# Aggravating matters: accounting for baryons in cosmological analyses

Debackere, S.N.B.

# 1 Introduction



In Greek mythology, the cosmos (κόσμος) is an ordered state that came from chaos (χάος), the void[1]. In a similar vein, the field of cosmology attempts to fill up the metaphorical void in our understanding of the Universe: where did it come from and how does it evolve? While some form of cosmology has been practised for millennia, only in the last century has it evolved from a philosophical pastime into a fully fledged scientific subject. All it took was the development of large and powerful telescopes that allowed us to study the spectra of fuzzy nebulae—the nearest of which are visible with the naked eye in unpolluted skies—to come to the realization that these are actually conglomerations of stars, some of which are receding away from us at "unparalleled velocities" (Curtis, 1915; Wilson, 1915; Slipher, 1921, 1922; Stromberg, 1925). While this conclusion proved controversial initially as evidenced by the "Great Debate" between Shapley & Curtis (1921), the discovery of individual Cepheid variable stars in Andromeda and the Triangulum galaxy by Hubble (1925) provided a definitive proof for the immense distance and the extra-galactic nature of these and similar spiral nebulae. Lemaître (1927) was the first to interpret the high recession velocities to these galaxies as proof for an expanding Universe as predicted by general relativity. Hubble (1929) later explicitly showed that galaxies at larger distances recede away at higher velocities, as required for an expanding Universe. This breakthrough heralded the start of *observational* cosmology.

In the past almost century since this initial discovery of the increasing recession velocity of more distant galaxies, a wealth of disparate observations have elucidated the distribution of matter on cosmological scales, spanning millions to billions of lightyears. The picture emerged of an expanding Universe containing structure on different scales: galaxies bunch together inside groups and clusters, which form the filaments and nodes of an extensive network of overdensities, known as the cosmic web. Dark matter, which only feels gravity, forms the backbone of the cosmic web, constituting about $25\,\%$ of the total energy density of the Universe, whereas the ordinary matter that we are all familiar with, present in gas clouds, stars, planets and their inhabitants, only accounts for $5\,\%$. The final $70\,\%$ is contributed by the mysterious and unknown dark energy, which counteracts gravity and forces the Universe into a seemingly never-ending phase of accelerated expansion.

To shed light on the properties of dark energy, a slew of telescopes will see their first light in the coming decade. Their aim? To observe over a billion galaxies and use them to map the evolution of the total matter distribution—both dark and ordinary—over the past 10 billion years. The real breakthrough expected from these surveys is their ability to directly probe the era in which dark energy starts dominating the energy content of the Universe, about 3.6 billion years ago. The accelerating expansion of space counteracts the gravitational attraction of matter, slowing down the formation of new structures, leaving a clear mark on the distribution of galaxies. However, inferring the distribution of the total matter while only observing the galaxies is a challenge.

[1] Hesiod, Theogony, 110

Using weak gravitational lensing, a statistical method that measures the tiny but coherent distortion of galaxy shapes due to the curvature of space-time caused by intervening matter, it is possible to probe the full cosmologically evolving matter distribution. From these distortions, the total mass between us, the observers, and those galaxies can be inferred. When used to directly study the large-scale distribution of matter, these distortions are called cosmic shear, as the cosmic matter is shearing the images of galaxies. Weak lensing can also measure the total mass, including dark and ordinary matter, of individual objects, such as clusters of galaxies. Clusters are rare objects: they take a long time to form and they contain enormous amounts of matter. The change in the cluster abundance over time strongly constrains the evolution of the matter content and dark energy.

The problem this thesis focusses on, is interpreting these weak lensing observations to obtain the correct mass. A complex issue, since we need to know how the ordinary matter, known as baryons to cosmologists, traces the dominant dark matter. Only normal matter feels the electromagnetic force which substantially changes its behaviour on the smaller scales, cosmologically speaking, of galaxies—as a matter of fact, galaxies can only exist because of this fundamentally different behaviour. Paradoxically, however, not all of this *normal* matter is visible to us as a significant fraction of the total normal matter in the Universe resides in hot, low-density gas, which we are unable to observe directly with our current X-ray telescopes. Hence, we need to find a way to disentangle the unknown contribution of the normal matter from the measured weak lensing signal. Our suggested solutions combine the predictions of computer-simulated universes, evolving billions of particles in time, with simplified models that reproduce the observed distribution of hot gas inside clusters of galaxies, but that freely vary the amount of matter where no observations are available, to quantify how strongly our ignorance of the relation between ordinary and dark matter will affect the analysis of the aforementioned planned surveys. Additionally, we study how baryons affect cluster mass determinations and we suggest a new analysis method for cluster abundance studies that is less sensitive to our lack of knowledge of the exact distribution of normal matter in the outskirts of galaxy clusters.

## 1.1 The challenges of future surveys

Three major missions are planned for the next decade: the Euclid space telescope[2] (Laureijs et al., 2011), the Vera C. Rubin Observatory[3] in Chile (LSST Science Collaboration et al., 2009), and the Nancy Grace Roman Space Telescope[4] (Spergel et al., 2015). Collectively, they are called stage IV surveys by the community. Peaking back in time, these surveys aim to detect over 1 billion galaxies over the area of the sky not blocked by our own Milky Way. Different from the Hubble Space Telescope and large, ground-based telescopes such as the Very Large Telescope (VLT) or the Keck telescopes, these planned observatories are designed to have large fields-of-view to image huge swathes of the sky in a short amount of time.

[2] https://www.euclid-ec.org
[3] https://www.lsst.org/
[4] https://roman.gsfc.nasa.gov/

The Euclid space telescope has a $1.2\,\mathrm{m}$ primary mirror with a field-of-view of $0.5\,\mathrm{deg}^2$ (approximately twice the area of the Sun on the sky). To compare, the Wide Field Camera 3 on the Hubble Space Telescope covers a mere $0.002\,\mathrm{deg}^2$ of the sky. The planned Euclid survey will cover an area $> 15\,000\,\mathrm{deg}^2$—more than a third of the entire celestial sphere—probing back in time over the past 10 billion years. It will yield a sample of more than 1.5 billion galaxies whose distorted shapes will constrain the intervening matter. Additionally, the observations will generate a sample of $> 100\,000$ galaxy clusters whose abundance will constrain the late-time evolution of matter and dark energy. The survey will reach limiting magnitudes of $\approx 24.5$ in a single, wide visual band, with a small point-spread function (PSF) $< 0.2\,\mathrm{arcsec}$ for optimal galaxy shape measurements. Similar limiting magnitudes are expected in the three near-infrared filters within the wavelength range of $1.1 - 2.0\,\mu\mathrm{m}$, which will be used to photometrically measure the distance to the observed galaxies. With these properties, an average number density of $n_\mathrm{gal} \approx 30\,\mathrm{arcmin}^{-2}$ background galaxies is expected in each Euclid image, approximately $54\,000$ galaxies in total. To improve its distance calibrations, Euclid will rely on large, ground-based surveys to supplement its visual band observations.

The Earth-based Rubin observatory, on the other hand, boasts an $8.4\,\mathrm{m}$ primary mirror with an impressive $9.6\,\mathrm{deg}^2$ field-of-view, allowing it to—theoretically—image its entire $20\,000\,\mathrm{deg}^2$ survey area in only $\approx 2000$ exposures of less than $1\,\mathrm{min}$ each. The more familiar name of this survey is the Legacy Survey of Space and Time (LSST). Similarly to Euclid, the Rubin observatory is expected to reach limiting magnitudes of $\approx 24.5$ in a single exposure resulting in $n_\mathrm{gal} \approx 30\,\mathrm{arcmin}^{-2}$. The fast turnover time of the observations means that the LSST will be able to probe its full area more than a hundred times, allowing it detect galaxies to very high depths, yielding a baffling sample of 10 billion galaxies, out of which approximately 3 billion are expected to be useful for weak lensing. We will undoubtedly discover many unknown transient phenomena thanks to the high-cadence observations enabled by the massive field-of-view. Since the survey is ground-based, however, it will face varying observing conditions and seeing, significantly affecting its angular resolution and resulting in the blending of galaxies. Due to the large overlap between LSST and the Euclid survey, a close collaboration will benefit both surveys (Rhodes et al., 2017, 2019).

Finally, the Nancy Grace Roman Space Telescope consists of a $2.4\,\mathrm{m}$ primary mirror, just like Hubble, but with a significantly wider field-of-view of $0.28\,\mathrm{deg}^2$. The telescope housing and mirrors were a gift from the National Reconnaissance Office to NASA—the original mission was planned to house a $1.3\,\mathrm{m}$ primary mirror (Hand, 2012). While the final mission became more expensive, the large diameter will allow high-resolution observations of fainter objects in its four near-infrared bands down to magnitudes of $\approx 26.5$. The expected number of detected background galaxies is impressive at $n_\mathrm{gal} \approx 45\,\mathrm{arcmin}^{-2}$. However, the planned survey area for large-scale structure studies will only cover $2200\,\mathrm{deg}^2$ on the sky, yielding an expected 380 million galaxies to measure the cosmic shear and $40\,000$ galaxy clusters to constrain the abundance. The overlapping area with Euclid and LSST will boost the performance of those missions significantly, as it can be used as a powerful calibration tool.

These surveys will revolutionize our understanding of the evolution of the late-time matter distribution due to the immense size of their detected galaxy samples, their high

spatial resolution and the depth to which they can detect galaxies. No current survey comes close: the completed Kilo-Degree Survey (KiDS), carried out by the $2.6\,\mathrm{m}$ VLT Survey Telescope covers an area of $1000\,\mathrm{deg}^2$, but at a lower angular resolution limited by the seeing to $\approx 0.7\,\mathrm{arcsec}$ and a limiting magnitude between $23.5$ in the $i$ filter to $25$ in the $ugr$ filters (Kuijken et al., 2019). Importantly, the whole survey area is additionally covered by five band near-infrared bands using the $4.1\,\mathrm{m}$ Visible and Infrared Survey Telescope for Astronomy (VISTA), which significantly improves the distance measurements to galaxies. The final KiDS sample contains about 21 million of galaxies with accurate distance measurements with $n_{\mathrm{gal}} \approx 6\,\mathrm{arcmin}^{-2}$ (Giblin et al., 2021). The Dark Energy Survey (DES), using the $4\,\mathrm{m}$ Blanco telescope in Chile, probes a larger area of $5000\,\mathrm{deg}^2$, but to a lower depth than the KiDS data (Abbott et al., 2021). The DES sample contains about 390 million galaxies (Sevilla-Noarbe et al., 2021). Finally, the Hyper Suprime-Cam Survey (HSC) has the largest mirror size, observing with the $8.2\,\mathrm{m}$ Subaru telescope, probing an area of $1400\,\mathrm{deg}^2$ down to a deeper limiting magnitude of $\approx 26$ (Aihara et al., 2018). Due to the large collecting area of the telescope, a number of background galaxies $n_{\mathrm{gal}} \approx 20\,\mathrm{arcmin}^{-2}$ can be reached (Mandelbaum et al., 2018). The whole area has not yet been observed, so a final sample size is not available for the HSC survey.

