Multiplicity of timescales: Insight into galaxy formation

Peter Laursen[1,2]

[1] Cosmic Dawn Center (DAWN). e-mail: `pela@nbi.ku.dk`.
[2] Niels Bohr Institute, University of Copenhagen, Jagtvej 128, 2200 Copenhagen N, Denmark.

ABSTRACT

Timescales in astronomy comprise the largest range of any scientific discipline. When constructing physical models, this circumstance is both a blessing and a curse. Galaxy evolution occurs on typical timescales of hundreds of millions of years, but is affected by atomic processes on sub-second timescales. On the other hand, the vast dynamic range implies that we can often make meaningful predictions by simply comparing characteristic timescales of the physical processes involved. This brief review, aimed primarily at non-astronomer scientists, attempts to highlight a few occasions in the context of galaxy formation and evolution in which comparing timescales can shed light on astrophysical phenomena, as well as some of the challenges that may be encountered.

Key words. galaxies: kinematics and dynamics — galaxies: formation — cosmology: theory.

1. Introduction

No scientific field deals with a longer span of timescales than astrophysics. From explosion mechanisms in dying stars and oscillations of neutron stars, to cosmic structure formation and the age of the Universe itself, more than twenty orders of magnitude prevail. Counting Big Bang physics as "astrophysics" doubles or even triples this amount[1]. However, while this enormous dynamic range can be challenging to model under one umbrella, it may also be an advantage: Because the evolution of a system in general happens at a rate determined by the longer timescale, modified by shorter timescales, many characteristic features of various astrophysical phenomena can be understood simply by comparing relevant physical timescales.

Whereas the short end of this range of timescales is in many cases observable, an inherent problem in astronomy is that many processes occur on timescales much longer than a human life, or even a human civilization that would be able to pass knowledge down through generations. How, then, may we say anything about how stars form, how galaxies evolve, or which phases the Universe has gone through?

Several approaches exist to answering this question: Observationally, while we cannot wait to witness the evolution of any given galaxy — which to us puny humans seems frozen in time — we are lucky to live in a universe where the speed of light is finite, but fast. As we peer deeper into the cosmos, we probe earlier and earlier epochs. Observing large samples of galaxies throughout cosmic history then allows us to study their properties in a statistical sense — numerous defining properties such as their star formation rates, the build-up of heavy elements and dust, and changes in their morphology.

On the theoretical side, as in all other fields of science we make use of *models*. Models that predict the evolution of these various physical properties, while being (largely) consistent with already known physical properties. Models can be analytical, numerical, or a mixture thereof; the so-called semi-analytical models. In this chiefly theoretical review we will have a look at how

timescales can be used to make meaningful predictions about the most beautiful structures in the Universe; the galaxies.

Ever since the revelation that the Milky Way does not comprise the whole Universe, but that we are surrounded by countless similar "island universes" (Shapley & Curtis 1921; Hubble 1926), we have strived to comprehend the origin of these magnificent entities. Which physical processes may lead to their formation and govern their evolution, while at the same time resulting in the astounding observed heterogeneity in their properties?

Progress in this captivating field is not only an exquisite interplay between the increasing power of telescopes and detectors, improved observational techniques, and an increasingly better theoretical understanding of cosmology and the Universe in general; advances are also stimulated by breakthroughs in such diverse fields as particle physics, chemistry, and computer science. No man is an island (Jovi 1990), and the same can be said about fields of science — interdisciplinarity is essential to progress!

A brief outline of the physical processes leading to a galaxy can be summarized as follows:

2. Galaxy formation and evolution

In the early, expanding Universe, sufficiently large overdensities are able to withstand and detach themselves from this expansion, turn over, and collapse. With more than five times as much dark matter as baryonic matter, the dynamics are initially dominated by the former. Eventually the cloud will *virialize* — astronomers' term for reaching a dynamical equilibrium — and come to a halt. Gas, which unlike the collisionless dark matter is able to cool and fragment further, condenses in the center of more extended dark matter halos.

In the very center, supermassive black holes form which accrete mass, ejecting excess energy as so-called *active galactic nuclei* (AGN). Meanwhile, dying stars inject not only energy but also heavy elements (in astronomy, everything heavier than helium is collectively called "metals"). With time, the interstellar

---

[1] The age of the Universe is $\sim 10^{61}$ Planck times!

Fig. 1. Relationship between observed redshift of light emitted from a distant object, and various properties of the Universe. Because the relative expansion was larger in the past when the Universe was smaller, redshift increased more at that time asymptotically approaching infinity for light emitted at the time of the Big Bang, 13.8 Gyr ago. The red dot marks the current record for the most distant known galaxy, GN-z11, with $z = 11.09$ (Oesch et al. 2016).

Cosmological redshift

Arguably, the most essential concept in astronomy is the *redshift z* of light. Well-known to other physicist as a result of Doppler shift caused by the relative motion of an emitter and an observer, in extragalactic astronomy there is second effect, dominating on all but the most local scales, caused by the cosmological expansion of the Universe.

As light travels through expanding space, its wavelength increases by a factor $1 + z$, equal to ratio of the size of the Universe at the time of observation and emission, respectively. Because the Universe has always been expanding, the observed redshift becomes a measure of both the age of the Universe at emission, the current distance, the "lookback time", the observed volume, and several other useful aspects. We therefore simply use the term redshift to talk about all these quantities, collectively. For instance, we refer to galaxy seen 1 billion years after the Big Bang as "*a redshift 2 galaxy*", we observe "*the high-z Universe*" (with disparate definitions among different researchers), or we speak of "*the evolution of galaxies with redshift*".

