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.

October 11, 2022

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 dynamical 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, 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 — properties such as their star formation rates, build-up of heavy elements, and changes in morphology.

On the theoretical side, as in most 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 say something meaningful about the most beautiful structures in the Universe — the galaxies.

2. Timescales in galaxy evolution

Brief intro to this section Brief intro to this section Brief intro to this section Brief intro to this section

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

2.1. Galaxy formation

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 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.

---

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

Fig. 2. Luminosity functions...

2.2. Using timescales to explain the properties of galaxies

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 doesn't 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.

The Universe, it turns out, is more complex than this. Observationally, we 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", incidentally (?) 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.

The evolution of the LF with time has proved an invaluable tool for teaching us how galaxies have evolved since their forma-

---

Cosmological redshift

Arguably, the most essential concept in astronomy is the *redshift z* of light. Known to all 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). Partly for the same reason, however, the timescales of evolution — not only of galaxies but in general — was shorter at earlier times.

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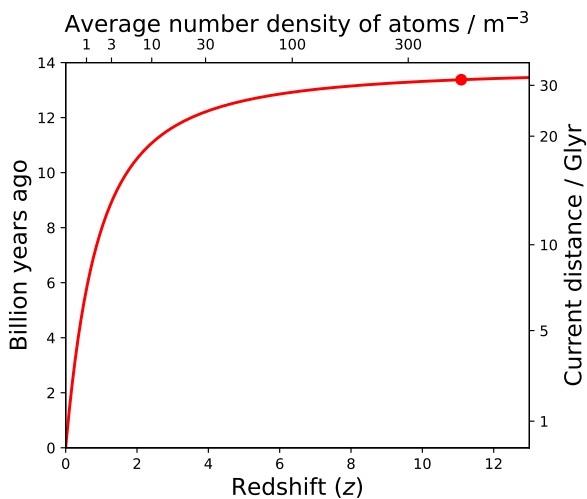

Fig. 1. Relationship between observed redshift of light emitted from a distant object, and various properties of the Universe. Because the relative expansion was larger in the past when the Universe was smaller, redshift increased more at that time asymptotically approaching infinity for light emitted at the time of the Big Bang, 13.8 Gyr ago.

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?
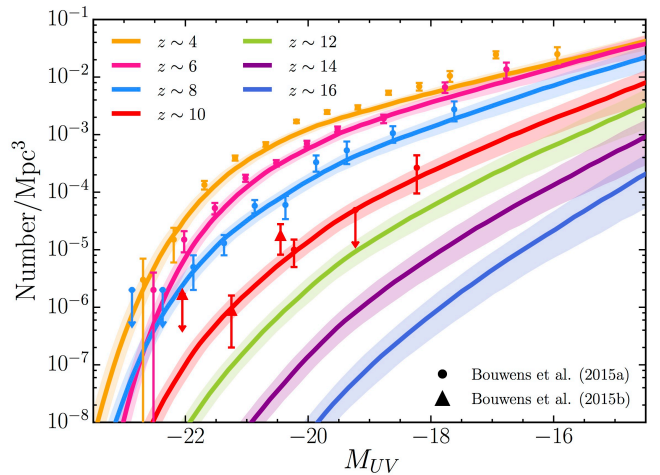
tion. Unlike earlier attempts to characterize galaxy populations, the Schechter function is not purely phenomenological, but has its roots in the underlying distribution of matter. In the following I will attempt to shed light on how from a timescale perspective we can understand, to a certain extent, the physics of the LF.

2.2.1. The halo mass function

One of the revelations 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.

2.2.2. 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) and White & Rees (1978) 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_{ff} given by

$$t_{ff} = \left(\frac{3\pi}{32G\rho_m}\right)^{1/2} = \left(\frac{3\pi f_b}{32Gn\mu m_H}\right)^{1/2}, \qquad (2)$$

where G is the gravitational constant, and ρ_m is the mass density. In the early Universe, before dark energy played any role, the

density simply evolved as $\rho_m(z) = \rho_{m,0}(1 + z)^3$ where $\rho_{m,0}$ is the present-day density. In the second step we have expressed this density in terms of the gas through the universal baryonic fraction $f_b = \rho_b/\rho_m = 0.16$ (Planck Collaboration et al. 2018) and the particle density n.

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

$$\begin{aligned} T_{vir} &= \frac{\mu m_H}{2k_B} \frac{GM_{vir}}{R_{vir}} \\ &\sim 2 \times 10^4 \left(\frac{M_{vir}}{10^8 M_\odot}\right)^{2/3} \left(\frac{1+z}{10}\right) \text{K}, \end{aligned} \qquad (3)$$

where μ is the mean molecular mass in terms of the hydrogen mass m_H, k_B is Boltzmann's constant, and M_{vir} and R_{vir} are the (virial) mass and radius of the cloud, respectively. In the second line we have used that a collapsing could reaches an overdensity with respect to the average density $\bar{\rho_m}$ of ~ 200, as well as the fact that $\bar{\rho_m}$ 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. Luckily, the gas is able 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 $T \sim 10^7$ K where everything is ionized, cooling takes place through free-free emission, or Bremsstrahlung, where electrons emit photons. If the gas 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 (higher) temperatures, collisions excite (ionize) the atoms with subsequent de-excitation (recombination) leading to emission. The effectiveness of these pristine elements peaks at $T \sim 10^4$ K and $T \sim 10^5$ K for hydrogen and helium, respectively.

