Structure and Evolution of Galaxies

Tsutomu T. Takeuchi[1]

1. Division of Particle and Astrophysical Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya 464–8602, Japan

We present an overview of the basics of galaxy evolution. Especially we focus on the theoretical aspects of the chemical evolution of galaxies and structure formation in the Universe.

1 Introduction

Galaxies evolve. This fact has been a strong driving force of the studies of galaxies. However, how galaxies formed and evolved is still far from well-understood. The evolution of galaxies is a change of their various physical properties as a function of galaxy age, or, from a more cosmological point of view, a function of the cosmic age.

From a cosmological point of view, dark matter interacts gravitationally to form dark halos, and galaxies form and evolve in the halos. The most characteristic aspect of galaxy evolution is the evolution of star formation (SF) activities. Stars formed in a primordial gas clump, then this first generation of stars died and expelled the first heavy elements in the Universe. This made the formation of the next generation of stars much more efficient, and triggered the collective formation of stars, i.e. galaxy formation. Subsequently, the chemical composition of the interstellar medium (ISM) in galaxies changed with time, along with the cycle of the birth and death of stars. This is referred to as the chemical evolution of galaxies (Tinsley, 1980). It is also related to the formation and evolution of dust, another major player in galaxy evolution (e.g., Asano et al., 2013a,b, 2014).

In addition to the internal evolution of galaxies, mergers drastically change the morphology of galaxies. Next, environmental effects become more significant in, for example, clusters of galaxies. Recently, in less dense environments the evolution of galaxies has been found to be affected by the infall of small galaxies or cold gas infall from the large-scale structure. Such "external" factors are also one of the important aspects of galaxy evolution.

It is known that there exist massive black holes at the centers of most of galaxies. The activity of an active galactic nucleus (AGN) depends on the mass accretion rate to the black hole. There is an empirical correlation between the mass of the spheroid component and the central black hole in galaxies (Magorrian et al., 1998). This suggests a hidden connection of the evolution of black holes and galaxies, known as "the co-evolution".

As seen above, galaxy evolution has various aspects and it is difficult to cover all the topics in this lecture. Here, we restrict ourselves to topics related to the chemical evolution and structure formation in the Universe. Other related topics can be found in other articles in this volume.

2 Chemical Evolution of Galaxies

2.1 Nucleosynthesis in stars

The life of a star is almost uniquely determined by its initial mass when it was born. The lowest mass stars live very long, and cool down at the end of their life. Slightly more massive stars like the Sun live shorter than them, experience a thermal pulse during the late evolutionary phase referred to as the asymptotic giant branch (AGB), and finally end by forming white dwarfs and planetary nebulae (PNe). Massive stars heavier than 8 $M_{\odot}$ live very short lives, end them with a violent explosion known as a Type'II supernova (SN II). As for binary stars, a certain fraction of them end their life as a different type of supernova, Type Ia (SN Ia).

The maximum atomic number of elements synthesized at the center of stars also depends on the initial mass of stars. Solar-mass stars synthesize C and O at the end, through nuclear fusion. More massive stars can produce Si and Mg. The most massive stars finally synthesize Fe and iron-family elements. Since the binding energy of the nucleus is the largest at Fe, stars cannot synthesize elements heavier than Fe. Elements heavier than Fe are produced during the final phase of stellar life, through a neutron capture process.

We should note that we need a physical process to get back the synthesized heavy elements from the center of stars. Most stars blow away their mass with the stellar wind, but this process is not very efficient except for the most massive stars. A much more efficient way to expel mass from stars are the PNe and SNe. The elements provided to the ISM depend on the mass ejection process, and, when combined with the different lifetimes of stars, this makes the chemical evolution process very complicated. In this lecture, we formulate the classical chemical evolution in its simplest form.

2.2 The basic framework of chemical evolution

The ISM of galaxies is compressed and cooled by various mechanisms, and forms dense gas clouds. At the center of these clouds, nuclear reactions begin and the clouds become stars. Stars produce heavy elements and return them back to the ISM. Gas mixed with the heavy elements forms stars more easily than pristine gas because they can cool by various line radiation processes, and accelerate subsequent star formation of the galaxy in which they reside. However, less massive stars live long and hardly return gas into the ISM. Hence, the remaining gas is reduced and finally runs out. This is the end of the star formation of a galaxy. Star formation can revive when new gas is provided, e.g. via merging or infall of other galaxies.

2.2.1 Initial mass function (IMF)

One of the most fundamental ingredients of the chemical evolution, but at the same time the most uncertain part, is the stellar initial mass function (IMF). The mass distribution function of stars at their birth is referred to as the IMF. Since the IMF determines the ratio between massive and less massive stars, it plays a crucial role in the chemical evolution of galaxies. However the IMF shape is not well determined, either observationally or theoretically. Thus in the studies of galaxies, we usually assume a certain functional form for the IMF. Classically a single power-law IMF proposed by Salpeter (1955) was used, but recently more realistic a IMF with a flattened low-mass end is becoming popular For a review, see Kroupa (2002). Two different conventions are used to define the IMF. Here we define the IMF through the normalization as

$$\int m\Phi(m)\mathrm{d}m = 1 \tag{1}$$

where m is the stellar mass and $\Phi(m)$ stands for the IMF.