Because the Universe was smaller in the past, the relative expansion of the Universe decreases. Hence, the time span between, say, $z = 0$ (today) and $z = 1$ is much longer than between $z = 10$ and $z = 11$, namely 8 billion years (Gyr) and 60 million years (Myr), respectively. Partly for the same reason, however, the timescales of evolution — not only of galaxies but for all physicall processes — were shorter at earlier times.

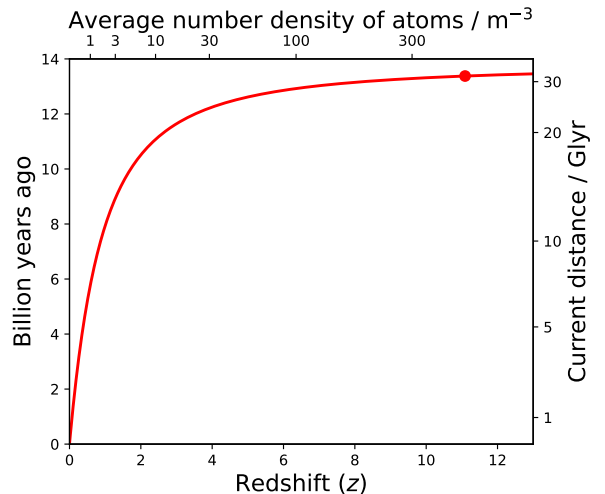The relationship between redshift and other quantities is shown in Fig. 1.

medium (ISM) is therefore enriched with metals, part of which condenses to dust.

Thermal and kinetic feedback from stars and AGN drive strong winds which may exceed the galaxy's escape velocity, enriching also the intergalactic medium (IGM) with metals. Star formation typically declines after an initial starburst, but may be sustained by continuous accretion of new material from the IGM, while new starbursts may be initiated by collisions with other galaxies, a process known as *merging*. Galactic winds, merging, and gas depletion is also responsible for some galaxies ceasing to form new stars.

Which processes dominate will determine the nature of the galaxy, in particular its morphology: will it end up as a disky spiral galaxy, a featureless elliptical galaxy, or something else?

2.1. The luminosity function

One of most popular ways of describing statistically a population of galaxies and its evolution through cosmic time is the *luminosity function* (LF) — the probability distribution function of the luminosities of galaxies in a given wavelength region, e.g. rest-frame ultraviolet light. Carefully taking into consideration observational pitfalls such as differential dust extinction and various selection biases (e.g. faint galaxies are progressively more difficult to see at high redshift than bright galaxies), the physical properties and the evolution of a given galaxy type can then be studied through time.
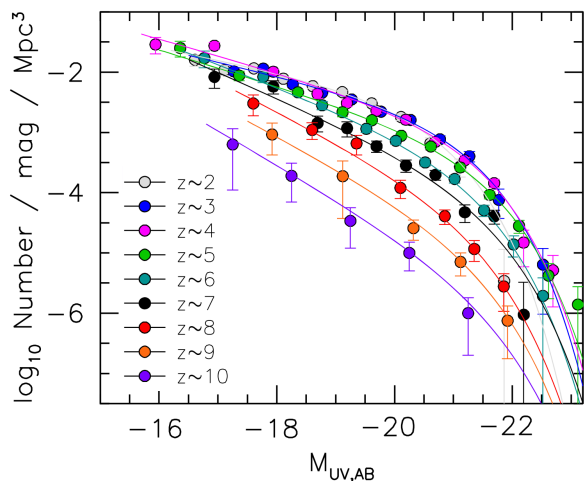
All structure in the Universe is, ultimately, born out ouf primordial (quantum) fluctuations in the density field. Observations of the cosmic microwave background (CMB) show that apparently these fluctuations can be described by an almost scale-free power spectrum (Planck Collaboration et al. 2018). In other words, the Universe does not have a preferred scale, and neither does gravity (there are no characteristic masses in the Einstein equations). We might therefore expect structure to form in a self-similar way: If the ratio between the number of structures with masses M and $10M$ is f, then the ratio between the number of structures with masses $\frac{1}{10}M$ and M is also f. An if all halos contain the same baryonic fraction and are equally effective in converting gas to stars, we might expect the LF to resemble the distribution of masses in the Universe.

The Universe, it turns out, is more complex than this, as will be elucidated further down. Observationally, we do however see a LF that is remarkably universal — that is, it can be rather accurately parametrized by same functional form for all galaxy types, and all redshifts: As seen in Fig. 2, at low luminosities the LF is characterized by a power law, while above a certain characteristic luminosity there is an exponential cut-off. Observed luminosities are therefore often fitted with a so-called Schechter function (Schechter 1976) with three parameters that are a function of redshift: The faint-end slope α, the normalization ϕ^*, and the "knee" luminosity, L^* (pronounced "L-star"):

$$N(L)dL = \phi^* \left(\frac{L}{L^*}\right)^{\alpha} e^{-L/L^*} \frac{dL}{L^*}. \tag{1}$$

The cut-off implies that there is in fact such a thing as a "typical galaxy", roughly comparable to our own Milky Way. The fact that the number density of galaxies larger than the Milky Way declines exponentially fast means that, if one integrates the total stellar mass of all galaxies, Milky Way-sized galaxies dominate the total stellar budget in the Universe. It is thus not entirely coincidental that we live in an L^* galaxy, although other factors than just the sheer number of stars are important (not all galaxies are equally habitable).