The ambitious goal of future stage IV surveys is to constrain the statistical distribution of matter, quantified by the matter power spectrum, to $1\,\%$ precision, and to pin down the evolving behaviour of dark energy, specifically, the constant and the time-varying equation of state parameters to $2\,\%$ and $10\,\%$, respectively. The gargantuan galaxy samples will provide sufficient statistical power to reach such high precision. However, biases in the analysis can arise if we do not control the systematic uncertainties in the theoretical models we use to analyse the observations to the same level of accuracy (e.g. Hearin et al., 2012; Taylor et al., 2018). Hence, we need to scrutinize our models to identify all possible sources of uncertainty that would bias the data analysis and the inferred cosmological model. In this thesis we will consider the impact of uncertainties in the normal matter distribution for both cosmic shear analyses and galaxy cluster abundance studies.

## 1.2 A brief history of our Universe

Before diving into the cosmologist's mathematical toolkit that allows them to model the evolution of the Universe, a brief status overview of our knowledge of the Universe is in order.

The current standard model, the Lambda Cold Dark Matter ($\Lambda$CDM) cosmology, is an immense achievement both theoretically and observationally: a theory tweaked and polished over the past century. Einstein's theory of general relativity provides the framework to model the evolution of an expanding, isotropic and homogeneous universe composed of ordinary matter contributing a mere $5\,\%$ of the total energy content of the Universe, dark matter ($\approx 25\,\%$), and dark energy ($\approx 70\,\%$), lending its name to the model through the cosmological constant, $\Lambda$ (Planck Collaboration et al., 2020b). Radiation, in terms of photons emitted by gas and stars, and neutrinos generated in radioactive decays, amounts to a negligible fraction of $\approx 0.01\,\%$ of the total energy density of our Universe. Against

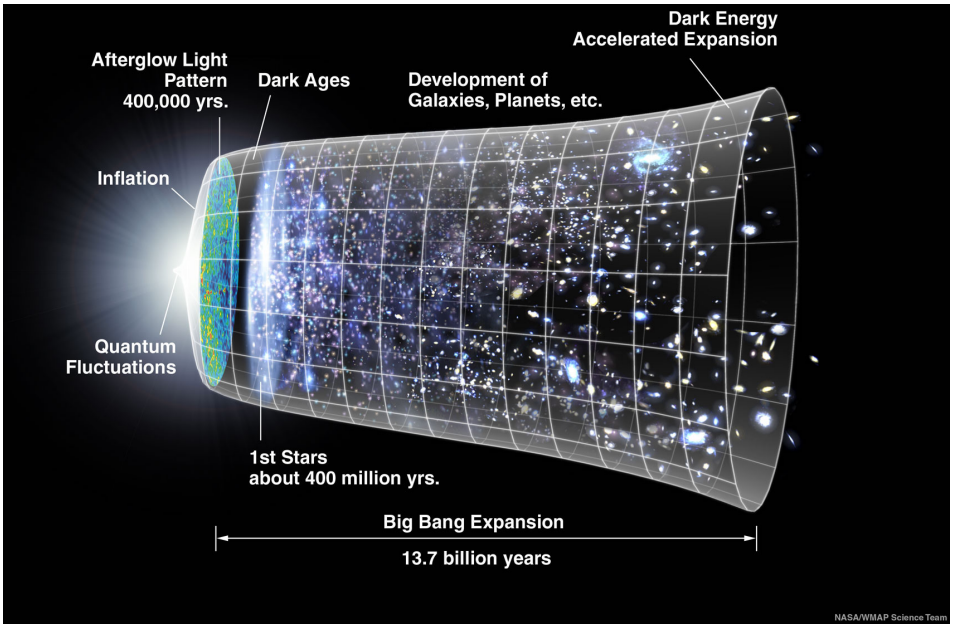


Figure 1.1: Timeline of the Universe from the beginning, on the left, to today, on the right. The vertical size represents the spatial expansion (not to scale). Inflation generates macroscopic density perturbations from the quantum fluctuations present in the beginning of the Universe. The dark matter fluctuations grow as soon as matter dominates the energy content, eventually attracting ordinary matter when it decouples from radiation around 400 000 years after the Big Bang. The ordinary matter collapses, forming stars that make up the galaxies that we see. Over time, these galaxies coalesce into more massive groups and clusters, forming the cosmic web. In the past 3.6 billion years, dark energy has started to dominate the expansion of the Universe, inaugurating a phase of accelerated expansion. *Figure by the NASA/WMAP Science Team.*

this smooth backdrop, tiny density perturbations generated during the early inflationary phase of exponentially fast expansion, will coalesce into ever more massive dark matter overdensities under the influence of gravity, eventually drawing in the ordinary matter, sparking star formation about 400 million years after the Big Bang and forming galaxies. After about 10 billion years, the expansion of the Universe has sufficiently diluted the matter density for dark energy to become the dominant energy density component, toppling the Universe into a supposedly everlasting accelerated expansion. This is where we are now, about 13.8 billion years after the Big Bang. Fig. 1.1 shows an illustration of the expansion history of the Universe with several milestones in the development of structure.

The standard cosmological model has been immensely successful in simultaneously explaining an eclectic range of observations, including the typical size of pressure waves propagating in the primordial plasma imprinted as fluctuations in both the cosmic mi-

crowave background (CMB), approximately 380 000 years after the Big Bang (e.g. Spergel et al., 2003) and the late-time distribution of luminous red galaxies, about 10 billion years after the Big Bang, known as baryon acoustic oscillations (BAO, e.g. Cole et al., 2005; Eisenstein et al., 2005), the absence of a decrease in the velocity with which galaxies encircle massive agglomerations known as galaxy clusters (Zwicky, 1933), and also of stars in the outskirts of galaxies (Rubin et al., 1980) due to an additional mass contribution of a non-directly observable, "dark" matter, the accelerating expansion of the Universe as probed by supernovae at cosmological distances (e.g. Riess et al., 1998; Perlmutter et al., 1999), the clustering of galaxies (e.g. Peacock et al., 2001; Percival et al., 2001), the statistical distortion of observed galaxy shapes due to the weak lensing effect of intervening matter (e.g. Wittman et al., 2000; Van Waerbeke et al., 2000; Bacon et al., 2000; Hoekstra et al., 2002), the abundance of massive clusters of galaxies (e.g. Bahcall & Cen, 1993) and more.

There are two hidden elephants in the room, however. Neither dark matter nor dark energy fit comfortably within the standard model of particle physics. While there are many smoking guns pointing to a consistent dark matter component, dark energy still remains a mystery. The inferred amount of dark matter is required to explain a range of different observations. Peebles (2017) provides an interesting personal and historical overview of how the evidence and the acceptance of cold dark matter has grown over the past century—another fascinating account on the history of cosmology in general is given in Peebles (2012). We provide a small summary: Firstly, the ratio of the acoustic peaks in the statistical distribution of CMB anisotropies indicate an extra gravitational forcing due to a non-baryonic component (e.g. Hu & Dodelson, 2002; Planck Collaboration et al., 2020b)[5]. Ubiquitous already established dark matter structures are also required at the time of the CMB to explain how the tiny observed ordinary matter perturbations were able to quickly collapse and form galaxies by the current time. Additionally, the flat rotation curves of stars in galaxies (e.g. Rubin & Ford, 1970; Rubin et al., 1980) and the lack of decreasing velocity of galaxies toward the outskirts of clusters indicate the need for an additional invisible mass (e.g. Zwicky, 1933). Finally, gravitational lensing observations of colliding clusters indicate that most of the (dark) mass stays confined to the cluster while the hot gas is left behind in the collision, as observed in the Bullet Cluster (Clowe et al., 2006). However, so far no dark matter particle candidates have been found in any direct searches (e.g. Undagoitia & Rauch, 2016).

Dark energy, on the other hand, remains elusive. From quantum field theory, we expect the vacuum of space to have an associated energy, but the measured cosmological constant comes out about 120 orders of magnitude lower than the predicted value (Weinberg, 1989). However, the energy density of the cosmological constant inferred from observations simultaneously explains both the flatness of the Universe and the late-time accelerating expansion (for a review, see Frieman et al., 2008). Many different theoretical dark energy models exist, ranging from scalar fields with negative pressure to modifications to the laws of gravity (e.g. Joyce et al., 2016). So far, however, the data do not show a preference for more complicated models over a standard cosmological constant (e.g. Chiba et al., 2013; Wang et al., 2018; Gerardi et al., 2019).

[5]Clear animated plots can be found on Wayne Hu's personal website

Now that late-time observational probes of the cosmological expansion history and the matter distribution are gaining statistical power, another potential problem is appearing for ΛCDM in the guise of tensions between cosmological parameters inferred from the CMB and late-time probes (for a recent overview, see Abdalla et al., 2022). Most famously, the current value of the Hubble parameter inferred from distances to supernovae in galaxies flowing along with the expanding Universe, is larger and in a $5\,\sigma$ tension with that extrapolated from the CMB—a discrepancy known as the Hubble tension (Planck Collaboration et al., 2020b; Riess et al., 2021). Additionally, late-time probes of the matter distribution consistently find values of the matter clustering between $2-3\,\sigma$ lower than predicted by the CMB—dubbed the $S_8$ tension (e.g. Bocquet et al., 2019; Heymans et al., 2021). There are two possible ways to solve these problems: either there are unidentified sources of systematic error in at least one of the analysis methods or there are unaccounted for physical effects present in the data—colloquially, we speak of "new physics".