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

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

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

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

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 plot. 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 $t_{ff} = t_{cool}$ locus (black) resembles the cooling function, with $n \propto (T/\Lambda)^{1/2}$, and is calculated assuming a

---

[2] The expansion rate is measures the recession velocity at a given distance, typically in km/s per Mpc, or "mega-parsec", where 1 pc $\simeq$ 3.3 lightyears. The current value of $H(t)$ is called the Hubble constant, H_0, and is roughly equal to $H_0 \simeq 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$.
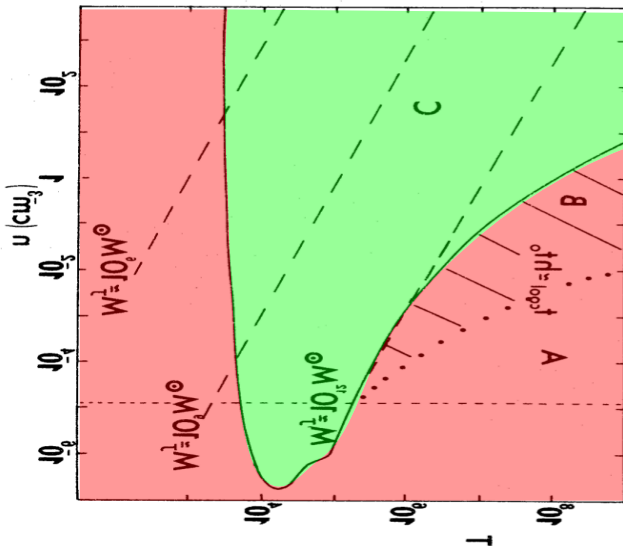
Fig. 3. Gas density n as a function of virial temperature T_{vir}. The black locus marks the values where the cooling timescale t_{cool} equals the dynamical timescale t_{ff}. Above this curve, in the green region, clouds are able to cool sufficiently to spark star formation. **Will make my own, pretty version, and also explain better.**

metal-free gas. For a Solar metallicity cloud, 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 whilw at higher and lower masses, cooling is less efficient. In other words, the galaxy mass function should be suppressed on either side of $M \sim 10^{12} M_\odot$; the mass of the Milky Way halo. At lower (higher) masses, the slope is shallower (steeper), resulting in a knee in the mass function.

2.2.3. Photoionization

The differential cooling described in Sec. 2.2.2 explains why we should not expect the galaxy mass function to simply mirror the HMF, shifted by a factor $1/f_{\mathrm{b}}$. But observed LFs are shallower (steeper) at low (high) masses than predicted by the simple timescale arguments above. The reason is thought be, mainly, three physical processes known as *photoionization*, *merging*, and *feedback*.

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 acceretion of gas onto halos, and reduces 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 exact timescales involved in this so-called "epoch of reionization" (EoR) are still not entirely clear: What was its du-

ration (when did it start and end; what 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 2014) A paramount problem in such calculations is the enormous range in spatial scales involved: 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).

2.2.4. 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 collisionless 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) find that a simple model with

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

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) using the more appropriate dynamical timescale, which is essentially the free-fall timescale to within factors of order unity, yield similar conclusions.

2.2.5. Feedback

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

Fig. 4. Caption caption

Fig. 5. Cippety cappety caption. The arrow indicates how much masses may decrease if using a T-dependent IMF.

lating them to the timescales of early galaxy formation, and then show Fig. 5.

ducing too many stars (Reference 1234). 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 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 especially in low-mass galaxies which have a shallower gravitational potential, allowing gas to escape more easily (?Kauffmann et al. 1993; ?). 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

Will write a bit more here, and end by referring to Fig. 4 which is supposed to summarize all of Sec. 2.2.

2.3. Galaxy quenching

Not sure if I'll include a section on this

—
—
—

3. The first galaxies

I thought I'd end with a section on the preliminary JWST results, along the lines of the news story I wrote recently, re-

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

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

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

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

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

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

Hutter, A. 2014, PhD thesis

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

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

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

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

Reference. 1234, Some J., 666, 666

Rosdahl, J., Blaizot, J., Katz, H., et al. 2022, MNRAS, 515, 2386

Schechter, P. 1976, ApJ, 203, 297

Sheth, R. K. & Tormen, G. 2002, MNRAS, 329, 61

Somerville, R. S. & Primack, J. R. 1999, MNRAS, 310, 1087
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
Wyithe, J. S. B. & Loeb, A. 2004, Nature, 432, 194