Fig. 2. The evolving luminosity function of galaxies from the Universe was ∼ 500 Myr (z ∼ 10; purple) till it was 3.3 Gyr old (z ∼ 2; gray). Along the *y* axis is differential number density, while the luminosity along the *x* axis is expressed as *magnitudes*; astronomers' notorious units for brightness, where a more negative number means a brighter galaxy. The lines show the best-fit Schechter functions. The figure is a slightly modified version of Fig. 5 in (Bouwens et al. 2021). **Should ask Rychard for permission to reproduce.**

The evolution of the LF with time has proved an invaluable tool for teaching us how galaxies have evolved since their formation. Unlike earlier attempts to characterize galaxy populations, the Schechter function is not purely phenomenological, but has its roots in the underlying distribution of matter. In the following I will attempt to shed light on how from a timescale perspective we can understand, to a certain extent, the physics of the LF.

2.2. The halo mass function

One of the big discoveries of the 1970s was that, in contrast to the prevailing picture at the time (Eggen et al. 1962), galaxies do not form from huge, "monolithically collapsing" clouds that later fragment to stars. Rather, it seems to be the other way round, with small structures forming first, later building up to larger structures in a hierarchical manner. This was first realized by Press & Schechter (1974), and later refined by e.g. Sheth & Tormen (2002) and Tinker et al. (2008), who calculated the collapse of gravitating clouds from an initial smoothed density field. The resulting *halo mass function* (HMF) is analogous to the LF, but for DM halo masses rather than galaxy luminosities. Since baryons and dark matter was originally mixed in the ratio ∼1:5, one might expect that the LF would mimic the HMF, albeit shifted. But this is not at all what we see! Although they are both characterized by a power law and an exponential cut-off, the HMF cut-off is at much high masses than the halos hosting L^* galaxies.

The exponential departure of the HMF from a universal power law derives from the Gaussian (or at least near-Gaussian) distribution of clumps in the inital density field. Structure forms on all scales, but because no information can be gained from outside the cosmic event horizon — i.e. the "boundary" marking the sphere within which light has had the time to travel since the Big Bang — small structures form before larger ones. The relevant timescale that sets the maximum size of structures in the Universe at a given time is the *Hubble time*, defined at a time t as $t_H(t) = 1/H(t)$, where $H(t)$ is the Hubble parameter that measures the expansion rate of the Universe[2]. If the Universe had always expanded at its current rate, t_H would be equal to the age of the Universe. In most cases we can ignore physical processes that happen on timescales longer than t_H. For instance, this timescale sets the primordial abundances of elements created through nucleosynthesis in the first 20 minutes after the Big Bang.

2.3. Gas cooling in dark matter halos

How may we understand the shape of the LF? The first clues to this problem were provided by Rees & Ostriker (1977) who simply compared two timescales:

A cloud of gas and DM that meets the Jeans criterion and is able to withstand the expansion of the Universe collapses first in a free fall. In the simple case of a spherically symmetric collapse, all particles of a pressureless gas meet in the center in the free-fall time $t_{\rm ff}$ given by

$$t_{\rm ff} = \left(\frac{3\pi}{32G\rho_{\rm m}}\right)^{1/2} = \left(\frac{3\pi f_{\rm b}}{32Gn\mu m_{\rm H}}\right)^{1/2}, \tag{2}$$

where G is the gravitational constant, and $\rho_{\rm m}$ is the mass density. In the early Universe, before dark energy played any role (see info box 2.3), the density simply evolved as $\rho_{\rm m}(z) = \rho_{\rm m,0}(1 + z)^3$ where $\rho_{\rm m,0} = \rho_{\rm DM,0} + \rho_{\rm b,0}$ is the present-day density. In the second step we have expressed this density in terms of the gas through the universal baryonic fraction $f_{\rm b} = \rho_{\rm b}/\rho_{\rm m} = 0.16$ (Planck Collaboration et al. 2018), the particle density n, and the mean molecular mass μ in terms of the hydrogen mass $m_{\rm H}$.

But the gas is *not* pressureless, and an initially adiabatic collapse therefore increases the gas temperature until thermal pressure prevents the cloud from contracting further and the gas is shock-heated to the *virial temperature*:

$$\begin{aligned} T_{\rm vir} &= \frac{\mu m_{\rm H}}{2k_{\rm B}}\frac{GM_{\rm vir}}{R_{\rm vir}} \\ &\sim 2 \times 10^4 \left(\frac{M_{\rm vir}}{10^8 M_\odot}\right)^{2/3}\left(\frac{1+z}{10}\right) \rm K, \end{aligned} \tag{3}$$

where $M_{\rm vir}$ and $R_{\rm vir}$ are the virial mass and radius of the cloud, respectively (i.e. the subscript "vir" is used to denote values that the cloud reaches once it has virialized). In the second line we have used that a collapsing could reaches an overdensity with respect to the average density $\bar\rho_{\rm m}$ of $\Delta \sim 180$ (e.g. Binney & Tremaine 2008), as well as the fact that $\bar\rho_{\rm m}$ increases with redshift as $(1 + z)^3$.

In order to form stars, the gas needs to be much colder, $T \sim 10^2$ K. In other words, only the smallest halos would be able to form any stars at all. The key to more massive halos being able to host stars is the gas' ability to *cool* through radiative processes, that is, they can convert their kinetic energy to radiation which is then able to leave the system.

