is $\lambda_0 = c\delta t_0$, and the emitted one is $\lambda_e = c\delta t_e$. We will now find the relationship between δt_e and δt_0. Since $ds = 0$ for light, we have

$$\int_{t_e}^{t_0} \frac{dt}{R(t)} = \frac{1}{c}\int_0^{\chi} \frac{d\chi}{\sqrt{1 - k\chi^2}}$$

$$\int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{R(t)} = \frac{1}{c}\int_0^{\chi} \frac{d\chi}{\sqrt{1 - k\chi^2}}$$

for the first and second maximum, respectively. The right-hand side of both equations is the same, therefore we can write

$$\int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{R(t)} - \int_{t_e}^{t_0} \frac{dt}{R(t)} = 0$$

The first term on the left-hand side of the previous equation can be rewritten as

$$\int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{R(t)} = \int_{t_e}^{t_0} \frac{dt}{R(t)} + \int_{t_0}^{t_0+\delta t_0} \frac{dt}{R(t)} - \int_{t_e}^{t_e+\delta t_e} \frac{dt}{R(t)}$$

and therefore

$$\int_{t_0}^{t_0+\delta t_0} \frac{dt}{R(t)} - \int_{t_e}^{t_e+\delta t_e} \frac{dt}{R(t)} = 0$$

The intervals δt_e and δt_0 are negligible compared with the expansion timescale of the universe, and therefore $R(t)$ is to a good approximation constant during these two time intervals; inserting this condition into the previous equation provides

$$\frac{\delta t_0}{R(t_0)} = \frac{\delta t_e}{R(t_e)}$$

The redshift $z = (\lambda_0 - \lambda_e)/\lambda_e$ is therefore given by

$$z = \frac{\delta t_0}{\delta t_e} - 1 = \frac{R(t_0)}{R(t_e)} - 1 \tag{1.4}$$

In an expanding universe $z > 0$ (since $R(t_0) > R(t_e)$) as observed. If the redshift is small enough, i.e. t_e is close to t_0 in cosmological terms, we can expand $R(t_e)$ about t_0 using the Taylor formula, and retain only the terms up to the second order:

$$R(t_e) = R(t_0) + (t_e - t_0)\frac{dR(t_0)}{dt} + \frac{1}{2}(t_e - t_0)^2 \frac{d^2R(t_0)}{dt^2}$$

We can now define H_0 as

$$H_0 \equiv H(t_0) = \frac{dR(t_0)}{dt}\frac{1}{R(t_0)}$$

i.e. the present value of the Hubble constant, and the so-called deceleration parameter

$$q_0 \equiv -\frac{d^2R(t_0)}{dt^2}\frac{1}{R(t_0)H_0^2} \tag{1.5}$$

Both H_0 and q_0 are related to the present rate of expansion of the universe. H_0 measures the actual expansion rate, whilst q_0 is positive if the expansion is slowing down (hence the name deceleration parameter) or negative if the opposite is true. With these definitions the second-order expansion of $R(t_e)$ can be rewritten as

$$R(t_e) = R(t_0)\left[1 + H_0(t_e - t_0) - \frac{1}{2}q_0 H_0^2 (t_e - t_0)^2\right]$$

and after additional manipulations one obtains the following useful results:

$$z = H_0(t_0 - t_e) + H_0^2(t_0 - t_e)^2\left(1 + \frac{1}{2}q_0\right) \tag{1.6}$$

$$t_0 - t_e = \frac{1}{H_0}\left[z - \left(1 + \frac{1}{2}q_0\right)z^2\right] \tag{1.7}$$

$$\chi = \frac{c}{R(t_0)H_0}\left[z - \frac{1}{2}(1 + q_0)z^2\right] \tag{1.8}$$

These relationships between z, H_0 and q_0 hold in the case of a redshift due to the expansion of the universe. Superimposed on the expansion of the universe are local peculiar velocities (e.g. blue- and redshifts) due to the motions caused by local anisotropies in the matter distribution; an example is local motions in clusters and groups of galaxies due to the gravitational potential of the cluster itself. These effects are minimized by observing suitably distant objects, where the velocities corresponding to the expansion of the universe become so large that they make local peculiar motions negligible.

From an observational point of view, the Hubble law needs, in addition to the measurements of the redshift z, an estimate of galaxy distances. This is usually done by comparing the observed flux l received from certain standard candles (i.e. objects of known intrinsic luminosity L) with their intrinsic luminosities. Traditionally one uses the inverse square law to determine the distance:

$$d = \left(\frac{L}{4\pi l}\right)^{1/2} \tag{1.9}$$

This result is based on the conservation of energy and assumes a flat static space. In cosmology, the distance obtained through Equation (1.9) is called the luminosity distance, and is denoted by d_{L}.