2.2.2 Basic equations

We start from a set of basic equations to describe the chemical evolution.

$$\begin{aligned} M_{\mathrm{tot}} &= M_{\mathrm{gas}} + M_* & (2)\\ \frac{\mathrm{d}M}{\mathrm{d}t} &= f - e & (3)\\ \frac{\mathrm{d}M_*}{\mathrm{d}t} &= \Psi - E & (4)\\ \frac{\mathrm{d}M_{\mathrm{gas}}}{\mathrm{d}t} &= -\Psi + E + f - e\ , & (5)\end{aligned}$$

where M_{tot} is the total baryon mass of a galaxy (mass of dark matter is not included), M_{gas} is the gas mass, M_* is the stellar mass, f is the gas infall rate from intergalactic space, e is the gas outflow rate from the galaxy, Ψ is the SFR of the galaxy at time t, and E is the gas mass ejection rate of all stars.

Total gas ejection rate of stars is expressed as

$$E(t) = \int_{m_t}^{\infty} (m - w_m)\,\Psi\,(t - \tau_m)\,\Phi(m)\mathrm{d}m \tag{6}$$

where m_t: turnoff mass at time t, i.e,. the lowest mass of stars dying at time t, $m - w_m$: ejected mass, $\Psi(t-\tau_m)\Phi(m)$: birth rate at $t - \tau_m$, i.e., death rate at time t, τ_m: main sequence lifetime for a star with a mass m, and w_m: remnant mass:

$$w_m = \begin{cases} 0.11m + 0.45M_\odot & (m < 0.68\ M_\odot) \\ \\ 1.5\ M_\odot & (m > 0.68\ M_\odot) \end{cases} \quad . \tag{7}$$

Evolution of the metal abundance Z is written as

$$\frac{\mathrm{d}\,(ZM_{\mathrm{gas}})}{\mathrm{d}t} = -Z\Psi + E_Z + Z_f f - Ze\ , \tag{8}$$

where E_Z: ejection rate of metal(s) from stars (main sequence stars, Wolf-Rayet stars, SNe, etc.), Z_f: infalling metals per time, and ZM_{gas}; mass of metal(s) in the gas. The ejection rate of metals E_Z reads

$$E_Z(t) = \int_{m_t}^{\infty} \left[(m - w_m)\,Z(t-\tau_m) + mp_{Z_m}\right]\Psi\,(t-\tau_m)\,\Phi(m)\mathrm{d}m \tag{9}$$

where $(m - w_m)Z(t - \tau_m)$: mass of metal(s) which was locked in a star of mass m at time $t - \tau_m$ and is now ejected at time t, mp_{Z_m}: new metal(s) produced by a star of mass m with originally formed from gas with metallicity Z. We note that we adopt an ansatz: the instantaneous mixing of the produced metal(s) with the ISM in eqs. (8) and (9).

Here we define two (rather obscure) quantities. The returned mass per mass of stars formed is expressed as

$$R \equiv \int_{m_l}^{\infty} (m - w_m)\, \Phi(m) \mathrm{d}m \; . \tag{10}$$

This is independent of the star formation rate, thus is only valid for a single generation of stars. The mass of produced metal(s) per remaining stellar mass (including the stellar remnant) is called *yield*. This is expressed as

$$y \equiv \frac{1}{1-R} \int_{m_l}^{\infty} mp_{Z_m} \Phi(m) \mathrm{d}m \; . \tag{11}$$

Recall eq. (1), $1-R$ is always positive. In the next step we will see how these quantities work in the chemical evolution equations.

Before going ahead, we explicitly introduce important approximations.

1. Massive stars die immediately after their birth and less massive stars live forever,
2. Produced elements are instantaneously mixed with the ISM, and are used.

The second one was already introduced above. The first one is valid if the SFR is almost constant over a timescale of 10^7 yr for lighter elements like O, C, N, Mg, etc. (of SN II origin), and of 10^{8-9} yr for heavier elements like Fe (of SN Ia origin).

If the instantaneous recycling applies, using R and y and assuming the IMF is constant with time (meaning $R = \text{const.}$) we obtain

$$E(t) = R\Psi(t) \tag{12}$$

$$E_Z(t) = RZ(t)\Psi(t) + (1-R)\, y(t)\Psi(t) \; . \tag{13}$$

Inserting eq. (13) into eq.(4), we have

$$\frac{\mathrm{d}ZM_{\mathrm{gas}}}{\mathrm{d}t} = -Z\Psi(t) + RZ(t)\Psi(t) + (1-R)\, y(t)\Psi(t) + Z_f f - Ze \; , \tag{14}$$

thus

$$\frac{\mathrm{d}\,(ZM_{\mathrm{gas}})}{\mathrm{d}t} = (1-R)\,[-Z + y(t)]\,\Psi(t) + Z_f f - Ze \; . \tag{15}$$

As for stellar mass, with eq. (12),

$$\frac{\mathrm{d}M_*}{\mathrm{d}t} = (1-R)\,\Psi(t) \tag{16}$$

and for gas mass,

$$\frac{\mathrm{d}M_{\mathrm{gas}}}{\mathrm{d}t} = -\,(1-R)\,\Psi(t) + f - e \; . \tag{17}$$

Then, we have

$$M_{\rm gas}\frac{{\rm d}Z}{{\rm d}t} = (1-R)\,y(t)\Psi(t) + (Z_f - Z)\,f\ . \tag{18}$$

These equations are the framework under the instantaneous recycling.