Various physical mechanisms dominate the cooling process at different temperatures. At very high temperatures, above $T \sim 10^{6\text{-}7}$ K where everything is ionized, cooling takes place through free-free emission, or Bremsstrahlung, where accelerated charged particles (mostly electrons) emit photons. If the gas

---

[2] The expansion rate measures the recession velocity at a given distance, typically in km/s per Mpc, or "mega-parsec", where 1 pc ≃ 3.3 lightyears. The current value of $H(t)$ is called the Hubble constant, H_0, and is roughly equal to $H_0 \simeq 70\,\rm km\,s^{-1}\,Mpc^{-1}$.

Expansion of the Universe

The Universe consists of several different components which affect its dynamics differently: Dark energy (DE) has a constant energy density; more DE is therefore created along with the expansion of the Universe. Matter (m) consists of dark matter (DM) and baryons (b) which both act "normally"; its density decreases proportionally to the volume because the total amount is conserved (except when converted to radiation). Because redshift is inversely proportional to the "size", or *scale factor* a — i.e. the redshift of an object observed when the scale of the Universe was a factor a compared to today — the density of matter evolves with redshift as $\rho_m(z) = \rho_{m,0}(1 + z)^3$. Radiation (r) may of course be created and destroyed, but the number of photons is largely conserved. However, in addition its energy density is diluted due to its redshift, and the density therefore goes as $\rho_r(z) = \rho_{r,0}(1 + z)^4$.

Because of the differing exponents, the different components dominate on different timescales: In the very early Universe, dynamics were dominated by radiation until an age of roughly $t = 50\,000$ yr, after which we entered the matter-dominated era. The future will bring an exponential expansion due to DE taking over around $t \sim 10$ Gyr, but it is a rather remarkable fact that we live in an era where both matter and DE governs the expansion.

Fig. 3. Particle density n as a function of virial temperature $T_{\rm vir}$. The black locus marks the values where the cooling timescale $t_{\rm cool}$ equals the dynamical timescale $t_{\rm ff}$. Above this curve, in the green region, clouds are able to cool sufficiently to spark star formation. Slanted, dashed lines show the n-$T_{\rm vir}$ values of halos with the indicated masses. Horisontal blue lines mark typical densities of virialized halos at the indicated redshifts.

is rich in *metals* — astronomer's notion for all elements heavier than helium, many electronic transitions at various wavelengths exist that may cool the gas efficiently. But in the early Universe, before the gas was polluted with metals, the only efficient cooling mechanisms at lower temperatures was through hydrogen and helium: At the lowest temperatures, collisions excite the atoms with subsequent de-excitation leading to emission. At slightly higher temperatures, the atoms are ionized; subsequent recombination that do not go directly to the ground state (which would just result in a new ionizing photon) cascade down through intermediate states, emitting low-energy photons. The effectiveness of these pristine elements peaks at $T \sim 2 \times 10^4$ K and $T \sim 10^5$ K for hydrogen and helium, respectively (see e.g. Katz et al. 1996).

The total cooling function resulting from these processes is denoted $\Lambda(T)$ and can be calculated from quantum physics (e.g. Sutherland & Dopita 1993); it is defined such that the cooling *rate* is

$$\frac{dE_{\rm cool}}{dt} = n^2 \Lambda(T). \qquad (4)$$

The time it takes the cloud to radiate away its kinetic energy $K = \frac{3}{2} n k_B T$ is thus

$$t_{\rm cool} = \frac{E}{dE_{\rm cool}/dt} = \frac{3}{2} \frac{k_B T}{n \Lambda(T)}. \qquad (5)$$

From Eqs. 2 and 5 we see that the question of whether or not a cloud will be able to form stars efficiently boils down to an interplay between its density and its temperature. Figure 3 compares the cooling timescale to the free-fall timescale in an n vs. T plane. The plot is divided into two regions showing where the cooling timescale is *less than* (green) and *greater than* (red) the free-fall timescale. In the green domain, the gas collapses
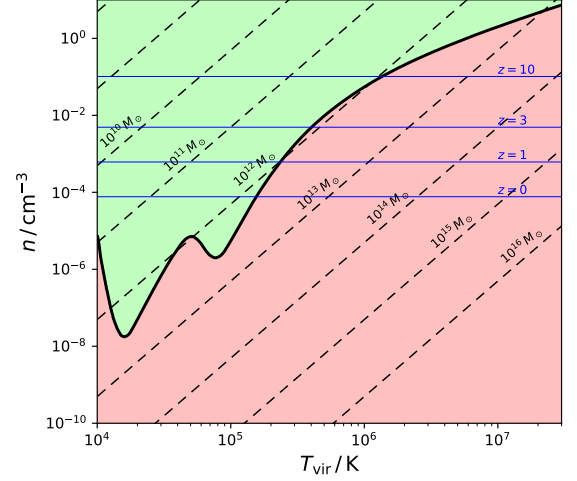
more or less freely with negligible thermal pressure while in the red domain, the cloud is pressure-supported in a quasi-static state where cooling can take longer than the Hubble time.

The shape of the $t_{\rm ff} = t_{\rm cool}$ locus (black) resembles the cooling function, with $n \propto (T/\Lambda)^{1/2}$, and is calculated assuming a metal-free gas. For a Solar metallicity cloud (i.e. characteristic of the present-day Universe), cooling is more efficient by more than an order of magnitude in the $T \sim 10^{4.5-6.5}$ K regime, but the low-density trough of the locus still lies at $M \sim 10^{12} M_\odot$. The slanted lines show densities and temperatures of halos of constant mass. We see that clouds of $M \sim 10^{12} M_\odot$ can cool even at low densities while at higher and lower masses, cooling is less efficient. In other words, the stellar mass function should be suppressed on either side of $M \sim 10^{12} M_\odot$; the mass of the Milky Way halo.