Consider a light source located at a radial co-moving coordinate χ; at a given cosmic time t_{e} the source emits photons that reach the observer located at $\chi = 0$ at time t_0. By the time the light reaches the observer it is distributed uniformly across a sphere of coordinate radius $R(t_0)\chi$. The area of the spherical surface at the observer location centred at the source is therefore given by $4\pi R(t_0)^2\chi^2$. The photons emitted by the source are redshifted by the expansion of the universe, and their energy is therefore reduced by a factor $(1 + z)$ when measured by the observer; this is because the wavelength is increased by a factor $(1 + z)$ and the photon energy is proportional to the inverse of its wavelength. There is also an additional reduction by a factor $(1 + z)$ due to the so-called time dilation effect, i.e. the observer receives less photons per unit time than emitted at the source. This can easily be understood by means of the same arguments as were applied in the case of the wave maxima, that led to the notion of redshift. We found before that the time between two consecutive maxima at emission is different from that at reception; the same holds for the time interval between photons emitted by the source, and implies that the rate of reception of photons is different from the rate of emission. Taking into account these two effects, conservation of energy dictates that:

$$l = \frac{1}{(1+z)^2}\frac{L}{4\pi R(t_0)^2\chi^2}$$

We now define the luminosity distance d_{L} of the observed source, according to Equation (1.9); one obtains $d_{\mathrm{L}} = R(t_0)\chi(1+z)$, which can be rewritten using Equation (1.8) as (retaining the terms up to the second order in z):

$$d_L = \frac{cz}{H_0}\left[1 + \frac{1}{2}(1 - q_0)z\right] \tag{1.10}$$

The first term is the empirical Hubble law, with the recession velocity given by the product cz. The higher-order correction term is proportional to the deceleration parameter q_0 and starts to play a role when $z > 0.1$.

Another way to determine cosmological distances is to consider objects (e.g. galaxies) with known diameter D_p, and compare the measured angular diameters Θ with the intrinsic ones. One can define a diameter distance d_{D_p} as

$$d_{D_p} = \frac{D_p}{\Theta} \tag{1.11}$$

which is equal to d_L for a flat static space. Consider an object located at the radial co-moving coordinate χ, that emits light at time t_e; if the observer is located at $\chi = 0$ and receives the light from the object at t_0, the relationship between D_p and Θ can easily be obtained by determining $\sqrt{\Delta s^2}$ where Δs is obtained integrating the FRW metric with $dt = d\chi = d\phi = 0$. This provides $d_{D_p} = R(t_e)\chi$. By comparing the latter equation with $d_L = R(t_0)\chi(1 + z)$ obtained before and using the definition of z we obtain

$$d_{D_p} = \frac{d_L}{(1 + z)^2}$$

In principle d_{D_p} is different from d_L, but the two distances converge to the same value when $z \to 0$.

It should be clear from this brief discussion that the empirical study of the trends of d_L and d_{D_p} with redshift z provides an estimate of the kinematical parameters H_0 and q_0. A third possible method to determine the kinematical status of the universe involves number counts of galaxies with a flux greater than some specified value l $(N(l))$. Assuming there are n galaxies per unit volume, in a static flat universe (with uniform distribution of galaxies) one expects

$$N(l) = \frac{4}{3}\pi n\left(\frac{L}{4\pi l}\right)^{3/2}$$

where L is the intrinsic galaxy luminosity, supposed constant. For an expanding universe it can be shown that (as a second-order approximation in z)

$$N(l) = \frac{4\pi n(t_0)}{3}\left(\frac{L}{4\pi l}\right)^{3/2}\left[1 - \frac{3H_0}{c}\left(\frac{L}{4\pi l}\right)^{1/2}\right]$$

where $n(t_0)$ is the number density at the present time (e.g. in the low redshift universe); notice that by a fortuitous cancellation of terms this relationship does not depend on q_0. The correction term to the static flat case is always negative, so that in principle one should always observe fewer sources than predicted by the simple $l^{-3/2}$ formula.