For Planck, there seem to be inconsistencies in the data between measurements when using only small or large scales (e.g. Addison et al., 2016), partially manifesting as a smoothing of the small scale peaks, similar to the signal expected from a stronger lensing of the CMB (Planck Collaboration et al., 2017). However, these features are stronger than expected from direct constraints inferred from the higher-order correlations that the lensing induces in the anisotropies of the CMB (e.g. Obied et al., 2017; Motloch & Hu, 2018, 2020). Tantalizingly, excluding the small-scale information shifts the best-fitting values of both the Hubble parameter and the matter clustering closer to the late-time measurements (Planck Collaboration et al., 2017). However, given the high dimensionality of the parameter space, Planck Collaboration et al. (2017) finds shifts in the cosmological parameters with a similar significance in about $10\,\%$ of their simulated data, consistent with statistical fluctuations. The inconsistencies could also point to some unknown systematic effects in the Planck analysis, or to unaccounted for physical processes.

The supernovae analyses calibrate distances to nearby galaxies that have hosted a supernova explosion, using a standard candle, such as Cepheid variable stars (e.g. Riess et al., 2021). This method has several requirements. Firstly, it requires an accurate calibration of the standard candle, for example, the Cepheid period–luminosity relation, extrapolated to more distant, SN-hosting galaxies. Additionally, geometric distance calibrations to local galaxies are needed. Finally, the decay in the supernova brightness must follow a universal, redshift-independent relation. Riess et al. (2021) carefully investigate and dispel most currently hypothesized sources of systematic error for Cepheid distance calibrations, such as biases induced by crowding of background stars in the Cepheid photometry, inconsistencies in the period–luminosity relation for different galaxies, or the impact of using different nearby anchors for the distance ladder (as detailed in the responses to Efstathiou 2020, and also in the discussion of Riess et al. 2021). However, the complex analysis, the possible astrophysical sources of systematic uncertainty and the fact that different distance calibrations, using, for example, the tip of the red giant branch (e.g. Freedman et al., 2019; Freedman, 2021; Anand et al., 2022), or lensing time delays (e.g. Birrer et al., 2020), result in parameter determinations consistent with the CMB prediction, should serve as a caution before concluding that a modification to ΛCDM is needed. In a few years, a completely independent measurement using standard gravitational sirens will be able to shed light on this issue (e.g. The LIGO Scientific Collaboration and The

Virgo Collaboration et al., 2017; Feeney et al., 2019; Soares-Santos et al., 2019).

Finally, the most significant discrepancy in the inferred matter clustering compared to the CMB data is found in cosmic shear analyses (e.g. Asgari et al., 2021; Amon et al., 2022b,a). Inferring cosmological parameters from measuring the statistical shape distortion of galaxies caused by the intervening matter is a complicated problem beset by possible systematic uncertainties. First of all, for a fixed lensing signal over or underestimating the distance to the source galaxies results in higher or lower values for the inferred clustering of matter, respectively (Joudaki et al., 2020). Additionally, intrinsic alignments of galaxies in each other's neighbourhood can masquerade as a cosmological signal (e.g. Croft & Metzler, 2000; Heavens et al., 2000; Hirata & Seljak, 2004). The details of this alignment signal will depend on galaxy formation processes and the galaxy population, but current models are likely too simplistic to capture the full behaviour (e.g. Joachimi et al., 2015). Different analysis methods and different surveys find significantly different values for the strength of the intrinsic alignment signal (e.g. Efstathiou & Lemos, 2018; Asgari et al., 2021), which could indicate that the parameter is accounting for another systematic uncertainty in the data. Calibrating the intrinsic alignment signal observationally, as suggested by Fortuna et al. (2021), could reduce this effect. However, tests of the internal consistency of the observations and variations in the systematic uncertainty modelling have not identified any significant possible reductions in the tension which could point to the need for modifications to the cosmological model (e.g. Asgari et al., 2021; Joachimi et al., 2021; Amon et al., 2022b).

As the saying goes: "May you live in interesting times", and we certainly do. The advent of future surveys aims to constrain the expansion history and the matter content of the Universe with revolutionary precision. Any inconsistencies in the standard model of cosmology will be tremendously magnified with the future data, that is, if we are able to adequately control the systematic uncertainties in the challenging data analysis.

## 1.3 The basics of cosmology

To understand how weak lensing can probe the matter distribution of the Universe, we first need to introduce the tools of modern cosmology. We start by introducing how we can model the smooth background evolution of the average density of the Universe in Section 1.3.1, then, we will detail in Section 1.3.2 how small density fluctuations generate structure on top of this smooth background, providing the seed locations to form galaxies. Finally, we will describe how baryons condense into the galaxies that we observe in Section 1.3.3.

### 1.3.1 Smooth background evolution

The current cosmological framework rests on the foundational assumption that our Universe is homogeneous and isotropic when averaged over sufficiently large length scales, also known as the cosmological principle. The distribution of galaxies and quasars on scales $\gtrsim 100\,\mathrm{Mpc}$ (about 330 million lightyears) provides evidence for homogeneity (e.g. Hogg et al., 2005; Laurent et al., 2016; Ntelis et al., 2017), and the remarkable isotropy of

the cosmic microwave background is taken as strong evidence for the latter. The maturing of cosmology as a scientific subject started with the development of Einstein's theory of general relativity (Einstein, 1916). In its full glory, the theory of general relativity involves solving the highly non-linear set of Einstein field equations relating the local curvature of four dimensional space-time to the matter and energy density. In compact index notation, the field equations can be written as

$$R_{\mu\nu} + \left(\Lambda - \frac{1}{2}R\right) g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \,, \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci curvature tensor, $R$ is the scalar curvature, i.e. the trace of the Ricci tensor, $\Lambda$ is the cosmological constant, $g_{\mu\nu}$ is the metric tensor, $G$ is Newton's gravitational constant, $c$ is the speed of light, and, finally, $T_{\mu\nu}$ is the stress-energy tensor. The different Greek subscript letters can be any of either the three spatial dimensions or the single time dimension, resulting in a set of $4 \times 4 = 16$ equations. The fundamental geometric quantity in general relativity is the metric, $g_{\mu\nu}$, since the Ricci tensor can be derived from it. The metric acts as a ruler, allowing distances in a curved space to be uniquely defined, independently of the assumed coordinate system, in accordance with the relativity principle promoted by Einstein. Physical properties of the matter enter through the stress-energy tensor.

The cosmological principle greatly simplifies the Einstein field equations. Under the assumption of homogeneity and isotropy, the space-time line element, defined by the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, can be written as

$$\begin{aligned} \mathrm{d}s^2 &= g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu \\ &= -c^2\mathrm{d}t^2 + a^2(t)\mathrm{d}\boldsymbol{\chi}^2 \\ &= -c^2\mathrm{d}t^2 + a^2(t)\left(\mathrm{d}\chi^2 + f_K(\chi)(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2)\right) \,, \end{aligned} \tag{1.2}$$

where we have introduced standard spherical coordinates, $(\chi, \theta, \phi)$, the scale factor, $a(t)$, and

$$f_K(\chi) = \begin{cases} 1/\sqrt{K}\sin(\sqrt{K}\chi) & K > 0 \\ \chi & K = 0 \\ 1/\sqrt{-K}\sinh(\sqrt{-K}\chi) & K < 0 \end{cases} \tag{1.3}$$

introduces the dependence on the spatial curvature, $K$, which is measured to be indistinguishable from 0 (Planck Collaboration et al., 2020b). The FLRW metric implies a Universe consisting of flat, three-dimensional hypersurfaces with an increasing physical distance $\mathrm{d}\mathbf{l} = a(t)\mathrm{d}\boldsymbol{\chi}$ between observers that move along with its expansion. The comoving distance, $\boldsymbol{\chi}$, provides a spatial distance measure that is unaffected by the expansion, and the scale factor, $a(t)$, which can be freely set to be 1 at our current time, functions as a clock, measuring time in terms of how much the Universe has expanded. In the beginning, $t = 0$, the FLRW metric implies a singularity, $a = 0$, where all matter was packed together so densely that our known laws of physics no longer hold. This initial hot and dense state of the Universe is known as the Big Bang. Since the expansion of space also stretches the wavelength of individual photons, another way of measuring time is by

identifying the observed wavelength, $\lambda_{\rm obs}$, of a known atomic emission line at wavelength $\lambda_{\rm em}$. The redshift, $z$, of the emission lines tells us how much smaller the Universe was at the time of emission from the relation

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm em}} = 1 + z = \frac{a(t_{\rm obs})}{a(t_{\rm em})} \,. \tag{1.4}$$

Since practically all the information from the distant Universe reaches us through photons (neglecting gravitational waves and neutrinos), the redshift is the best tool to correctly situate observed objects or events in the history of the Universe.

The finite speed of light allows us to look back in time, with photons leaving galaxies in the very early Universe travelling for billions of years before finally reaching our detectors. The finite speed of photons also results in a cosmological "horizon", the region from which light is able to reach us in the age of the Universe. Since photons travel along null geodesics, $\mathrm{d}s^2 = 0$, the metric in Eq. (1.2) gives the comoving horizon size,

$$\chi_{\rm h}(t) = c \int_0^t \frac{\mathrm{d}t'}{a(t')} \,, \tag{1.5}$$

which at the current time equals about $14\,\mathrm{Gpc}$.