This differential cooling explains why we should not expect the stellar mass function, and hence the LF, to simply mirror the HMF, shifted by the baryon fraction f_b. But observed stellar mass functions are shallower at low masses, and steeper at high masses, than predicted by these simple timescale arguments. In the calculations of the region of efficient cooling in Fig. 3 we assumed a halo overdensity of $\Delta = 180$. But this is the average overdensity inside the virial radius; in reality the halo will have a decreasing density profile with much higher densities in the center, so at least some of the gas should be able to cool.

Taking into account the fact that gas accumulates and condensates in the bottom of the potential well created by the halo, White & Rees (1978) calculated a LF. The result was still that the typical galaxy is hosted by a $M_{\rm vir} \sim 10^{12} M_\odot$ halo. But now astronomers faced another nuisance: The high density allowed galaxies to cool *too* efficiently, vastly overpredicting the number density of especially small galaxies.

The rescue for this "overcooling" problem is believed be, mainly, three physical processes known as *photoionization*, *merging*, and *feedback*.

2.4. Photoionization

In the "dark ages", before the advent of the first luminous sources, the Universe was filled with neutral gas which absorbed all light blueward of the ionization threshold at $\lambda = 912\,\text{Å}$ (Gnedin 2000; Barkana & Loeb 2001). Strong UV radiation created ionized bubbles around the first galaxies which percolated the Universe and eventually overlapped. The resulting metagalactic UV background permeates and heats the intergalactic medium (IGM) and increases its pressure, preventing accretion of gas onto halos, as well as reduced the rate of radiative cooling within the halos (Benson et al. 2002), inhibiting star formation. Small halos are particular vulnerable to this effect, resulting in a shallower slope in the LF. The UVB increases to a maximum and then decreases, with a characteristic value of $\Gamma \sim 10^{-12}$ photons s^{-1}, leading to a characteristic timescale for photoionization of $t_{\text{ph}} \sim 1/\Gamma \sim 3 \times 10$ yr; much shorter than the typical dynamical timescales involved.

The exact timescales involved in this so-called "epoch of reionization" (EoR) are still not entirely clear: What was its duration (when did it start and end; was it short and intense or more prolonged?), which sources contributed more (small galaxies, massive galaxies, or even quasars?), and what was its topology (small and fizzly vs. large bubbles)?

Analytical solutions exist for calculating the propagation of an ionized bubble around a source in a homogeneous gas, relating the output rate of ionizing photons to the timescale of recombination (Strömgren 1939; Dopita & Sutherland 2003). Relating these timescales, however, only puts weak constraints on the EoR; the largest advances in this field have been made through numerical simulations which can capture the highly non-linear nature of the involved densities, temperatures, velocities, etc. (see e.g. Hutter et al. 2020; Perez et al. 2022, for recent advances regarding the topology of reionization). A paramount problem in such calculations is the enormous range in spatial scales involved (see info box): Typical bubbles have sizes of the order of ten(s) of Mpc (Wyithe & Loeb 2004; Giri et al. 2018), but photons begin ionizing gas already on sub-parsec scales, in the molecular clouds enshrouding the stars. Most simulations have therefore had to focus either on the escape of ionizing radiation through the interstellar medium (ISM) of individual galaxies, or on the ionization of the large scales of the intergalactic medium (IGM), regarding individual galaxies simply as (unresolved) photon sources (e.g. Jensen et al. 2013, 2014), although intermediate-scale simulations also do exist (e.g. Rosdahl et al. 2022).

2.5. Galaxy merging

Even today, the separation between galaxies is not vastly larger than their cross section, and at $z \sim 10$, where number densities were 10^3 times larger, encounters between galaxies were frequent. Galaxies therefore interact gravitationally and occasionally merge with other galaxies. Minor mergers, where a galaxy of mass M_1 accretes a smaller galaxy of mass $M_2 \lesssim M_1/10$ are frequent at all times, but even major mergers events, with $M_1 \sim M_2$, happen on average once per Hubble time (Lacey & Cole 1993). The effect of this process is to increase the number of large galaxies, at the expense of small galaxies (e.g. Cole 1991; Kauffmann et al. 1993).

To first order, the relevant timescale for accretion of satellite galaxies is set by a physical process known as *dynamical friction*, described first by Chandrasekhar (1943), where the gravity of a large body moving through an "atmosphere of colli-

---

> **ⓘ Adaptive resolution in numerical simulations.**
>
> A fundamental challenge in numerical simulations — not only in astrophysics — is the vast dynamic range in both space and time. Computers are limited by both memory and processing power, and trade-offs must necessarily be made between simulating sufficiently large spatial and temporal scales, and resolving these scales sufficiently to capture the small-scale physics. Various approaches exist to alleviating these limitations; cosmological simulations, modeling a more or less representative chunk of the Universe, usually imploy an adaptive resolution.
>
> Structures can be simulated using *particles* that each represent a given amount of mass (gas, dark matter, and/or stars). This is analogous to simple *N*-body codes, but particles are "smoothed" in space (Gingold & Monaghan 1977; Lucy 1977). The particles follow trajectories dictated by forces (gravitational, hydrodynamical, magnetic, etc.), and resolution thus increases automatically in high-density regions. Alternatively, structure can be simulated on a fixed grid of cells, analogous to pixels in a picture but in 3D. Adaptive resolution can then be achieved by refining cells according to a pre-defined criterion (usually a density threshold), splitting a cell up in 8 cells, recursively (Berger & Oliger 1984; Berger & Colella 1989). The two schemes both have their pros and cons. Additionally, a relatively new technique is a "moving mesh" that incorporates the best of two worlds; here space is split up into cells, but the cells have arbitrary (polyhedral) shapes, and they are not fixed in space but move around (Springel 2010).
>
> Not only space, but also time may be refined. In dense regions where matter undegoes large accelerations, numerical errors will be introduced if the time step is too small. In order to avoid wasting time *when* matter is not dense, time steps may adapt to the density. Likewise, to avoid wasting time *where* matter is not dense, different regions may use different time steps such that the equations of motion are advanced first on the smallest time steps, and then on successively larger scales until the largest time step "catches up" with the smaller.