There are many practical difficulties in implementing these three tests; the reason is that we are assuming the existence of perfect standard candles and the absence of evolutionary effects on the size, and brightness of galaxies. Evolutionary effects are particularly important since a high redshift means a time far in the past, when galaxies had a very different age from the present one. A detailed discussion of these classical cosmological tests and the related observational problems can be found in [187]. In recent years the class of stellar objects called *Type Ia supernovae* (see Section 7.6) has been used as an effective standard candle and applied with great success to study the d_{L}–z relationship (see [146]).

We conclude this section by discussing briefly the concept of particle horizon in an FRW expanding universe. In general, as the universe expands and ages, a generic observer is able to see increasingly distant objects as the light they emitted has time to arrive at the observer's location. This implies that as time increases, increasingly larger regions of space come into causal contact with the observer, who will therefore be able to 'see' increasingly larger portions of the universe. We can ask ourselves what is the co-moving coordinate χ_H of the most distant galaxy we can see at a given cosmic time t. Increasing values of χ_H with time mean that we are actually seeing more and more distant galaxies (supposed to be at rest with respect to the cosmic expansion) as the time increases. Consider a radially travelling photon, for which $ds = 0$. From the FRW metric we obtain

$$\int_0^t \frac{dt'}{R(t')} = \frac{1}{c} \int_0^{\chi_H} \frac{d\chi}{\sqrt{1 - k\chi^2}}$$

and therefore

$$\begin{aligned} \chi_H &= \sin\left(c \int_0^t \frac{dt'}{R(t')} \right) \quad k = +1 \\ \chi_H &= c \int_0^t \frac{dt'}{R(t')} \quad k = 0 \\ \chi_H &= \sinh\left(c \int_0^t \frac{dt'}{R(t')} \right) \quad k = -1 \end{aligned} \tag{1.12}$$

If the space has $k = 0$ or $k = -1$ it is in principle possible, for specific forms of $R(t)$, to have an infinite χ_H; this means that all galaxies in the universe might eventually be visible at a certain time t for particular forms of the function $R(t)$. If $k = +1$ the behaviour of χ_H is periodic, and if the argument of the sine function is equal to or larger than π, one can sweep the entire universe.

1.3 Cosmic dynamics

The previous discussion about the kinematics of the cosmic fluid was based exclusively on the properties of the FRW metrics which, in turn, depend only on the hypothesis of homogeneous and isotropic cosmic fluid. To determine the behaviour of $R(t)$ with cosmic time t and the value of k we need to apply a theory for the physical force(s) governing the evolution of the cosmic fluid. The only fundamental interaction able to bridge the relevant cosmological scale is the gravitational force, therefore we need to use a theory of gravity – the general relativity theory – to describe the evolution of FRW universes.

The case of a space with the FRW metrics provides the equation

$$\left(\frac{dR(t)}{dt}\right)^2 = -kc^2 + \frac{8\pi G\rho(t)R(t)^2}{3} \tag{1.13}$$

where G is the gravitational constant ($6.6742 \times 10^{-8}\,\mathrm{dyn\,cm^2\,g^{-2}}$) and ρ is the matter density. Equation (1.13) was obtained in 1922 by Friedmann, who solved Einstein's field equations for an isotropic and homogeneous universe. As we will see in a moment, these equations predict an expanding universe. A more general form of the field equations contains the constant Λ – called the cosmological constant – introduced by Einstein in 1917 in order to obtain static universes (the expansion of the universe had not been discovered yet). It is important to notice that the value of Λ must be small in absolute terms, since the planetary motions in the solar system are well described by the Einstein field equations with $\Lambda = 0$. Including Λ in the gravitational field equations provides

$$\left(\frac{dR(t)}{dt}\right)^2 = -kc^2 + \frac{(8\pi G\rho(t) + \Lambda)R(t)^2}{3} \tag{1.14}$$

It is clear that the evolution of $R(t)$ is controlled by the density (ρ), the geometry (k) and the cosmological constant (Λ). By using the definition of $H(t)$ one can rewrite Equation (1.14) as

$$H(t)^2 = -\frac{kc^2}{R(t)^2} + \frac{8\pi G\rho(t)}{3} + \frac{\Lambda}{3} \tag{1.15}$$

It is customary to introduce the critical density $\rho_c \equiv 3H(t)^2/(8\pi G)$ and define the density parameter $\Omega_\rho = \rho/\rho_c$, an equivalent for the cosmological constant $\Omega_\Lambda = \Lambda/(3H(t)^2)$, and the sum $\Omega = \Omega_\rho + \Omega_\Lambda$. With these definitions Equation (1.15) becomes

$$(1 - \Omega)H(t)^2 R(t)^2 = -kc^2 \tag{1.16}$$

We can immediately see from this form of the Friedmann equation that there is an intimate connection between the density of matter plus the cosmological constant, and the geometry of space. $\Omega = 1$ gives a flat space, $\Omega > 1$ a positive curvature, and $\Omega < 1$ a negative curvature. It is also important to notice the obvious fact that Ω changes with time, since $H(t)$ and $R(t)$ both change with t, but the product kc^2 is a constant.