2.2.3 Closed-box model

To go further analytically, we introduce another simplifying assumption. Assume we have a closed system containing only gas with zero metallicity (not essential) and no stars. Since $f = e = 0$, $M_{\rm gas}(t=0) = M_{\rm tot}$, $M_*(t=0) = 0$,

$$\begin{aligned} \frac{{\rm d}M_*}{{\rm d}t} &= (1-R)\,\Psi(t)\ , \\ M_{\rm gas}\frac{{\rm d}Z}{{\rm d}t} &= (1-R)\,y(t)\Psi(t) \end{aligned} \tag{19}$$

which leads to:

$$\frac{1}{M_{\rm gas}}{\rm d}M_{\rm gas}{\rm d}Z = -\frac{1}{y}\ , \tag{20}$$

or equivalently:

$$\ln M_{\rm gas}\big|_{M}^{M_{\rm gas}(t)} = -\int_0^{Z(t)}\frac{{\rm d}Z}{y} \simeq -\frac{Z}{\bar{y}}\ . \tag{21}$$

The metallicity of the gas is

$$Z(t) = \bar{y}\ln\frac{M_{\rm gas}(t=0)}{M_{\rm gas}(t)}\ . \tag{22}$$

Here we should note that $Z(t)$ depends only on $M_{\rm gas}(t)/M_{\rm tot}$, thus not explicitly on t or SFR.

By a similar line of discussion, we can also obtain an important result for the stellar metallicity Z_*. After some arithmetic, we obtain:

$$Z_* M_* \simeq \bar{y}M_* - ZM_{\rm gas}\ . \tag{23}$$

This leads to an important result:

$$M_{\rm gas} \ll M_* \quad \Rightarrow \quad Z_* \simeq \bar{y} \tag{24}$$

i.e., the average stellar metallicity cannot exceed the average yield.

2.3 Evolutionary synthesis model of galaxy spectra

By making use of the chemical evolution result, we can construct a more practical type of model, the spectral synthesis model. If the model treats the evolutionary aspect, it is called the evolutionary synthesis of galaxy spectra. Stars have different spectra depending on their effective temperature (often referred to as spectral type) and metallicity. Let there be a population of stars which were born at the same moment and with a certain IMF and metallicity. Then, the total spectrum of this population at age t is expressed as an IMF-weighted sum of the spectra of each stellar mass (equivalently spectral type) with age t after their birth. This hypothetical population is called a single (or simple) stellar population (SSP), and plays a fundamental role in theoretical modelling of galaxy spectra. The SSPs vary as a function of age, metallicity, and the adopted IMF.

If we convolve the SSP spectra with arbitrary star formation history $\Psi(t)$, we obtain the observed galaxy spectra at age t,

$$F_\lambda(t) = \int_0^t \Psi(t-\tau) f_\lambda(\tau) \mathrm{d}\tau \ , \tag{25}$$

where $f_\lambda(t)$: an SSP of age t. A commonly used SF history is the exponentially decaying SFR:

$$\Psi(t) = \frac{1}{\tau} \exp\left(\frac{t}{\tau}\right) \ . \tag{26}$$

This simple model actually reproduces the observed spectra of various galaxies from ellipticals to intensively SF dwarf galaxies, if we take into account the emission lines from gas and dust extinction. Observationally, spiral galaxies have a slowly continuous decaying SF history (SFH) after a short burst. From earlier to later types, the timescale of SF becomes longer. This trend also holds including ellipticals and irregulars: Sandage law (Sandage, 1986). Hence, empirically the exponentially decaying SFR is a good approximation.

However, the SFH plays a central role in controlling the chemical evolution. Though in the simplest model the SFH did not appear explicitly, generally we have to put in a certain physical model to specify the SFH. In reality, this part is complicated and still poorly understood; we often adopt the following empirical law, known as the Schmidt law (Schmidt, 1959) or its variant:

$$\mathrm{SFR} \propto \rho^n \ , \quad n = 1\text{–}2 \ . \tag{27}$$

On the observational side, it is expressed as a function of *surface* gas density, while in the chemical evolution modelling a different form is often used. In any case, applying the Schmidt law to the chemical evolution we can reproduce an exponential-type SFH in a self-consistent manner (not by hand).

In modern galactic astrophysics, these models are very extensively used and applied to various problems. However, it is strongly recommended to use the spectral synthesis models after understanding the above theoretical background, rather than merely as a black box.

2.4 Chemical evolution with dust

The produced heavy elements are not always in a gaseous state: indeed, more than a half of the heavy elements form tiny solid grains, called dust. Dust grains are usually suspended within the ISM in galaxies. Star formation activity is closely related to heavy element production. This, therefore, means that the star formation is also connected to the production of dust. Hence, intense star formation always accompanies active dust production. On the other hand, dust grains also accelerate the efficiency of star formation. The interplay between the star formation and dust is very complex and nonlinear.

The existence of dust affects the spectral energy distribution (SED) of galaxies. As mentioned above, when we apply the spectral synthesis model to fit the galaxy SEDs, dust extinction always introduces a large uncertainty, especially at high z. Observationally, the SFR of galaxies is measured through ultraviolet (UV) radiation or related observables (recombination lines, forbidden lines from HII regions, non-ionizing UV etc.) However, short-wavelength photons like UV are easily scattered and absorbed by dust (referred to as "extinction"), and re-emitted in the far-infrared (FIR). Obviously the amount of UV light obscured by dust is only measured at FIR wavelengths. Otherwise this leads us to a serious underestimation of the SFR.

Therefore, the chemical evolution including dust is very important in modern discussions of galaxy evolution. However, since it is beyond the level of this lecture note, interested readers are recommended to refer to Asano et al. (2013a,b, 2014).

2.5 Galactic wind model for dwarf galaxies

Dwarf galaxies are known to have a large variety of physical properties. Also, their global properties are distinctly different from those of giant galaxies (e.g., Mateo, 1998). Especially we should note that dwarfs have high mass-to-light ratios. This is interpreted such that their SF efficiency is low compared to giant galaxies. This peculiar aspect of dwarfs is often discussed in relation to their shallow gravitational potential, which makes them susceptible to both internal and external disturbance.