---

sionless gas" (in this case the DM halo) creates a "wake" of increased mass, exterting a drag on the body. For minor mergers, the timescale with which the orbit of a satellite decays decreases with increasing mass; Cole et al. (1994) found that a simple model with

$$t_{\text{merge}} \sim 0.5 t_{\text{H}} \left(\frac{M_1}{M_2} \right)^{0.25} \tag{6}$$

follows the trend of simulated mergers. For sufficiently large ratios between M_1 and M_2 (in this case $M_2/M_1 \gtrsim 16$), the merging timescale exceeds the Hubble time, consistent with the fact that massive galaxies still have satellites today. More sophisticated models (e.g Boylan-Kolchin et al. 2008) yield similar conclusions.

2.6. Feedback

Even with the physical mechanisms described above, galaxy formation models struggled to match observed LFs, apparently pro-

ducing too many stars. This annoyance entailed the need for *feedback* processes which could regulate star formation. Early models (e.g. White & Frenk 1991; Somerville & Primack 1999; Efstathiou 2000) employed only supernova feedback which inject both thermal energy and momentum into the ISM. The efficacy of this process can be understood from a simple timescale perspective: Star formation in galaxies occurs basically on the dynamical timescale which is of the order 100 Myr. In contrast, supernovae result from the explosions of the most massive stars which has life times of $\lesssim 10$ Myr.

Supernova feedback not only reheats the cold gas but may also expel it altogether in galactic "winds", and is thus very efficient at inhibiting star formation. The effect is particular pronounced in low-mass galaxies which have a shallower gravitational potential, allowing gas to escape more easily (Dekel & Silk 1986; Kauffmann et al. 1993; Natarajan 1999). In fact, it turned out to be a bit too efficient, because the expelled gas then becomes available for accretion onto more massive halos (galaxy groups and clusters) where at late times it cools and forms galaxies much more massive than observed (Benson et al. 2003). Various mechanisms were proposed to solve this overcooling problem. Arguably the most promising mechanism to prevent too many massive galaxies is feedback from active galactic nuclei — that is, gas accreting onto supermassive black holes in the center of galaxies, radiating excess energy away at extreme energies. This effect has been successful at matching LFs at low and high redshifts (e.g. Croton et al. 2006; Bower et al. 2006; Somerville et al. 2008; Mccarthy et al. 2011).

2.7. From halo mass functions to luminosity functions

The physical processes described in Sec. 2.1 are shown in Fig. 4. The blue line shows the HMF at $z = 0.1$ ("almost today"); a Sheth & Tormen (2002) HMF with a z-dependent correction factor which brings the analytical prediction very close to results from a large-scale, ~ 8 billion particle, cosmological simulation Klypin et al. (2011); its functional form is described in Laursen et al. (2019). Multiplying the mass by the cosmic baryon fraction f_b shifts the HMF to the left, shown by the magenta line. That is, if galaxies at all masses were able to cool and convert all their gas to stars, this line is what we would observe. Reality is quite different: Observed stellar mass functions at $z \sim 0.1$ at the low-mass end (Wright et al. 2017) and the high-mass end (Bernardi et al. 2013) are shown with olive green and steel blue points, respectively. The physical mechanisms responsible for this drastic suppression are indicated by the green and orange arrows.

3. The first galaxies

On Christmas day 2021, the long awaited James Webb Space Telescope (JWST) was finally launched. Half a year later, after a journey to its "destination" L2, 1.5 million kilometers from Earth, cooling to 6–40 K, commissioning and calibration of its four instruments, the astronomy community finally saw science data in July 2022. A proclaimed goal of JWST is to look for some of the very first galaxies, and it is therefore expected that we will eventually find galaxies with redshifts higher than the current "record-holder", GN-z11, which has a redshift of 11.09 (Oesch et al. 2016).

Already five days after the science data started flowing, the first report of a redshift record-breaking galaxy came, when Naidu et al. (2022) announced, on the preprint server arXiv, the discovery of two galaxies with apparent redshifts of $z \simeq 11$ and $z \simeq 13$. The two galaxies were dubbed GLASS-z11 and
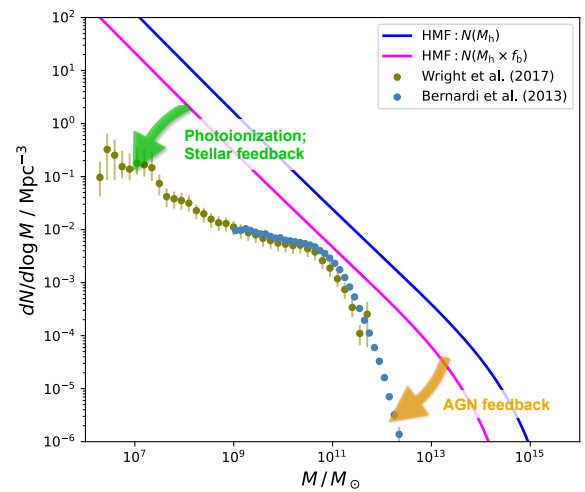
Fig. 4. Number density of halo masses and stellar masses as a function of mass. Theoretical prediction from a HMF are shown in blue and magenta for total halo masses and baryon masses, respectively. They are compared to observed stellar mass functions at the low-mass end (olive) and high-mass end (steel-blue). Photoionization and stellar feedback is believed to suppress the mass function at the low-mass end (green arrow), while AGN activity suppresses the high-mass end (orange arrow).