Now, we only need Einstein's field equations to solve for the scale factor and to capture the expansion history of the Universe (e.g. Weinberg, 1972). Friedmann (1922) was the first to show that the scale factor in the FLRW metric must obey the relation

$$H^2(t) \equiv \left(\frac{\dot{a}(t)}{a(t)}\right)^2 = \frac{8\pi G}{3}\rho(t) - \frac{Kc^2}{a^2(t)} \tag{1.6}$$

also known as the Friedmann equation, which follows from the space-space and time-time components of Einstein's equations. We have introduced the common notation $\mathrm{d}/\mathrm{d}t \equiv \dot{}$ and the density,

$$\rho(t) = \rho_{\rm m}(t) + \rho_\gamma(t) + \rho_\Lambda \,, \tag{1.7}$$

consisting of a matter (m), radiation ($\gamma$), and cosmological constant ($\Lambda$) contribution. Additionally, the time-time component of Einstein's equations gives

$$\frac{\ddot{a}(t)}{a(t)} = -\frac{4\pi G}{3}\left(\rho(t) + \frac{3p(t)}{c^2}\right) \,, \tag{1.8}$$

where the pressure,

$$p(t) = \sum_i w_i \rho_i(t) \,, \tag{1.9}$$

depends on the equation of state parameter, $w_i$, of each density component. Eqs. (1.6) and (1.8) explicitly show how the expansion of the Universe depends on the physical properties, that is the density and the pressure, of its constituent matter and energy components. Defining Eq. (1.6) in terms of the critical density for which the Universe is flat, that is, for which space is Euclidean,

$$\rho_{\rm crit}(t) = \frac{3H^2(t)}{8\pi G} \,, \tag{1.10}$$

we obtain the most-widely used form for a flat universe,

$$H^2(z) = H_0^2(\Omega_{\mathrm{m},0}(1+z)^3 + \Omega_{\gamma,0}(1+z)^4 + \Omega_{\Lambda,0})\,, \tag{1.11}$$

where $\Omega_i \equiv \rho_i/\rho_{\mathrm{crit}}$ and the subscript 0 indicates that the quantity is measured at the current time $t_0$, or, identically, $a = 1$ and $z = 0$. Observationally, the parameters are constrained to $H_0 = 67.44 \pm 0.42\,\mathrm{km\,s^{-1}\,Mpc}$ and $(\Omega_{\mathrm{m},0}, \Omega_{\Lambda,0}) = (0.311, 0.689) \pm 0.006$, with $\Omega_{\gamma,0} < 10^{-4}$ (Planck Collaboration et al., 2020b). Here, we have used Eq. (1.8) with the equation of state parameters for a photon gas, $w_\gamma = 1/3$, pressureless dark and ordinary matter, $w_{\mathrm{m}} = 0$, and dark energy in the form of a cosmological constant, $w_\Lambda = -1$. The radiation density decreases strongly with time since the spatial expansion decreases the number density of photons and additionally stretches their wavelength, reducing the total energy density. Matter, on the other hand, is only affected by the decreasing number density. Finally, the cosmological constant permeates space with a fixed background density, unaffected by the expansion of the Universe.

According to the most recent Planck Collaboration et al. (2020b) measurements of the temperature fluctuations imprinted in the cosmic microwave background (CMB), the Universe is currently approximately 13.8 billion years old. It has passed through different epochs in which different density components dominate its expansion history. The early Universe was dominated by radiation, until about 20 000 years after the Big Bang (or $z \approx 5700$) when matter took over. Interestingly, about 3.6 billion years ago (or $z \approx 0.3$)—recent history, cosmologically speaking—dark energy started dominating over matter, initiating a phase of seemingly never-ending, accelerated expansion. However, until we fully understand what causes the observed acceleration, that is, until we have a satisfactory model of dark energy or modified gravity that is observationally preferred, reports about the of final fate of the Universe are greatly exaggerated.

A final important consequence of our expanding Universe for observational cosmologists is that measuring distances to objects is more complex. In the neighbourhood of our own galaxy, we can infer distances, $D$, to objects from their flux, $F_{\mathrm{tot}}$, if we know their total bolometric luminosity, $L_{\mathrm{tot}}$, using the fact that

$$D_L = \sqrt{\frac{L_{\mathrm{tot}}}{4\pi F_{\mathrm{tot}}}}\,. \tag{1.12}$$

Alternatively, if we know some physical size, $\delta l$, we can infer the distance from the angle, $\delta\theta$, it subtends on the sky since

$$D_\theta = \frac{\delta l}{\delta\theta}\,. \tag{1.13}$$

In a flat, expanding Universe, the expansion of space need to be taken into account when inferring distances in a similar way. It is useful to first define the line-of-sight comoving distance, $\chi$, between us, the observers, and a point located at a particular redshift, $z$, as

$$\chi(z) = \int_0^z \mathrm{d}z \frac{c}{H(z)}\,, \tag{1.14}$$

which follows from the metric in Eq. (1.2). This comoving distance defines the spherical surface over which photons emitted at redshift $z$ will be spread out by $z = 0$. Additionally,

photons are redshifted by the time they reach the detector, decreasing the energy and increasing the arrival time interval, resulting in a measured flux

$$F_{\mathrm{tot,obs}} = \frac{L_{\mathrm{tot}}}{4\pi\chi^2(z)(1+z)^2} \tag{1.15}$$

from which we get the luminosity distance

$$D_L = (1+z)\chi(z)\,. \tag{1.16}$$

The angular diameter distance, which is the generalization of Eq. (1.13), is easily derived from the metric, Eq. (1.2). The proper size, $\delta l$, of an object or feature at redshift $z$, a comoving distance $\chi$ from an observer at the origin, fixes the angular size, $\delta\theta$, seen by the observer to

$$\delta l^2 = a^2(t_{\mathrm{em}}) f_K^2(\chi)\delta\theta^2\,, \tag{1.17}$$

giving

$$D_A = \frac{f_K(\chi)}{1+z}\,. \tag{1.18}$$

A fascinating property of our expanding Universe is that objects of fixed physical size, $\delta l$, will reach a minimum angular size between $z \simeq 1-2$, but for higher redshifts their angular sizes actually become larger because the physical distance to the observer was significantly smaller at the time of emission, resulting in a larger angular size.

Being able to convert angles on the sky or luminosities on a detector into distances, requires so-called standard rulers or standard candles. With a ruler or a candle in hand, observational cosmologists are able to constrain the expansion history and the geometry of the Universe through the inferred comoving distance, $\chi(z)$, to said ruler or candle, which depends on the Hubble parameter, $H(z)$. Some of the strongest current constraints on the properties of our Universe were derived this way. Anisotropies in the cosmic microwave background provide a fossil record of the sound horizon, the maximum distance that pressure waves were able to travel since the start of the Universe, at the time of recombination of protons and electrons around 380 000 years ($z \approx 1100$) after the Big Bang. The physical size of the sound horizon depends on the matter-to-photon ratio, due to the balance between gravitational collapse and radiation pressure that generate the oscillation responsible for the pressure waves (e.g. Peebles & Yu, 1970; Hu et al., 1997). The angular scale of these perturbations has been measured to be $0.59643\,\mathrm{deg}$ (about the angular size of the Sun on the sky) with an exquisite accuracy of $0.05\,\%$ by the Planck satellite (Planck Collaboration et al., 2020b). The angular size additionally depends on the expansion history of the Universe through $\chi(z)$, as can be seen from Eq. (1.17), making the CMB a powerful cosmological probe. Since these perturbations are also imprinted—although strongly damped—in the baryons which collapse under gravity to form galaxies, another standard ruler can be found in the statistical distribution of galaxies at much later times. This baryon acoustic oscillation (BAO) feature was first detected by Eisenstein et al. (2005) and Cole et al. (2005). Finally, type Ia supernova explosions can function as standard candles, since they are generated when a white dwarf reaches the Chandrasekhar limit of $\approx 1.44\,\mathrm{M}_\odot$ through mass accretion from a companion, which results in a fixed

luminosity explosion event. These standard candles have been used by Riess et al. (1998) and Perlmutter et al. (1999) to constrain the distance–redshift relation out to $z \approx 1$ and show that the Universe is entering a phase of accelerated expansion.



### 1.3.2 Growth of structure

Even though the modelling of the evolution of the average density of the entire Universe is an impressive feat, it does not bring us closer to understanding how the observed galaxies form and cluster together in the cosmic web and how we can predict the lensing signal that this matter distribution generates. For that, we need to know the initial distribution of density perturbations and their evolution in time.

The stage for galaxy formation is set extremely early when quantum fluctuations get inflated to scales orders of magnitude larger than the causal horizon—the region where particles have been able to communicate—in a phase of exponential expansion, known as inflation (Guth, 1981; Linde, 1982). Inflation ends in a reheating period that spawns the standard model particles and initiates the radiation-dominated period of our Universe. The inflationary scenario provides an explanation for the counter-intuitive observed isotropy of the CMB: patches on opposite parts of the sky that cannot possibly be in thermal equilibrium, are observed to have the same temperature to within a fractional difference of $\Delta T/T \approx 10^{-5}$. Inflation explains this isotropy since these regions were actually in causal contact before they were stretched outside of each other's horizons, meaning they were in thermal equilibrium before inflation. Additionally, inflation also smooths out any initial curvature present in the early Universe, explaining why the current Universe is observed to be flat even though any tiny deviation from flatness exacerbates with time.

Due to the stochastic nature of the initial quantum fluctuations, we can only ever describe the spatial distribution of density perturbations in a statistical sense. Standard inflationary models predict that the initial density perturbation compared to the mean density, $\bar{\rho}$,

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t)}{\bar{\rho}(t)} - 1\,, \tag{1.19}$$

is a zero-mean Gaussian random field with correlations on length scales $\lambda = 2\pi/k$ described by a power spectrum,

$$P(k) = \langle |\delta_k|^2 \rangle \propto k^{n_\mathrm{s}}\,, \tag{1.20}$$

that is almost scale-free, i.e. $n_\mathrm{s} \approx 1$. Here, scale-free means that all density perturbations have the same amplitude when they re-enter the horizon after inflation has ended and radiation dominates the energy density (e.g. Bardeen et al., 1983). As a zero-mean Gaussian random field, the initial density perturbations are fully specified by the power spectrum since all even, non-zero higher-order moments can be derived from it.