Dekel & Silk (1986) proposed a refined idea of how to explain the low luminosity, low surface brightness, low metallicity, and lack of cold ISM in dwarf elliptical/spheroidal galaxies from the point of view of SN feedback. We outline their model – it is an interesting application of the chemical evolution theory in combination with a dynamical phenomenon. They considered the galactic wind induced by the SNe in dwarf galaxies. Once the first generation of stars is born, subsequent SN explosions cause a galaxy-wide wind phenomenon: galactic wind. If the mass of a galaxy is small, gas is accelerated easily and its velocity exceeds the escape velocity of the galaxy (e.g. Mac Low & Ferrara, 1999).

The basic idea is simple: galactic wind is induced by SNe. Then, the energy input into the ISM by SNe should be approximately proportional to the number of SNe which have ever occurred in a galaxy. The number of SNe is roughly proportional to the mass of stars in a galaxy. The binding energy of baryons in the dark halo potential well is expressed as:

$$E_{\rm bind} \propto bM\frac{M}{R} \propto \frac{M_*}{s}\frac{M}{R} \; , \tag{28}$$

where s: fraction of stellar mass to total baryon mass, M: halo mass, M_*: stellar mass, b: mass ratio of baryon to M, and R: dark halo size. We assume that the mass density is the same all over the galaxy, then we have

$$E_{\rm bind} \propto M_*^{5/3} s^{-5/3} \; . \tag{29}$$

The critical value such that the energy of SNe exceeds the binding energy of the ISM and the gas is expelled out of the galaxy is

$$E_{\mathrm{SN}} \propto M_* \propto M_*^{5/3} s^{-5/3} \,, \tag{30}$$

hence

$$s \propto M_*^{2/5} \,. \tag{31}$$

The surface brightness I is expresses as

$$I \propto \frac{M_*}{R^2} \propto M_*^{3/5} \,. \tag{32}$$

This implies that if M_* is small, the surface brightness and luminosity become small. Galaxies with higher gas mass fraction has lower metallicity, namely low-luminosity galaxies are metal-poor. When the stellar mass fraction s is small, from eq. (22), we can approximate the metallicity as

$$Z = y \ln \left(\frac{bM}{M_{\mathrm{gas}}} \right) = y \ln \left(\frac{M_{\mathrm{gas}} + M_*}{M_{\mathrm{gas}}} \right) \propto \frac{M_*}{M_{\mathrm{gas}}} \simeq s \,. \tag{33}$$

thus

$$Z \propto s \propto M_*^{2/5} \propto L^{2/5} \,. \tag{34}$$

Indeed, this reproduces the observed relation well.

3 Structure Formation in the Universe

Now we move on to the formation of cosmic structures and dark halos of galaxies. The treatment is divided into two regimes: linear and nonlinear evolution of structures. In the early Universe, the distribution of matter and radiation was almost homogeneous, with tiny fluctuations. These fluctuations might have been input by the supposed inflation period. Inflation models predict that the initial fluctuations were (almost) random Gaussian.

Since (as far as we know) there is no repulsion in gravitational interaction, such tiny fluctuations can gather surrounding matter by gravity and grow monotonically. In the early phase, the amplitude of such fluctuations is still very small compared to the average density. In this case, second- and higher-order terms related to the fluctuation can be ignored in the basic equations, i.e. the equations can be linearized. Under this condition, fluctuations with different spatial scales can be treated independently since there is no interaction between fluctuations with different scales, and the Fourier transformation plays a central role in the formulation. In contrast, when the amplitude of the density fluctuations approaches the average, we cannot apply this approximation any more. In this case, special treatment is required. This is the nonlinear regime. We briefly introduce these theoretical formulations in the following.

3.1 Linear theory of density fluctuations

3.1.1 Basics from cosmology

When we discuss the evolution of the Universe, the theory of general relativity is required. However, as for the structure formation, we need only a basic part of general relativity. The line element of spacetime is described by a four-by-four symmetric tensor (the metric tensor).

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a^2(t)\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right] , \tag{35}$$

where k is the (normalized) Gaussian curvature, and $a(t)$ is the scale factor. We set $a_0 \equiv a(t_0) = 1$. Einstein's field equation relates the spacetime curvature and the energy–momentum density. We start from this equation. The energy–momentum tensor of the homogeneous and isotropic Universe should take the form of a perfect fluid (e.g., Weinberg, 1972):

$$T^\mu_\nu = \mathrm{diag}\left(-\rho c^2, p, p, p\right) \tag{36}$$

where ρc^2 is the energy density and p is the pressure. The Einstein field equations read

$$R^\mu_\nu - \frac{1}{2}Rg^\mu_\nu + \Lambda g^\mu_\nu = \frac{8\pi G}{3}T^\mu_\nu \tag{37}$$

where R^μ_ν is the Ricci tensor, $R \equiv R^\alpha_\alpha$ is the Ricci scalar, and Λ is the cosmological constant. Substituting eq. (36) into eq. (37) and performing some algebra, we obtain the Friedmann equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} , \tag{38}$$

$$\frac{\ddot{a}}{a} = \frac{4\pi G\rho}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} . \tag{39}$$

Equation (38) comes from the 00 component of the Einstein equations, and eq. (39) is obtained from the trace.