GLASS-z13, respectively. Simultaneously with this, Castellano et al. (2022) reported a sample of galaxies with redshifts up to $z \simeq 15$. In the days to follow, more astounding discoveries came, with galaxies at $z \sim 17$ (Donnan et al. 2022; Harikane et al. 2022), and even $z \sim 20$ (Yan et al. 2022). Since the first galaxies are believed to have formed around $z \sim 20$–30, some 100–200 Myr after the Big Bang, could it really be some of the very first galaxies, consisting of stars made of pristine gas?

What is even more remarkable is that the galaxies seemingly are very massive; indeed the inferred stellar masses are much higher than expected from our understanding of hierarchical clustering and galaxy assembly. The $z \simeq 17$ galaxy, CEERS-1749, has a stellar mass of almost $10^{10} M_\odot$, a tenth of the Milky Way.

Figure 5 shows the four galaxies — GN-z11, GLASS-z11, GLASS-z13, and CEERS-1749 — in a stellar mass vs. age of the Universe plot. The blue lines shows the maximum expected mass of a galaxy at a given time in the volume surveyed, if we assume that 1) galaxies at that time were 100% efficient at converting gas to stars (which is a factor of at least 10 more efficient than in the less high-z Universe), and 2) the standard model of the structure and evolution of the Universe, the so-called ΛCDM model, is correct. Above this line, in the red domain, galaxies are unlikely to have been able to build up so much mass. Three of the galaxies are on the verge of being probable (if we assumed a star formation efficiency similar to lower redshifts) while CEERS-1749 seemingly defies the timescales of ΛCDM.

There are however a number of caveats that must be considered before claiming that the record has been broken. First of all, at the time of writing the papers have not been peer-reviewed. More importantly, the method by which the redshifts have been measured is a somewhat unreliable way. None of them have spectra, which would be needed to measure exact redshifts. Instead the redshifts are estimated through the so-called "dropout technique": At these high redshifts, the Universe is predominantly neutral (we are deep in the epoch of reionization), so all light blueward of the Lyman break at 912 Å is absorbed, and galaxies will appear black — or *drop out* — in images taken with
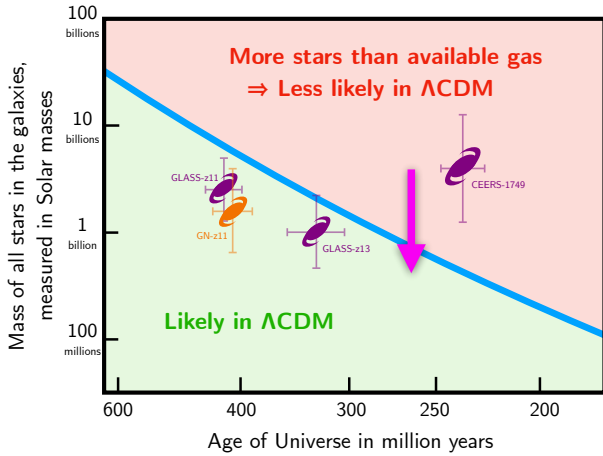
Fig. 5. Stellar masses of some galaxies recently observed with JWST, vs. the age of the Universe at the time we see them.

filters blueward of this break. The higher the redshift a galaxy has, the redder the filter it will drop out of, so imaging a field with multiple filters gives a crude estimate for the redshift.

However, without a spectroscopical redshift to confirm spectral features, we cannot be certain that the galaxy is not red for another reason than redshift. Dusty galaxies also appear red, because dust absorbs more easily shorter wavelengths. Galaxies with old stellar populations, where the massive blue stars have died, similar appear red. Even brown dwarfs in the Milky Way may sometimes allude the observer. For this reason the galaxies will be followed up spectroscopically.

Other explanations have also been proposed. Mason et al. (2022) showed that stochastic star formation, where stars sometimes form slowly and other times fast, as well as more dust-free environments, could make a fraction of the galaxies appear much brighter than average, and that we are seeing "the tip of the iceberg". There is also the possibility that we have wimply been lucky and pointed the telescope in a lucky direction with galactic overdensities. Finally, Steinhardt et al. (2022) point out that the templates that are used to calculate redshifts from the series of images through different filters are based on physics known from the local Universe. But there are several reasons to believe that this is not appropriate at very high redshifts. For instance they argue that, because the Universe was hotter in the past, the mass distribution of stellar populations was significantly different from today. As domenstrated by (Sneppen et al. 2022), accounting for the different gas temperature in the early Universe, the calculated masses of the galaxies seems to be significantly smaller, probably at least ten times. In other words, the galaxies in Fig. 5 would move down at least by an amount indicated by the magenta arrow, placing them in the more likely domain.

In conclusion, will it is not impossible that we will have to remedy ΛCDM, we will have to wait for more data, in particular spectroscopic data and more wide-field surveys.