1.3.1 Histories of $R(t)$

Equation (1.14) enables us to perform a simple analysis of the behaviour of $R(t)$ for various model universes, once an additional equation for the density is obtained; this equation can be determined by applying the first principle of thermodynamics to the cosmic fluid. In an isolated system the first law of thermodynamics states that $dU = -PdV$ where U is the internal energy of the system, V its volume and P the pressure. The internal energy is ρc^2 times the volume V (i.e. the energy associated with the rest mass of the matter) so that the time evolution of the system according to the first law is

$$\frac{d(\rho(t)c^2V(t))}{dt} = -P\frac{dV(t)}{dt}$$

which can also be rewritten as

$$\frac{d(\rho(t)c^2R(t)^3)}{dt} = -P\frac{dR(t)^3}{dt} \tag{1.17}$$

using the fact that the volume V scales as $R(t)^3$. Let us now assume that the density is dominated by matter and not by radiation; this is a very good assumption since observationally – as discussed before – one finds that at the present time the matter density is about three orders of magnitude larger than the density associated with radiation ($\rho_{\rm r} = \epsilon_{\rm r}/c^2$, where $\epsilon_{\rm r}$ is the photon energy density). If the matter is non-relativistic (a correct assumption for almost the whole evolution of the universe) its pressure is negligible with respect to ρc^2 and Equation (1.17) provides

$$\frac{d\rho(t)}{dt}\frac{1}{\rho(t)} = -3\frac{dR(t)}{dt}\frac{1}{R(t)} \tag{1.18}$$

which implies

$$\rho(t)R(t)^3 = \rho(t_0)R(t_0)^3$$

where t_0 is the present cosmic time and t a generic value. This reflects the simple fact that the density of non-relativistic matter is decreasing because of dilution as space is expanding. If photons were to be the dominant contributor to the total density, the previous relationship would be different. In fact, for photons (and more generally for relativistic particles) $P = (\rho_r c^2)/3$ and P is no longer negligible with respect to U. Therefore Equation (1.17) would provide

$$\frac{d\rho_r(t)}{dt}\frac{1}{\rho_r(t)} = -4\frac{dR(t)}{dt}\frac{1}{R(t)} \tag{1.19}$$

and

$$\rho_r(t)R(t)^4 = \rho_r(t_0)R(t_0)^4$$

The scaling of $\rho(t)$ with $R(t)^4$ is firstly due to the decrease of the number density of photons as $R(t)^{-3}$ when the universe expands (since the volume increases as $R(t)^3$). In addition, the energy of individual photons decreases as $R(t)^{-1}$ because of the cosmological redshift and therefore both ϵ_r and ρ_r decrease with time as $R(t)^{-4}$, faster than the matter density.

By considering a matter dominated universe one now can rewrite Equation (1.14) as

$$\left(\frac{dR(t)}{dt}\right)^2 = -kc^2 + \frac{8\pi G\rho(t_0)R(t_0)^3}{3R(t)} + \frac{\Lambda R(t)^2}{3} \tag{1.20}$$

Differentiation of this equation with respect to t provides:

$$\frac{d^2R(t)}{dt^2} = -\frac{4\pi G\rho(t_0)R(t_0)^3}{3R(t)^2} + \frac{\Lambda R(t)}{3} \tag{1.21}$$

This equation shows clearly how the self gravitation of matter (represented by ρ) acts to slow down the expansion of the universe, because it appears as a negative contribution to the acceleration of $R(t)$. On the other hand, a positive Λ acts like a negative density and tends to accelerate the expansion of the universe; a particular choice of Λ makes the universe static (although in a situation of unstable equilibrium). The term $(\Lambda R(t))/3$ is often called the cosmic repulsion term.