3.1.2 From physical coordinates to comoving coordinates

On a scale much smaller than the horizon scale, or when the motion of matter is much slower than the speed of light, we can safely approximate that matter behaves like a classical fluid. This is called "the Newtonian fluid approximation". Under this assumption, we have

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0 , \tag{40}$$

$$\frac{\partial\mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = -\frac{\nabla p}{\rho} - \nabla\phi , \tag{41}$$

$$\nabla^2\phi = 4\pi G\rho , \tag{42}$$

where ρ is the matter density, p and $\mathbf{v}$ are the pressure and velocity of matter, and ϕ is the gravitational potential produced by matter. Equations (40)–(41) are written in *physical* coordinates.

We introduce comoving coordinates so that we can treat structure formation in the expanding Universe:

$$\mathbf{r} = a(t)\mathbf{x} , \tag{43}$$

where $a(t)$ is the scale factor describing the cosmic expansion (e.g., Weinberg, 1972; Peebles, 1993). It directly leads

$$\begin{aligned}\frac{\partial \mathbf{r}}{\partial t} &= \dot{\mathbf{r}} = \dot{a}\mathbf{x} + a\dot{\mathbf{x}} \equiv H(t)a\mathbf{x} + a\dot{\mathbf{x}} \\ &= H(t)\mathbf{r} + \mathbf{u} \qquad (44)\end{aligned}$$

where $\mathbf{u}$ is a peculiar velocity. Further, we obtain for the differential operators as follows:

$$\left(\frac{\partial}{\partial t}\right)_r = \left(\frac{\partial}{\partial t}\right)_x - \frac{\dot{a}}{a}\mathbf{x} \cdot \nabla_x \,, \qquad (45)$$

$$\nabla_r = \frac{1}{a}\nabla_x \,. \qquad (46)$$

Hereafter, we drop the subscript x.

By changing the coordinates from physical to comoving, we have the continuity equation [eq. (40)]

$$\frac{\partial \rho}{\partial t} + 3H\rho + \frac{1}{a}\nabla \cdot (\rho \mathbf{v}) = 0 \,, \qquad (47)$$

and Euler equation [eq. (41)]

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}\left(\mathbf{u} \cdot \nabla\right)\mathbf{u} = -\frac{\nabla p}{a\rho} - \nabla\phi - \ddot{a}\mathbf{x} \,. \qquad (48)$$

By defining a new potential

$$\Phi \equiv \nabla\phi + \frac{a\ddot{a}\mathbf{x}}{2} \,, \qquad (49)$$

we have

$$\frac{\partial \mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}\left(\mathbf{u} \cdot \nabla\right)\mathbf{u} = -\frac{\nabla p}{a\rho} - \frac{1}{a}\nabla\Phi \,. \qquad (50)$$

The Poisson equation [eq. (42)] leads

$$\nabla^2\phi = 4\pi G a^2 \rho \,.$$

From the Friedmann equation (e.g., Peebles, 1993), we have

$$3a\ddot{a} = (\nabla\mathbf{x}) = \left(\nabla^2 \frac{x^2}{2}\right) = -4\pi G\bar{\rho} \,.$$

Then, we obtain

$$\nabla^2\Phi = 4\pi G a^2 \left(\rho - \bar{\rho}\right) \,, \qquad (51)$$

with a solution

$$\Phi(\mathbf{x}) = -Ga^2 \int \mathrm{d}^3\mathbf{x}' \frac{\rho - \bar{\rho}}{|\mathbf{x}' - \mathbf{x}|} \,. \qquad (52)$$

Consider a small fluctuation from the background universe:

$$\begin{aligned}\rho_{\rm b} &= \bar{\rho}(t)\ ,\\ \mathbf{u} &= \mathbf{0}, \nabla\Phi = \mathbf{0}, \nabla p = \mathbf{0}\ .\end{aligned}$$

We introduce fluctuations from the homogeneous background as

$$\delta\left(\mathbf{x}, t\right) \equiv \frac{\rho - \bar{\rho}}{\bar{\rho}}\ , \tag{53}$$

$$\delta p\left(\mathbf{x}, t\right) \equiv p\left(\mathbf{x}, t\right) - \bar{p}(t)\ . \tag{54}$$

Then, eq. (51) becomes

$$\nabla^2\Phi = 4\pi G a^2 \bar{\rho}\delta\ . \tag{55}$$

The continuity equation [eq. (47)] leads

$$\frac{\partial}{\partial t}\left[\bar{\rho}(1+\delta)\right] + 3H\bar{\rho}(1+\delta) + \frac{\bar{\rho}}{a}\nabla\cdot\left[(1+\delta)\mathbf{u}\right] = 0\ . \tag{56}$$

Here we should note that, from the mass conservation $\partial(a^3\bar{\rho})/\partial t = 0$,

$$\frac{\partial\rho}{\partial t} + 3H\rho = 0\ . \tag{57}$$

From eqs. (56) and (57),

$$\frac{\partial}{\partial t}\left(\bar{\rho}\delta\right) + 3H\bar{\rho}\delta + \frac{\bar{\rho}}{a}\nabla\cdot\left[(1+\delta)\mathbf{u}\right] = \bar{\rho}\frac{\partial\delta}{\partial t} + \frac{\bar{\rho}}{a}\nabla\cdot\left[(1+\delta)\mathbf{u}\right] = 0\ . \tag{58}$$

Thus, the continuity equation [eq. (47)] becomes

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\left[(1+\delta)\mathbf{u}\right] = 0\ . \tag{59}$$

and the Euler equation [eq. (50)]

$$\frac{\partial\mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}\left(\mathbf{u}\cdot\nabla\right)\mathbf{u} = -\frac{\nabla\delta p}{a\rho(1+\delta)} - \frac{1}{a}\nabla\Phi\ . \tag{60}$$

Equations (55), (59), and (60) are the starting point for deriving the solutions which describe the linear growth of fluctuations.