The initial growth of density perturbations depends on the dominant energy component of the Universe. Perturbations that re-enter the horizon during the radiation-dominated epoch will stall their growth until matter becomes the dominant component and dark matter perturbations start to collapse and establish the potential wells in which galaxies will eventually form. This stalled growth imprints the comoving horizon size at

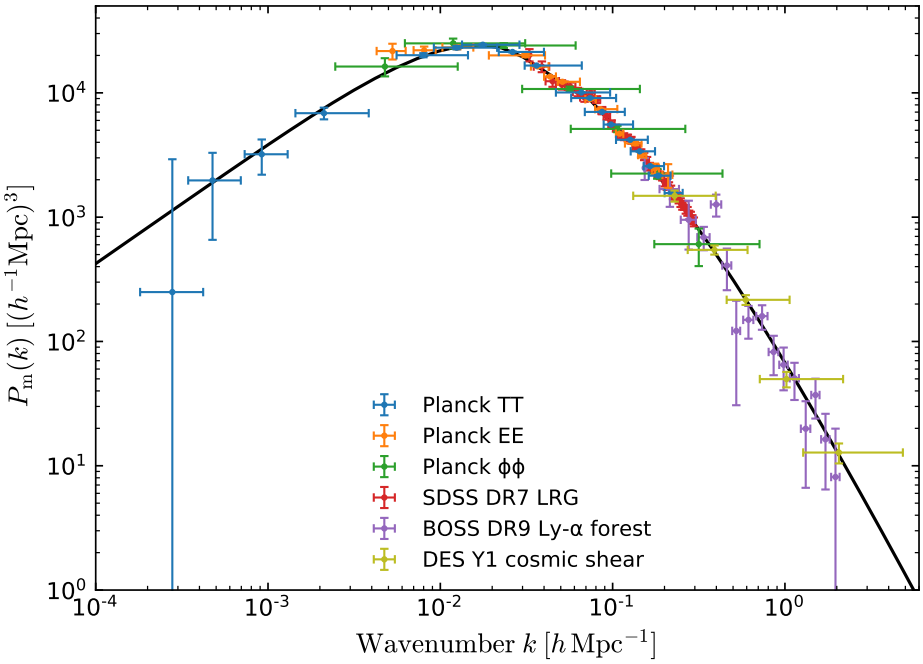


Figure 1.2: The linear matter power spectrum at $z = 0$ inferred from different observational probes. The Planck satellite measures the temperature fluctuations (TT) and the polarization (EE) of the cosmic microwave background at $z \approx 1100$, but also how it is lensed by the intermediate large-scale structure around $z \simeq 0.5 - 1$ ($\phi\phi$). The Sloan Digital Sky Survey (SDSS) measures the clustering of luminous red galaxies (LRGs) with $z < 0.5$. The Baryon Acoustic Oscillation Spectroscopic Survey (BOSS) Ly-$\alpha$ forest probes Ly-$\alpha$ absorption features in spectra of quasars that closely trace the underlying matter distribution. Finally, cosmic shear uses the weak gravitational lensing shape distortion of large samples of galaxies to constrain the statistical matter distribution between the observer and the galaxies. *Figure taken from Planck Collaboration et al. (2020a).*

matter–radiation equality, $d_{\rm H,eq} \approx 100\,{\rm Mpc}$ or $k_{\rm eq} \approx 0.01\,{\rm Mpc}^{-1}$, as a characteristic scale in the power spectrum of small fluctuations, $\delta < 1$, also known as the linear power spectrum. Fig. 1.2 shows observational constraints on the linear matter power spectrum inferred from the CMB, galaxy clustering, the Ly-$\alpha$ forest absorption feature in quasar spectra, and cosmic shear, together with the linear power spectrum predicted from the best-fit Planck Collaboration et al. (2020b) constraints. On large scales ($k < k_{\rm eq}$), the power spectrum retains its initial $\propto k$ dependence, whereas on small scales ($k > k_{\rm eq}$), the scaling changes to a $\propto k^{-3}$ dependence since smaller scales entered the horizon earlier and have been stalled for longer.

The seed locations of structure formation can be determined by following the growth of small, linear dark matter perturbations on top of the smooth background density of

the Universe. The growth of subhorizon dark matter perturbations in a matter-dominated Universe is captured by the simplified, Newtonian fluid equation

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi G\bar{\rho}\delta\,, \tag{1.21}$$

with $\delta \ll 1$. Here, the right-hand side exerts a gravitational pull, increasing the overdensity, whereas the expansion of the Universe, represented by the Hubble parameter term on the left-hand side, opposes this growth, decreasing the perturbation growth rate from an exponential, in the case of no expansion ($H = 0$), to a power-law, $\delta \propto a$, in the matter-dominated regime. Hence, even in the presence of the expansion of the Universe, dark matter perturbations will grow, eventually reaching $\delta \approx 1$, where the linear approximation no longer holds and higher-order terms in $\delta$ need to be included in Eq. (1.21).

The evolution of non-linear overdensities can be modelled with a simplified spherical collapse model, treating spherical overdensities as their own separate universes with higher density (e.g. Peebles, 1965, 1967). This calculation shows that overdensities first decouple from the Hubble flow and start to collapse when they are about 4 times denser than the average density of the Universe. The resulting haloes are about 200 times more dense than the average background density at their formation time. From these models, we know that this collapse occurs when the linear density perturbations in Eq. (1.21) reach a critical value of $\delta_{\rm c} \approx 1.686$. Non-linearities become important on scales $\lambda \lesssim 10\,{\rm Mpc}$, or, equivalently, $k \gtrsim 0.1\,{\rm Mpc}^{-1}$, where the total matter power spectrum will start deviating from the linear matter power spectrum shown in Fig. 1.2.

Press & Schechter (1974) used similar analytic arguments to predict the distribution of collapsed objects consisting of cold gas—dark matter was not yet considered in the cosmological models back then—as a function of their mass. They argued, following the spherical collapse model, that when linear perturbations on a scale $R$ reach the threshold for collapse, they will form a structure of mass $m \propto \bar{\rho}R^3$. Assuming that the probability $P(\delta_R > \delta_{\rm crit})$ determines the fraction of the mass of the Universe present in objects of mass $> m$, the resulting halo distribution is fully determined by the possible cosmology dependence of the critical density for collapse, $\delta_{\rm crit}$, and the variance of the overdensity field. Impressively, their prediction based on linear theory was later shown to qualitatively reproduce the mass-dependent abundance of haloes in simulations (e.g. Efstathiou et al., 1988).

Our understanding of how structure forms was furthered by studying the clustering of the most significant peaks in the density field: clusters of galaxies. Kaiser (1984) noted that the larger correlation lengths observed for clusters compared to lower-mass galaxies could be explained if clusters are a biased tracer of the underlying density field. That is, clusters only form when the local density field is coherently boosted by a large-wavelength perturbation allowing large-scale overdensities to become significant and form massive haloes with correlations on the lengthscale of the wavelength. Smaller fluctuations behave more like noise and small patches can reach significant overdensities, resulting in the formation of low-mass haloes with smaller coherence lengths. The analytic population properties of the peaks of random Gaussian fields have been worked out meticulously by Bardeen et al. (1986).

Studying the details of halo formation requires N-body simulations that model the gravitational evolution of collisionless matter particles in an expanding Universe. The

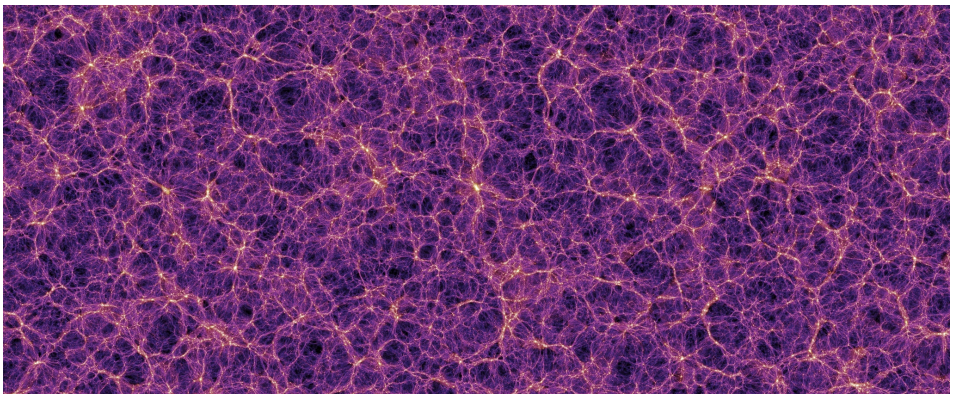

Figure 1.3: The cosmic dark matter distribution at $z = 0$ predicted by the Millennium simulation of Springel et al. (2005) in a $15\,h^{-1}\,\mathrm{Mpc}$ thick slice of the simulated volume. The cosmic web with dense, yellow nodes and connective filamentary structures surrounding empty voids is clearly visible. Galaxies predominantly form in haloes along the filaments, clustering in groups and clusters, which are located at the nodes of the cosmic web. *Figure by the Virgo consortium.*

first computer simulations in a cosmological context, carried out by Aarseth (1963), only followed the gravitational evolution of 100 collisionless particles, limited by the available computational power and the $N^2$ scaling of the direct force calculation between all the particles. In the early 80s, Efstathiou & Eastwood (1981) carried out the first cosmological simulation with 20 000 particles, using the more advanced particle–particle/particle–mesh (abbreviated as $\mathrm{P^3M}$) method, introduced by Eastwood (1975). The $\mathrm{P^3M}$ method reduces the computational expense of N-body simulations by introducing a mesh with $M^3$ cells over which the average gravitational potential is calculated. Forces on individual particles receive a large-scale contribution from the mesh and a short-range particle–particle contribution that only includes particles within a fixed, smaller distance from each other, saving a significant amount of computation and enabling much larger ensembles of particles to be modelled. Current simulation codes have upgraded their gravity calculations to a TreePM method that calculates short-range forces more efficiently using a tree structure while keeping the mesh for large scales (e.g. Barnes & Hut, 1986; Xu, 1995).

In the 90s, Navarro et al. (1996, 1997) showed that haloes in simulations modelling solely the evolution of dark matter in universes with different initial power spectra follow a universal density profile, now known as the Navarro-Frenk-White (NFW) profile, characterized by

$$\rho_{\mathrm{NFW}}(r) \propto \left(\frac{r}{r_\mathrm{s}}\right)^{-1} \left(1 + \frac{r}{r_\mathrm{s}}\right)^{-2}, \tag{1.22}$$

where $r_\mathrm{s}$ is the scale radius where the power-law slope $\mathrm{d}\ln\rho_{\mathrm{NFW}}/\mathrm{d}\ln r = -2$, transitioning between the $r^{-1}$ and $r^{-3}$ scaling of the inner and outer halo, respectively.