It is now easy to determine some general properties of $R(t)$ in a matter dominated universe. If Λ is zero or negative the acceleration of $R(t)$ is always negative; at some time in the past $R(t)$ must have reached zero and therefore ρ was infinite (i.e. a singularity is attained). It is natural to set the zero point of the cosmic time at this instant, which can also be considered the origin of the universe. As for the future evolution, if Λ is negative, $R(t)$ will also intersect the t axis some time in the future (hence a final implosion) since the expansion will slow down, eventually stop and then reverse to a contraction. If Λ is zero the acceleration can become zero in the future if $R(t)$ becomes infinite, and therefore the expansion can slow down without ever being followed by a contraction. The precise behaviour depends in this case on the value of k. If $k = -1$ or 0 the future collapse is avoided, but not if $k = +1$.

If Λ is positive then $R(t)$ is not always decelerating and there is the possibility of avoiding a singularity in the past. In fact, if $k = +1$ one can obtain from Equations (1.20) and (1.21) that in the past there has been a minimum of $R(t)$ different from zero, given by $R^3_{\text{min}} = (4\pi G\rho(t_0)R(t_0)^3)/\Lambda$ if the cosmological constant satisfies the following relation: $\Lambda < (c^6)/(4\pi G\rho(t_0)R(t_0)^3)^2$. As for the future evolution, if $k = 0$

or -1 the expansion continues forever, whereas if $k=+1$ the expansion may vanish and then be followed by a contraction, depending upon the value of Λ.

For historical interest we show briefly how it is possible to obtain a static universe by tuning the value of Λ. In a static universe both $R(t)$ and $\rho(t)$ are constant, and both velocity and acceleration of $R(t)$ are equal to zero. With these constraints Equations (1.20) and (1.21) provide $\Lambda=4\pi\rho(t_0)G$, $k/R^2=(4\pi\rho(t_0)G)/c^2$, where R denotes the constant value of $R(t)$. Since R has to be positive and k can be only equal to 0, +1, −1, we have that a static universe will have $k=+1$ and $R=c/\sqrt{4\pi\rho(t_0)G}$.

We conclude by providing analytical relationships between $R(t)$ and t for the case of flat geometry, i.e. $\Omega=1$ and $k=0$, and arbitrary values of Λ, which are relevant to the presently favoured cosmological model. With this choice of parameters the universe began from a singular state ($R=0$ and $\rho=\infty$ at $t=0$) and Equation (1.20) gives (see also Figure 1.3):

$$\begin{aligned}
R(t) &= R(t_0)\left(\frac{8\pi G\rho(t_0)}{\Lambda}\right)^{1/3}\sinh^{2/3}\left(\frac{1}{2}t\sqrt{3\Lambda}\right) \quad \Lambda>0\\
R(t) &= R(t_0)(6\pi G\rho(t_0))^{1/3}t^{2/3} \quad \Lambda=0\\
R(t) &= R(t_0)\left(\frac{8\pi G\rho(t_0)}{|\Lambda|}\right)^{1/3}\sin^{2/3}\left(\frac{1}{2}t\sqrt{3|\Lambda|}\right) \quad \Lambda<0
\end{aligned} \tag{1.22}$$

For $\Lambda=0$ one obtains the very simple result $q_0=1/2$, $H(t)=2/(3t)$ and therefore the age of the universe is $t_0=2/(3H_0)$. The quantity $1/H_0$ is often called Hubble time.

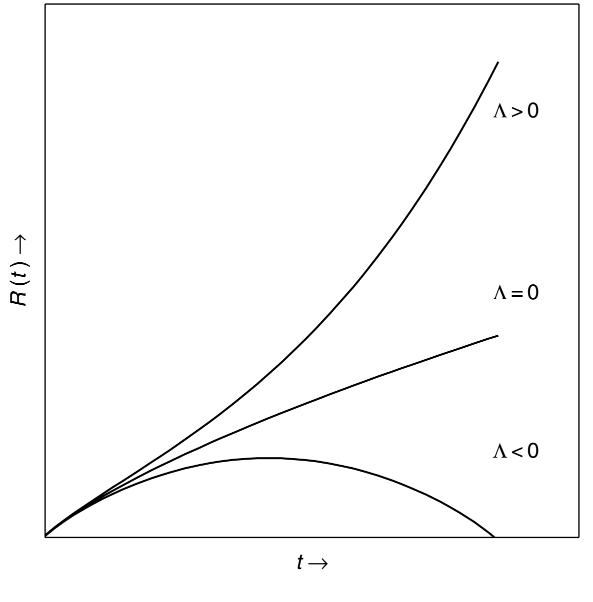

Figure 1.3 Qualitative behaviour of the scale factor $R(t)$ with respect to the cosmic time t for models with $\Omega=1$ and $k=0$