3.1.3 Jeans equation

Here, we neglect terms including multiplications of $\delta, \delta p$, and $\mathbf{u}$ (linearization). Then the continuity equation and the Euler equation now read

$$\frac{\partial\delta}{\partial t} + \frac{1}{a}\nabla\cdot\mathbf{u} = 0\ , \tag{61}$$

$$\frac{\partial\mathbf{u}}{\partial t} + H\mathbf{u} + \frac{1}{a}\nabla\Phi + \frac{\nabla\delta p}{a\rho} = 0\ . \tag{62}$$

Manipulating eqs. (61) and (62) gives

$$\frac{\partial^2 \delta}{\partial t^2} + 2\frac{\dot{a}}{a^2}(\nabla \cdot \mathbf{u}) - 4\pi G\bar{\rho}\delta - \frac{\nabla^2 \delta p}{a^2 \rho} = \frac{\partial^2 \delta}{\partial t^2} + 2H\frac{\partial \delta}{\partial t} - \left[4\pi G\bar{\rho}\delta + \frac{\nabla^2 \delta p}{a^2 \rho}\right] = 0 \ . \quad (63)$$

Consider the sound velocity of the fluid

$$c_s \equiv \sqrt{\frac{\partial p}{\partial \rho}}\Bigg|_S \quad (64)$$

where S is the entropy. Suppose that the entropy fluctuation is small, we have

$$\delta p = c_s \bar{\rho}\delta \ . \quad (65)$$

Since the basic equations are linear, we can deal with Fourier components for each wavenumber k. Consider

$$\delta(\mathbf{x}) = \sum_k \delta_k e^{i\mathbf{k}\cdot\mathbf{x}} \ . \quad (66)$$

Then,

$$\frac{\partial^2 \delta_k}{\partial t^2} + 2H\frac{\partial \delta_k}{\partial t} - \left[4\pi G\bar{\rho} - \frac{c_s^2 k^2}{a^2}\right]\delta_k = 0 \ . \quad (67)$$

This is the Jeans equation. The third term in the square brackets [ ] controls the evolution of fluctuation.

1. [ ] < 0 Fluctuations oscillate and decay. This happens when c_s is large. This also happens when k is large (small scale) because the fluctuation does not contain enough mass to contract gravitationally.

2. [ ] > 0 Fluctuations grow.

Here we define k_J so that

$$\frac{c_s^2 k_\mathrm{J}^2}{a^2} = 4\pi G\bar{\rho} \ , \quad (68)$$

then we can also define the Jeans length λ_J:

$$\lambda_\mathrm{J} = \frac{2\pi a}{k_\mathrm{J}} = c_s\sqrt{\frac{\pi}{G\bar{\rho}}} \ . \quad (69)$$

Fluctuations smaller than λ_J decay, while ones larger than λ_J can grow by gravitational instability. The corresponding mass

$$M_\mathrm{J} \equiv \frac{4\pi}{3}\bar{\rho}\left(\frac{\lambda_\mathrm{J}}{2}\right)^3 \quad (70)$$

is referred to as the Jeans mass.

3.2 Nonlinear theory for the formation of a collapsed object

After the fluctuation grows and the amplitude becomes of the same order as the average density, we cannot treat it using the linear theory. Thus we will now discuss how to consider the nonlinear dynamics of collapsed objects.

3.2.1 Spherical collapse in the expanding Universe

In comoving coordinates, a sphere centered on a local overdensity shrinks in time; the Hubble expansion is getting retarded by the overdensity. At some point, the sphere's expansion stops (turn-around), and the sphere starts to collapse. Consider the evolution of a spherical overdensity region as a simple model of the nonlinear evolution of fluctuation (Tomita, 1969; Gunn & Gott, 1972). Let $R(t)$ the radius of a shell, and M the mass contained inside the shell. Then, the equation of motion of the shell is

$$\frac{\mathrm{d}^2 R}{\mathrm{d}t^2} = -\frac{GM}{R^2} \tag{71}$$

which gives

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \frac{2GM}{R} + 2E\ . \tag{72}$$

As we know in classical mechanics, if the "binding energy" $E < 0$, the shell motion is bounded. In this case the solution is given parametrically as

$$R = C^2\left(1 - \cos\theta\right)\ , \quad t = \frac{C^3}{\sqrt{GM}}\left(1 - \cos\theta\right)\ , \tag{73}$$

where C is an integration constant, corresponding to the size of the shell. This curve is called a "cycloid". If the energy $E > 0$, the motion is unbounded. This corresponds to the dynamics of voids.

Densities of overdense and average regions are

$$\rho = M\left(\frac{4\pi R^3}{3}\right)^{-1}\ , \quad \bar{\rho} = \frac{1}{6\pi G t}\ , \tag{74}$$

respectively. Then we obtain the density contrast

$$\delta(t) = \frac{9GMt^2}{2R^3} - 1 = \frac{9}{2}\frac{\left(\theta - \sin\theta\right)^2}{\left(1 - \cos\theta\right)^3}\ . \tag{75}$$

We can characterize its motion as follows:

1. $\theta = \pi$ The turn-around point from expansion to contraction.

$$R_{\mathrm{turn}} = 2C^2, \quad t_{\mathrm{turn}} = \frac{\pi C^3}{\sqrt{GM}}\ . \tag{76}$$

2. $\theta = 2\pi$ Collapse ($R = 0$).