From the early 2000s to now, an increase in the computational power of supercomput-

ers has lead to ever larger simulations being run with increasing numbers of particles and for many different cosmological models. Springel et al. (2005) introduced the Millennium simulation, evolving over 10 billion dark matter particles, each weighing $\approx 10^9\,h^{-1}\,\mathrm{M}_\odot$, in a comoving volume of $500\,h^{-1}\,\mathrm{Mpc}$, where $h = H_0/100\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Such simulations aim to self-consistently model the full cosmological evolution of a large, representative patch of the Universe. They start at high redshift, well before the first galaxies formed, usually $z > 100$—less than $\approx 15$ million years after the Big Bang—with a particle distribution that follows the statistical distribution predicted for a chosen cosmology. Evolving these particles through time, including the effects of gravity and the expansion of the Universe, we end up with accurate predictions for the structure of the Universe, including the cosmic web and highly non-linear haloes, as shown in Fig. 1.3. Many features of the large-scale clustering of matter can be accurately inferred from such simulations since the ordinary matter—which is *not* included in these dark matter-only simulations—approximately follows the dark matter.

To predict the cosmology-dependence of quantities such as the halo abundance or matter clustering from simulations, we need a sample of simulations that vary the cosmological parameters and the resulting matter distribution and evolution. Heitmann et al. (2006) and Habib et al. (2007) suggested to run a set of simulations with cosmological parameters sampled on a grid that maximizes the minimum distance between all the parameter vectors. This enables an efficient interpolation between the results with sufficient accuracy while also minimizing the computational expense since the sampling can be relatively sparse. Such *emulator* approaches are gaining popularity owing to the stringent theoretical accuracy requirements of future galaxy surveys. Emulators have been developed for the halo abundance (e.g. McClintock et al., 2019; Nishimichi et al., 2019; Bocquet et al., 2020) and the matter power spectrum (e.g. Heitmann et al., 2009; Euclid Collaboration et al., 2018), both quantities of interest for this thesis. However, simulations can only be fully representative of our actual Universe with its stars, gas and galaxies, when they also include baryons and their associated physical processes, as we will discuss next.

### 1.3.3 Galaxy formation

The formation of galaxies is the next step after dark matter structures have collapsed. Dark matter has an approximately 350 000 year head start, being able to collapse as soon as matter dominates the energy budget of the Universe. Normal matter, on the other hand, only decouples from radiation when the Universe has cooled down to $T_\mathrm{dec} \approx 3000\,\mathrm{K}$ (at $z_\mathrm{dec} \approx 1100$) and photon collisions can no longer keep the primordial gas ionized. At this stage, the gas becomes neutral and can collapse into the already established dark matter potential wells.

The crucial difference between ordinary and dark matter is that ordinary matter takes part in electromagnetic interactions. As a result, gas particles can collide with each other, exerting a pressure that can balance gravitational collapse. Hence, collapsed dark matter haloes form a gaseous halo with an equilibrium, virial temperature and density profile set by the mass of the halo. This is not the full story, however: if the density and temperature are sufficiently high, the gas becomes ionized and collisions between electrons and atoms can excite electronic transitions or even ionize more atoms, resulting in radiative

cooling (which is more efficient for higher-metallicity gas due to the availability of more electrons). Gas at temperatures comparable to the ionization energy of the atomic lines, mostly between $10^4 - 10^6$ K, will be able to cool efficiently through these processes, fragmenting to form stars in low-mass haloes (e.g. Rees & Ostriker, 1977; Silk, 1977; White & Rees, 1978). The small initial angular momentum of the gas clouds, imparted by gravitational tidal torques from the surrounding large-scale structure, gets amplified in the collapse, resulting in the formation of a disk (e.g. Fall & Efstathiou, 1980). Subsequent halo mergers grow the total halo and stellar mass, possibly disturbing the disk, resulting in the formation of elliptical galaxies. The emerging picture is one of hierarchical growth of haloes, with the central galaxy mass increasing due to continuous accretion and mergers, and with large haloes accreting smaller haloes and their galaxies as a satellite population.

A successful galaxy formation model should be able to reproduce the observed abundance of galaxies as a function of properties such as their mass, colours, sizes, and their observed clustering. To self-consistently model the growth of dark matter haloes from their initial perturbations as well as the formation of galaxies, we need to resort to cosmological hydrodynamical simulations that simultaneously evolve dark matter and gas while accounting for the formation of stars and their associated feedback such as stellar winds and supernovae explosions (for a review, see Somerville & Davé, 2015). The immense dynamic range of galaxy formation requires a trade-off to be made between the mass resolution and the volume of the simulation. Processes that cannot be resolved due to the limited resolution of the simulations are included as subgrid physics recipes.

First of all, gas in the simulations needs to be able to cool. Hence, gas particles in cosmological hydrodynamical simulations typically track the abundance of the 11 dominant atomic species for cooling and interpolate the pre-calculated cooling and heating rates tabulated as a function of the density, temperature, metallicity and redshift of the gas (e.g. Wiersma et al., 2009a). When the gas becomes dense enough, it should be able to form stars. Gas particles are stochastically converted into stellar particles, representing stellar populations of thousands of stars, when their density exceeds the critical density for gravitational collapse (e.g. Schaye, 2004; Schaye & Dalla Vecchia, 2008). The stellar populations evolve, generating type II supernovae and stellar winds that enrich the surrounding gas (e.g. Wiersma et al., 2009b). The supernovae additionally kick or heat their neighbours, simulating the violent explosions that can shut down the star formation in low-mass galaxies by heating and dispersing the high-density star-forming gas (e.g. Dalla Vecchia & Schaye, 2008, 2012). This stellar feedback is important since it allows simulations to reproduce the observed low-mass slope of the galaxy stellar mass function (e.g. Puchwein & Springel, 2013).

Finally, most massive galaxies are observed to host central supermassive black holes (SMBHs) and hot haloes (e.g. Faber et al., 1997; Magorrian et al., 1998). The supermassive black holes are expected to form early on from the coherent collapse of massive gas clouds and grow through continuous gas accretion and mergers. Gas is funnelled onto the SMBH accretion disk, growing the black hole and fuelling powerful galactic-scale winds and jets that heat the surrounding gas, eventually unbinding it and ejecting metal-enriched gas into the hot halo while suppressing star formation (e.g. Silk & Rees, 1998; Blandford, 1999; Fabian, 1999). This AGN feedback is important to reproduce the high-mass end of the galaxy stellar mass function and also the observed baryon fractions in groups and

clusters of galaxies (e.g. Fabjan et al., 2010; McCarthy et al., 2010; Puchwein & Springel, 2013). In simulations, black hole particles can save up energy that they will eventually redistribute among their neighbours, heating them and suppressing star formation (e.g. Booth & Schaye, 2009).

To ensure that these physically inspired subgrid prescriptions result in realistic galaxy populations, their free parameters need to be calibrated to reproduce a chosen set of observations (see the discussion in Section 2.1 of Schaye et al., 2015). This necessary step does not detract from the predictive power of cosmological, hydrodynamical simulations: the highly non-linear relation between the subgrid model parameters and the simulated galaxy properties means that reproducing one observable does not imply that the simulation correctly predicts non-calibrated galaxy properties—if only. Hence, simulations should always be judged based on how well they can reproduce a wide range of observations. In this thesis, we mainly work with the BAHAMAS simulations, since they have been calibrated to reproduce the large-scale distribution of matter by fitting both the galaxy stellar mass function, ensuring that the abundance of galaxies as a function of mass is accurately reproduced, and the hot gas mass fraction of clusters, meaning that clusters contain the correct amount of baryons (McCarthy et al., 2017).

## 1.4 Weak gravitational lensing

One of the first confirmations of Einstein's theory of general relativity came from the stronger predicted lensing effect of massive objects compared to Newton's theory observed in the 1919 solar eclipse (e.g. Eddington, 1919; Dyson et al., 1920). In the 80s, Tyson et al. (1984) suggested to use this lensing effect to measure the mass of galaxies through the coherent statistical distortion that their mass will induce in the shape of randomly oriented background galaxies.

### 1.4.1 The basics

Weak gravitational lensing relies on the fact that a point mass, $M$, will deflect the path of photons travelling within a closest distance $\boldsymbol{\xi}$, by an angle

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \frac{4GM}{c^2}\frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|^2} \\ &= 12.5\left(\frac{M}{10^{14.5}\,\mathrm{M}_\odot}\right)\left(\frac{1\,\mathrm{Mpc}}{\xi}\right)\,\mathrm{arcsec}\,,\end{aligned} \tag{1.23}$$

where $G$ is the gravitational constant and $c$ the speed of light. The geometry of this deflection is illustrated in Fig. 1.4. This relation holds in the region where the gravitational field is weak, that is, $\xi$ is much larger than the Schwarzschild radius, $r_\mathrm{S} = 2GM/c^2$, of the point mass, and $\hat{\alpha} \ll 1$. In this regime, the apparent change in the source position in the source plane, $\delta\alpha$, for the observer can be written as

$$\delta\alpha = \theta - \beta = \frac{D_\mathrm{ds}}{D_\mathrm{s}}\hat{\alpha}\,, \tag{1.24}$$

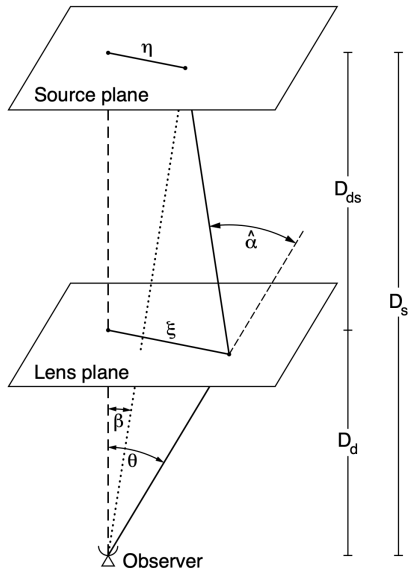


Figure 1.4: The geometry of the deflection of a ray of light by a point mass. The mass, $M$, bends a light ray emitted in the source plane at physical position $\boldsymbol{\eta}$, or angular position $\boldsymbol{\beta}$, by an angle $\hat{\boldsymbol{\alpha}}(\boldsymbol{\xi}, M)$, resulting in an apparent angular position $\boldsymbol{\theta}$ for the observer. In a curved Universe, the angular and physical positions can be related using the angular diameter distances $D_i$, with $i = \mathrm{d}$ indicating the deflector and $i = \mathrm{s}$ the source. *Figure taken from Bartelmann & Schneider (2001).*

where $D_{\mathrm{ds}}$ is the angular diameter distance to the source as seen from the deflector, and $D_{\mathrm{s}}$ is the angular diameter distance from the observer to the source plane.