References

Barkana, R. & Loeb, A. 2001, Phys. Rep., 349, 125

Benson, A. J., Bower, R. G., Frenk, C. S., et al. 2003, ApJ, 599, 38

Benson, A. J., Lacey, C. G., Baugh, C. M., Cole, S., & Frenk, C. S. 2002, MNRAS, 333, 156

Berger, M. & Colella, P. 1989, J. Comput. Phys., 82, 64

Berger, M. J. & Oliger, J. 1984, J. Comput. Phys., 53, 484

Bernardi, M., Meert, A., Sheth, R. K., et al. 2013, MNRAS, 436, 697

Binney, J. & Tremaine, S. 2008, Galactic Dynamics, second edi edn., Vol. 1 (Princeton University Press, Princeton, NJ USA)

Bouwens, R. J., Oesch, P. A., Stefanon, M., et al. 2021, AJ, 162, 47

Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645

Boylan-Kolchin, M., Ma, C. P., & Quataert, E. 2008, MNRAS, 383, 93

Castellano, M., Fontana, A., Treu, T., et al. 2022, 1324 [arXiv:2207.09436]

Chandrasekhar, S. 1943, ApJ, 97, 255

Cole, S. 1991, ApJ, 367, 45

Cole, S., Aragón-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E. 1994, MNRAS, 271, 781

Croton, D. J., Springel, V., White, S. D., et al. 2006, MNRAS, 365, 11

Dekel, A. & Silk, J. 1986, ApJ, 303, 39

Donnan, C. T., McLeod, D. J., Dunlop, J. S., et al. 2022, 12, 1

Dopita, M. A. & Sutherland, R. S. 2003, Astrophysics of the Diffuse Universe, Astronomy and Astrophysics Library (Berlin, Heidelberg: Springer Berlin Heidelberg)

Efstathiou, G. 2000, MNRAS, 317, 697

Eggen, O. J., Lynden-Bell, D., & Sandage, A. R. 1962, ApJ, 136, 748

Gingold, R. A. & Monaghan, J. J. 1977, MNRAS, 181, 375

Giri, S. K., Mellema, G., Dixon, K. L., & Iliev, I. T. 2018, MNRAS, 473, 2949

Gnedin, N. Y. 2000, ApJ, 535, 530

Harikane, Y., Ouchi, M., Oguri, M., et al. 2022 [arXiv:2208.01612]

Hubble, E. P. 1926, ApJ, 64, 321

Hutter, A., Watkinson, C. A., Seiler, J., et al. 2020, MNRAS, 492, 653

Jensen, H., Hayes, M., Iliev, I. T., et al. 2014, MNRAS, 444, 2114

Jensen, H., Laursen, P., Mellema, G., et al. 2013, MNRAS, 428, 1366

Jovi, J. B. 1990, St. Fe (Blaze Glory)

Katz, N., Weinberg, D. H., & Hernquist, L. 1996, ApJS, 105, 19

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102

Lacey, C. & Cole, S. 1993, MNRAS, 262, 627

Laursen, P., Sommer-Larsen, J., Milvang-Jensen, B., Fynbo, J. P. U., & Razoumov, A. O. 2019, A&A, 84

Lucy, L. B. 1977, AJ, 82, 1013

Mason, C. A., Trenti, M., & Treu, T. 2022, 6, 1

Mccarthy, I. G., Schaye, J., Bower, R. G., et al. 2011, MNRAS, 412, 1965

Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022 [arXiv:2207.09434]

Natarajan, P. 1999, ApJ, 512, L105

Oesch, P. A., Brammer, G., van Dokkum, P. G., et al. 2016, ApJ, 819, 129

Perez, L. A., Malhotra, S., Rhoads, J. E., Laursen, P., & Wold, I. G. B. 2022 [arXiv:2210.05722]

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, 1
Press, W. H. & Schechter, P. 1974, ApJ, 187, 425
Rees, M. J. & Ostriker, J. P. 1977, MNRAS, 179, 541
Rosdahl, J., Blaizot, J., Katz, H., et al. 2022, MNRAS, 515, 2386
Schechter, P. 1976, ApJ, 203, 297
Shapley, H. & Curtis, H. D. 1921, Bull. Natl. Res. Counc., 2, 171
Sheth, R. K. & Tormen, G. 2002, MNRAS, 329, 61
Sneppen, A., Steinhardt, C. L., Hensley, H., et al. 2022, ApJ, 931, 57
Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., & Hernquist, L. 2008, MNRAS, 391, 481
Somerville, R. S. & Primack, J. R. 1999, MNRAS, 310, 1087
Springel, V. 2010, MNRAS, 401, 791
Steinhardt, C. L., Kokorev, V., Rusakov, V., Garcia, E., & Sneppen, A. 2022 [arXiv:2208.07879]
Strömgren, B. 1939, ApJ, 89, 526
Sutherland, R. S. & Dopita, M. A. 1993, ApJS, 88, 253
Tinker, J. L., Kravtsov, A. V., Klypin, A., et al. 2008, ApJ, 688, 709
White, S. D. M. & Frenk, C. S. 1991, ApJ, 379, 52
White, S. D. M. & Rees, M. J. 1978, MNRAS, 183, 341
Wright, A. H., Robotham, A. S., Driver, S. P., et al. 2017, MNRAS, 470, 283
Wyithe, J. S. B. & Loeb, A. 2004, Nature, 432, 194
Yan, H., Ma, Z., Ling, C., et al. 2022 [arXiv:2207.11558]