$$t_{\mathrm{coll}} = \frac{2\pi C^3}{\sqrt{GM}}\ . \tag{77}$$

In reality, $R = 0$ is not established, but the overdense region becomes an object with $R_{\rm vir}$, via some mechanism like the violent relaxation. Suppose a mass M; then, from the conservation of energy, we have

$$\frac{GM^2}{R_{\rm turn}} = \frac{1}{2}\frac{GM^2}{R_{\rm coll}} \tag{78}$$

hence

$$R_{\rm coll} = \frac{R_{\rm turn}}{2} \ . \tag{79}$$

In this case, the density contrast of a collapsed object is given by

$$\delta(t) = \frac{M}{\left(\frac{4\pi R_{\rm vir}{}^2}{3}\right)\bar{\rho}(t_{\rm coll})} - 1 = 18\pi^2 - 1 \simeq 177 \ . \tag{80}$$

3.2.2 Connection to the linear theory

At early phases, the growth should agree with the linear growth. Expanding with θ, we have

$$\delta(t) = \frac{3}{20}\theta^2 + \mathcal{O}(\theta^4) \ , \quad t = \frac{C^3}{6\sqrt{GM}}\theta^3 + \mathcal{O}(\theta^5) \ ,$$

which give $\delta \propto t^{2/3}$. If we define this overdensity extrapolated from the linear theory, $\delta_{\rm L}$,

$$\delta_{\rm L}(t) = \frac{3}{20}\left(\frac{6\sqrt{GM}}{C^3}t\right)^3 \ . \tag{81}$$

Since both the nonlinear d of spherical collapse and linear $\delta_{\rm L}$ are monotonic functions of t, we can estimate the value of d from $\delta_{\rm L}$ if we have a relation between δ and $\delta_{\rm L}$, as

$$\delta_{\rm L}(t_{\rm coll}) = \frac{3(12\pi)^{2/3}}{20} \simeq 1.69 \ . \tag{82}$$

Namely, if we find a region with $\delta = 1.69$, it is actually a collapsed object.

The above discussion was based on the Einstein–de Sitter Universe, but for the flat Λ-dominated Universe, the correction is very small.

3.2.3 Press–Schechter formalism

Press & Schechter (1974) proposed an idea to connect the linear growth solution of density fluctuations and the extrapolation to the nonlinear regime through a spherical collapse model. This gives an analytic model of dark halo formation.

Let the number density of objects whose mass is between M and $M + \mathrm{d}M$ be $n(M)\mathrm{d}M$. Then, this $n(M)$ is called the mass function. The Press–Schechter (PS) formalism gives an analytical solution of $n(M)$.

$$M = \frac{4\pi R^3}{3}\bar{\rho} \ . \tag{83}$$

The smoothed (averaged) overdensity in a sphere whose radius R, corresponding to a mass M, is called a fluctuation δ_M of mass scale M.

As we mentioned above, the original matter fluctuation δ is Gaussian. In general, if a stochastic field is Gaussian, the Gaussianity is preserved by smoothing. Then the smoothed fluctuation δ_M is also Gaussian.

$$P(\delta_M)\mathrm{d}\delta_M = \frac{1}{\sqrt{2\pi\sigma(M)^2}} e^{\frac{\delta_M^2}{2\sigma(M)^2}} \mathrm{d}\delta_M \ , \tag{84}$$

where $\sigma(M)^2$ represents the variance of δ_M. Naturally the variance $\sigma(M)^2$ becomes smaller than the original σ^2 by smoothing. At a certain point, if the linear (extrapolated) δ_M exceeds the threshold value δ_c, a collapsed object with mass M is formed. We set $\delta_c = \delta_{\rm coll} = 1.69$ as the spherical collapse model. We note that recently a number of theoretical studies adopt a more complicated form for δ_c, reflecting more realistic physical conditions.

The spatial fraction of the regions with $\delta > \delta_c$ is

$$P(\delta_M > \delta_c)\mathrm{d}\delta_M = \frac{1}{\sqrt{2\pi\sigma(M)^2}} \int_{\delta_c}^{\infty} e^{-\frac{\delta_M^2}{2\sigma(M)^2}} \mathrm{d}\delta_M = \frac{1}{\sqrt{2\pi}} \int_{\frac{\delta_c}{\sigma(M)^2}}^{\infty} e^{\frac{x^2}{2}} \mathrm{d}x \ . \tag{85}$$

The amount of matter involved in an object with mass larger than M per unit volume is $\bar{\rho}P(\delta_c)(M)$. Hence

$$\bar{\rho}P(>\delta_c)(M+\mathrm{d}M) - \bar{\rho}P(>\delta_c)(M) = \frac{\mathrm{d}P(>\delta_c)}{\mathrm{d}M}\mathrm{d}M = n(M)M\mathrm{d}M \ . \tag{86}$$

The discussion above ignored the possibility that a once-collapsed object would be involved in a larger object (the cloud-in-cloud problem). And the region with $\delta < 0$ will never be involved in any collapsed object (i.e., $P(>\delta_c) \to 1/2$ as $\sigma(M) \to \infty$). Then, simply we multiply a factor 2 to avoid the problem.

$$n(M)M\mathrm{d}M = 2\bar{\rho}\left|\frac{\mathrm{d}P(>\delta_c)}{\mathrm{d}M}\right|_M \mathrm{d}M \ . \tag{87}$$

Hence

$$n(M) = \sqrt{\frac{2}{\pi}} \frac{\bar{\rho}}{M^2} \left|\frac{\mathrm{d}\ln\sigma(M)}{\mathrm{d}\ln M}\right| \frac{\delta_c}{\sigma(M)} e^{-\frac{\delta_c^2}{2\sigma(M)^2}} \ . \tag{88}$$

Since the power spectrum is approximated partially by a power-law like $P(k) \propto k^n \Leftrightarrow \sigma \propto M^{-\alpha}$ $(\alpha = (n+3)/6)$, we obtain the final form

$$n(M) = \frac{2\alpha}{\sqrt{\pi}} \frac{\bar{\rho}}{M_*^2} \left(\frac{M}{M_*}\right)^{\alpha-2} e^{-\left(\frac{M}{M_*}\right)^{2\alpha}} \ . \tag{89}$$

The Schechter function (Schechter, 1976), often used as an approximate form of the galaxy luminosity function, was originally inspired by the PS mass function [eq. (89)].