This principle also holds for mass distributions, such as galaxies or clusters of galaxies, that are located at a fixed redshift, but with a density profile that varies along the line-of-sight. As long as the total deflection along the mass distribution is small, $\sum_i \delta\alpha_i \ll 1$, photons will travel in approximately straight lines and their total deflection is simply the sum of all the thin lens contributions along a fixed line-of-sight, $l$, through the matter distribution

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{4G}{c^2} \int \mathrm{d}^2\boldsymbol{\theta}' \mathrm{d}l \, \frac{D_{\mathrm{ds}} D_{\mathrm{d}}}{D_{\mathrm{s}}} \rho(\boldsymbol{\theta}', l) \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \tag{1.25}$$

$$= \frac{1}{\pi} \int \mathrm{d}^2\boldsymbol{\theta}' \, \kappa(\boldsymbol{\theta}') \frac{\boldsymbol{\theta} - \boldsymbol{\theta}'}{|\boldsymbol{\theta} - \boldsymbol{\theta}'|^2} \,, \tag{1.26}$$

where we have changed to angular coordinates, using $\boldsymbol{\theta} = \boldsymbol{\xi}/D_{\mathrm{d}}$, we have assumed $\delta l \ll D_{\mathrm{d}}$, and we have introduced the convergence,

$$\kappa(\boldsymbol{\theta}) = \frac{\Sigma(\boldsymbol{\theta})}{\Sigma_{\mathrm{crit}}} \,, \tag{1.27}$$

where $\Sigma$ is the surface mass density of the mass distribution along the line-of-sight and

$$\Sigma_{\rm crit} = \frac{c^2}{4\pi G}\frac{D_{\rm s}}{D_{\rm ds}D_{\rm d}} \tag{1.28}$$

is the critical surface mass density. The deflection in Eq. (1.26) can be written as a force sourced by a two-dimensional potential given be the Poisson equation,

$$\nabla^2_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta}) = 2\kappa(\boldsymbol{\theta})\,, \tag{1.29}$$

with

$$\psi(\boldsymbol{\theta}) = \frac{1}{\pi}\int {\rm d}^2\boldsymbol{\theta}'\,\kappa(\boldsymbol{\theta}')\ln|\boldsymbol{\theta}-\boldsymbol{\theta}'|\,, \tag{1.30}$$

from which it is easy to see that

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}\psi(\boldsymbol{\theta})\,. \tag{1.31}$$

For strong gravitational lensing, which occurs when $\kappa > 1$, background galaxies can generate multiple images and these images can get spectacularly distorted, as shown for the galaxy cluster Abell 370 in Fig. 1.5. Abell 370 was one of the first clusters for which these arcs have been detected (Paczynski, 1987). In the weak lensing regime, on the other hand, this deformation is not so obvious: it can only be measured statistically by looking at the coherent distortion of background galaxy shapes equidistant from the cluster centre. The weak lensing distortion can be calculated by considering how much the emitted light ray position changes in the source plane for small changes in the lens plane, that is,

$$\mathcal{A}_{ij} = \frac{\partial\beta_i}{\partial\theta_j} = \delta_{ij} - \psi_{,ij}(\boldsymbol{\theta}) = \begin{pmatrix} 1-\kappa-\gamma_1 & -\gamma_2 \\ -\gamma_2 & 1-\kappa+\gamma_1 \end{pmatrix}_{ij}\,, \tag{1.32}$$

where we identify the $i$th component of the vector $\boldsymbol{x}$ as $x_i$, we have defined $\psi_{,i} \equiv \partial\psi/\partial\theta_i$, $\kappa$ is the convergence from Eq. (1.29), and we have introduced the components of the shear

$$\gamma_1(\boldsymbol{\theta}) = \frac{1}{2}(\psi_{,11}(\boldsymbol{\theta}) - \psi_{,22}(\boldsymbol{\theta})) \tag{1.33}$$

$$\gamma_2(\boldsymbol{\theta}) = \psi_{,12}(\boldsymbol{\theta})\,. \tag{1.34}$$

By diagonalizing the matrix $\mathcal{A}$, it is easy to see that the distortion matrix, $\mathcal{A}^{-1}$, has eigenvalues $\lambda_\pm = (1-\kappa\mp|\gamma|)^{-1}$, where $|\gamma| = \sqrt{\gamma_1^2+\gamma_2^2}$, showing that the convergence magnifies source images in the lens plane, whereas the shear distorts images, magnifying and compressing them along the eigenvectors located at an angle $\phi_+$ and $\phi_-$ from the $\theta_1$-axis, respectively, with $\phi_\pm = \gamma_1/\gamma_2 \mp \sqrt{1+(\gamma_1/\gamma_2)^2}$.

Weak lensing observations are unable to probe the shear and the convergence directly, instead they can only measure galaxy shapes. The coherent shape distortion caused by the intervening mass distribution can be measured under the assumption that the background galaxies are intrinsically randomly oriented following a specific ellipticity distribution. We show an illustration of this effect in Fig. 1.6. Observers measure galaxy ellipticities from the moments of the light distribution (e.g. Blandford et al., 1991; Bartelmann &

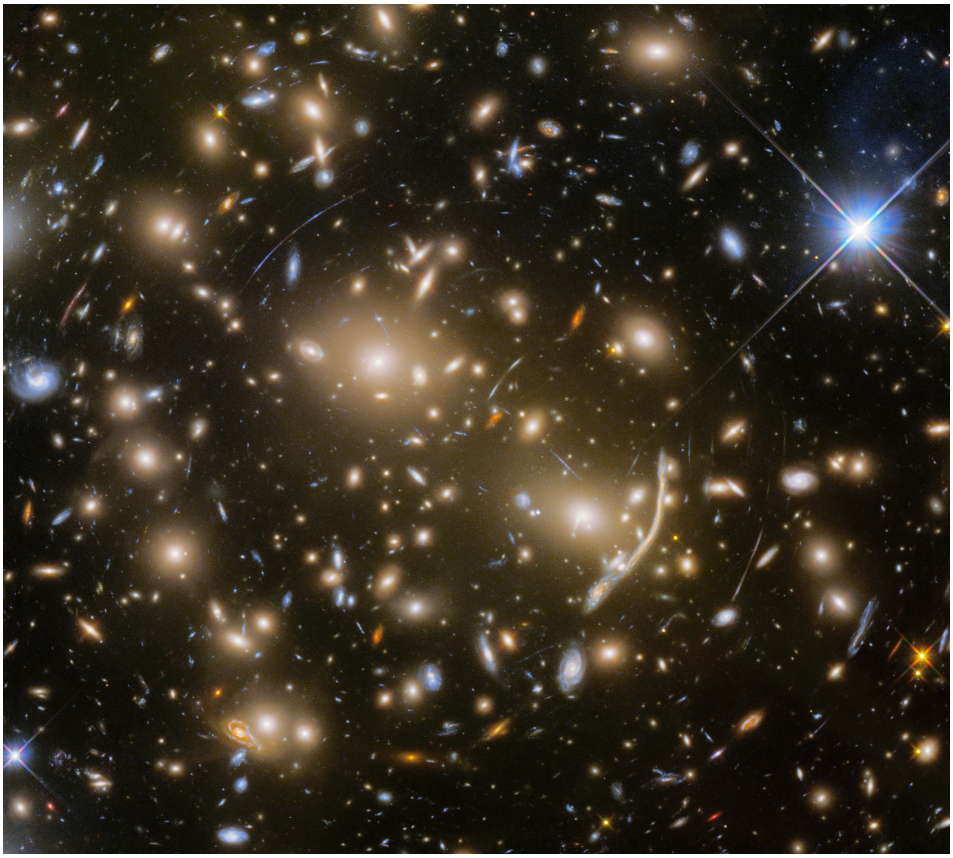

Figure 1.5: The galaxy cluster Abell 370, imaged with the Hubble Space Telescope by the Frontier Fields team. Circular arcs are clearly visible around the cluster centre: these are background galaxies that are strongly lensed and distorted by the mass of the cluster. Towards the cluster outskirts, this effect diminishes and it can only be detected statistically in the coherent distortion of large samples of background galaxies. *Picture from ESA/Hubble*.

Schneider, 2001). In the weak lensing regime, the intrinsic complex ellipticity of the source, $\epsilon^{\rm s}$, is related to the measured ellipticity after lensing, $\epsilon$, through a transformation that only depends on the reduced shear, $g$,

$$\epsilon \approx \epsilon^{\rm s} + g \tag{1.35}$$

with

$$g(\boldsymbol{\theta}) = \frac{\gamma(\boldsymbol{\theta})}{1-\kappa(\boldsymbol{\theta})}\,, \tag{1.36}$$

and $\gamma = \gamma_1 + i\gamma_2$. Since the sources are assumed to be randomly oriented, we expect $\langle \epsilon^{\rm s} \rangle = 0$ for small local patches containing sufficient galaxies. This means that the mea-