However, as we have already noticed, the original formulation by Press & Schechter contains some insufficient assumptions. The current mainstream of the mass function formulation is to derive the PS mass function by modelling the merging of galaxy halos (extended PS formalism: e.g., Lacey & Cole, 1993).

4 Further topics

In this lecture note, we mainly focused on the chemical evolution and structure formation. Dark matter forms gravitational potential, and, after decoupling, baryons fall into the potential. Then, in the dark matter haloes formed by a nonlinear collapse, the first stars and galaxies form. The key factor to form stars from gas is the cooling mechanism. In the very early Universe, cooling is very difficult because of the deficit of heavy elements. They are provided from the first supernovae in the Universe. Then galaxies form as a consequence of starbursts, and chemical evolution starts. The first stars and galaxies (and other sources if they exist) must have reionized the Universe, since the ionization rate is almost 100 % in the present day Universe. Galaxies merge and grow with merging of dark halos, as well as the internal SF in galaxies. How much of the stellar mass was contributed from SF and how much was from merging is still a matter of debate. The spatial distribution of galaxies is different from that of dark matter: this is the bias of galaxy distribution (e.g., Kaiser, 1984). This is partially explained as the result of peak bias, i.e., dark matter can collapse more easily at the top of a large fluctuation. However, even the halo and galaxy distributions do not agree, and we need further studies from the baryonic side. Galaxies are also affected by their environment through various mechanisms. The SF and AGN activities put energy into the ISM and regulate the subsequent SF, producing a feedback. In current studies of galaxy evolution, the effect of feedback is particularly important, since it is related to the evolution of dwarf galaxies, halting of too much SF at the central galaxy in clusters, etc. Many problems remain unsolved in the formation and evolution of galaxies. We hope this note will help in the understanding of the first step in the exciting studies in this field.

Acknowledgements. We have been supported by the Grant-in-Aid for the Scientific Research Fund (24111707) commissioned by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

References

Asano, R. S., Takeuchi, T. T., Hirashita, H., Inoue, A. K., *Dust formation history of galaxies: A critical role of metallicity for the dust mass growth by accreting materials in the interstellar medium*, *Earth, Planets, and Space* **65**, 213 (2013a), `1206.0817`

Asano, R. S., Takeuchi, T. T., Hirashita, H., Nozawa, T., *What determines the grain size distribution in galaxies?*, MNRAS **432**, 637 (2013b), `1303.5528`

Asano, R. S., Takeuchi, T. T., Hirashita, H., Nozawa, T., *Evolution of extinction curves in galaxies*, MNRAS **440**, 134 (2014), `1401.7121`

Dekel, A., Silk, J., *The origin of dwarf galaxies, cold dark matter, and biased galaxy formation*, ApJ **303**, 39 (1986)

Gunn, J. E., Gott, J. R., III, *On the Infall of Matter Into Clusters of Galaxies and Some Effects on Their Evolution*, ApJ **176**, 1 (1972)

Kaiser, N., *On the spatial correlations of Abell clusters*, ApJ **284**, L9 (1984)

Kroupa, P., *The Initial Mass Function of Stars: Evidence for Uniformity in Variable Systems*, *Science* **295**, 82 (2002), `astro-ph/0201098`

Lacey, C., Cole, S., *Merger rates in hierarchical models of galaxy formation*, MNRAS **262**, 627 (1993)

Mac Low, M.-M., Ferrara, A., *Starburst-driven Mass Loss from Dwarf Galaxies: Efficiency and Metal Ejection*, ApJ **513**, 142 (1999), `astro-ph/9801237`

Magorrian, J., et al., *The Demography of Massive Dark Objects in Galaxy Centers*, AJ **115**, 2285 (1998), `astro-ph/9708072`

Mateo, M. L., *Dwarf Galaxies of the Local Group*, ARA&A **36**, 435 (1998), `astro-ph/9810070`

Peebles, P. J. E., Principles of Physical Cosmology (1993)

Press, W. H., Schechter, P., *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, ApJ **187**, 425 (1974)

Salpeter, E. E., *The Luminosity Function and Stellar Evolution.*, ApJ **121**, 161 (1955)

Sandage, A., *Star formation rates, galaxy morphology, and the Hubble sequence*, A&A **161**, 89 (1986)

Schechter, P., *An analytic expression for the luminosity function for galaxies.*, ApJ **203**, 297 (1976)

Schmidt, M., *The Rate of Star Formation.*, ApJ **129**, 243 (1959)

Tinsley, B. M., *Evolution of the Stars and Gas in Galaxies*, Fund. Cosmic Phys. **5**, 287 (1980)

Tomita, K., *Formation of Gravitationally Bound Primordial Gas Clouds*, *Progress of Theoretical Physics* **42**, 9 (1969)

Weinberg, S., Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity (1972)