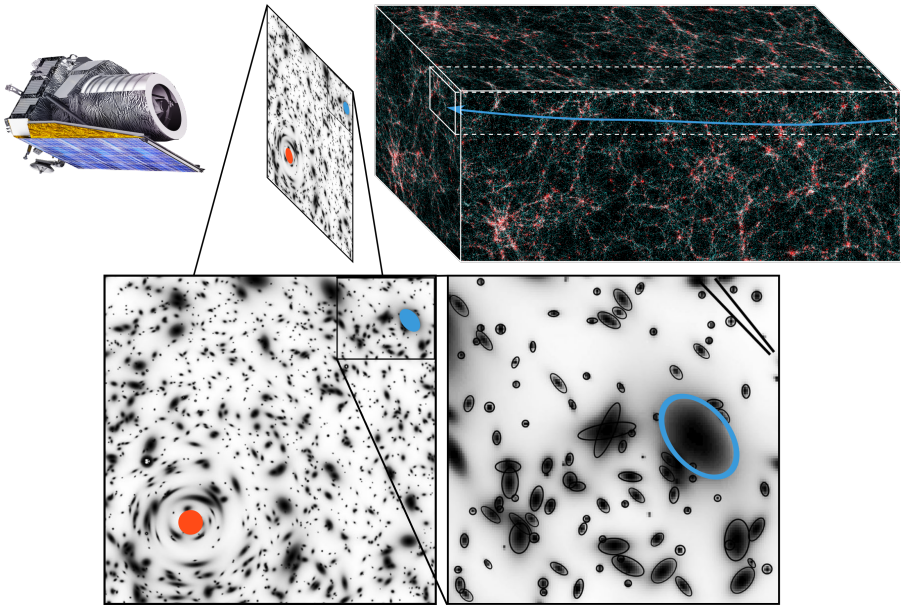

Figure 1.6: The differential bending of photon paths due to an intervening mass distribution, as illustrated in the top-right slice through the density field of the Euclid Flagship simulation, distorts the observed galaxy shapes (bottom panels). A massive cluster, indicated by the red circle, strongly lenses background galaxies close to the cluster centre, distorting them into clearly visible arcs (bottom-left image). Further out, the effect is smaller and the distortion can only be inferred by averaging over the shape of a large sample of galaxies (bottom-right inset). The thick bars in the top-right corner show the true shear signal (bottom bar) and the signal inferred from averaging the observed galaxy shapes (top bar). *Figure created by combining Euclid flagship mock data and illustrations by Mellier (1999).*

sured ellipticity directly probes the reduced shear,

$$\langle \epsilon \rangle \approx g \approx \gamma \,. \tag{1.37}$$

Since weak lensing is by definition a small effect, a sufficient number of galaxies is needed to obtain a statistically significant measurement of the reduced shear.

### 1.4.2 Cosmic shear

Blandford et al. (1991) extrapolated the weak lensing idea of galaxies and clusters to measure the mass fluctuations over large regions of the sky to constrain the typical clustering of matter between us and samples of background galaxies. This effect is known as "cosmic shear".

The power spectrum of the cosmic shear can be derived heuristically from the convergence defined in Eq. (1.27). First, we can define the convergence for a source at redshift $z_\mathrm{s}$ being lensed by an overdense thin matter slice at redshift $z_\mathrm{d}$ as follows

$$\delta\kappa(\boldsymbol{\theta}, z_\mathrm{s}, z_\mathrm{d}) = \frac{4\pi G}{c^2}\frac{D_\mathrm{ds}D_\mathrm{d}}{D_\mathrm{s}}(\rho(\boldsymbol{\theta}, z_\mathrm{d}) - \bar{\rho}(z_\mathrm{d}))\delta l\,. \tag{1.38}$$

We can integrate this expression along the line-of-sight to the source while switching to comoving coordinates and using the fact that $\rho(\boldsymbol{\theta}, z) - \bar{\rho} = \delta(\boldsymbol{\theta}, z)\bar{\rho}$ and

$$\bar{\rho} = \frac{3H_0^2\Omega_{\mathrm{m},0}}{8\pi G}\,, \tag{1.39}$$

giving the effective convergence

$$\kappa_\mathrm{eff}(\boldsymbol{\theta}, \chi_\mathrm{s}) = \frac{3H_0^2\Omega_{\mathrm{m},0}}{2c^2}\int_0^{\chi_\mathrm{s}}\mathrm{d}\chi\,\frac{f_K(\chi_\mathrm{s}-\chi)f_K(\chi)}{f_K(\chi_\mathrm{s})}\frac{\delta(f_K(\chi)\boldsymbol{\theta}, \chi)}{a(\chi)}\,, \tag{1.40}$$

where $f_K(\chi)$, defined in Eq. (1.3), enters through the comoving angular diameter distance as can be seen from Eq. (1.18). Realistically, the shapes of many source galaxies over a range of redshifts, or comoving distances, will be measured. Assuming a source redshift distribution, $n(z)$, with $n(z)\mathrm{d}z = n(\chi)\mathrm{d}\chi$, a selection limit, $\chi_\mathrm{lim}$, and accounting for the fact that only galaxies behind the matter overdensity will be lensed, we find

$$\kappa_\mathrm{eff}(\boldsymbol{\theta}) = \frac{3H_0^2\Omega_{\mathrm{m},0}}{2c^2}\int_0^{\chi_\mathrm{lim}}\mathrm{d}\chi W(\chi)f_K(\chi)\frac{\delta(f_K(\chi)\boldsymbol{\theta}, \chi)}{a(\chi)}\,, \tag{1.41}$$

with the lensing efficiency for the source distribution given by

$$W(\chi) = \int_\chi^{\chi_\mathrm{lim}}\mathrm{d}\chi_\mathrm{s}n(\chi_\mathrm{s})\frac{f_K(\chi_\mathrm{s}-\chi)}{f_K(\chi_\mathrm{s})}\,. \tag{1.42}$$

This simple argument gives the same result as a full general relativistic calculation of the deviation between neighbouring photons that reach the observer from minutely different angular positions after travelling through a perturbed homogeneous and isotropic universe (e.g. Bartelmann & Schneider, 2001).

The cosmic shear power spectrum, $P_\gamma(\boldsymbol{\ell})$, is then given by the correlations between galaxy shape distortions on different scales, $\boldsymbol{\ell}$, in the plane of the sky. This power spectrum can be calculated from the convergence power spectrum, since the Fourier components of the shear, $\gamma$, and the convergence, $\kappa$, only differ by a phase—which can easily be derived from Eqs. (1.29), (1.33), and (1.34), and the fact that $\gamma = \sqrt{\gamma_1^2+\gamma_2^2}e^{2i\phi}$. Using the Limber approximation (full details can be found in Section 2.4 of Bartelmann & Schneider, 2001), the projected correlations on the sky can be related to the 3D power spectrum of density fluctuations, $P_\mathrm{m}$, as follows

$$P_\kappa(\boldsymbol{\ell}) = \frac{9H_0^4\Omega_{\mathrm{m},0^2}}{4c^2}\int_0^{\chi_\mathrm{lim}}\mathrm{d}\chi\,\frac{W^2(\chi)}{a^2(\chi)}P_\mathrm{m}(\boldsymbol{\ell}/f_K(\chi), \chi)\,, \tag{1.43}$$

where $\boldsymbol{\ell} = f_K(\chi)\mathbf{k}_\perp$ and $\mathbf{k}_\perp$ is the wavevector in the plane of the sky. This result has some noteworthy implications: a fixed lengthscale, $\ell$, on the sky receives contributions from many different physical scales along the line-of-sight, since it corresponds to larger scales at higher redshifts. Interestingly, however, fluctuations *along* the line-of-sight do not contribute to the cosmic shear, which implies that we can limit the impact of small-scale non-linearities by choosing an appropriate cutoff $\boldsymbol{\ell}_{\rm max}$.

In conclusion, in cosmic shear analyses, the statistical shape distortion of galaxies at different distances from us, constrains the typical clustering of matter between us and those galaxies (e.g. Tyson et al., 1984; Blandford et al., 1991). A successful weak lensing analysis is observationally challenging since it needs to be able to accurately measure the galaxy shapes, which requires a small and well-characterized point-spread function, a good understanding of the image noise, the blending of sources, and the impact of undetected background galaxies (e.g. Hoekstra et al., 2017, 2021). Another difficulty arises in the distance measurement to galaxies: due to the massive size of the survey, not all galaxies will have reliable spectroscopic measurements to determine distances from their spectra. Instead, their distance will have to be inferred photometrically with the information available from broadband filters, which results in significantly larger uncertainties (for a recent review, see Salvato et al., 2019).

Additionally, background galaxies that are part of coherent large-scale structures along the line-of-sight can have intrinsic correlations in their shape which are not of cosmological origin, but that depend on galaxy formation processes (e.g. Croft & Metzler, 2000; Heavens et al., 2000). Hirata & Seljak (2004) identified that lensing structures additionally generate anti-correlated shape distortions between galaxies local to the lens that are stretched along the tidal field towards overdensities, and background galaxies that will be distorted tangentially to the tidal field generated by the overdensity. These intrinsic galaxy alignments pose a significant challenge for weak lensing surveys with different surveys finding different values for the strength of this signal (e.g. Efstathiou & Lemos, 2018). Fortunately, Fortuna et al. (2021) found that it is possible to constrain the strength of the intrinsic alignment of galaxies observationally, which can prevent the parameter being high-jacked to hide non-related systematic uncertainties in the analysis.

On the theoretical side, we also need accurate predictions of the average non-linear matter clustering, $P_{\rm m}(k)$ and its time evolution in order to correctly interpret the observed signal and infer the cosmological matter distribution (e.g. Jain & Seljak, 1997; Schneider et al., 1998). The non-linear behaviour can be predicted accurately by dark matter-only (DMO) simulations. However, cosmological hydrodynamical simulations that evolve the ordinary and dark mater distribution jointly in an expanding universe, have shown that violent AGN feedback processes significantly alter the large-scale structure compared to universes that only contain dark matter (e.g. Rudd et al., 2008; Semboloni et al., 2011; van Daalen et al., 2011, 2020). While these effects can be mitigated by only considering large scales (as is done by the Dark Energy Survey, e.g. Amon et al., 2022b), future surveys would be throwing away a considerable amount of useful information by neglecting small scales (e.g. Taylor et al., 2018).

In the absence of observational data that can fully constrain the normal matter distribution, including small-scale baryonic effects in the matter power spectrum necessarily relies on cosmological, hydrodynamical simulations to predict the expected effect. As

McCarthy et al. (2017) stresses, it is important that the simulations correctly reproduce the observed properties of the large scale matter distribution. The main idea is to come up with an analysis method that can either add the baryonic contribution to the matter power spectrum predicted from suites of dark matter-only simulations, or remove the baryonic signal from the data. Eifler et al. (2015) and Huang et al. (2019) perform a principal component analysis of the matter power spectrum including baryonic effects which results in a flexible model for the expected signal that can be marginalized over in the cosmological analysis. Mead et al. (2015) use hydrodynamical simulations to calibrate a halo model with additional freedom in the halo density profiles to flexibly capture the total matter power spectrum, including baryons. Schneider et al. (2019) phenomenologically modify the matter distribution in DMO simulations by shifting particles in haloes such that the density profiles more closely match those in observations.

The need to account for all the aforementioned effects makes cosmic shear a complex cosmological observable. However, the potential gain in information about the Universe from directly probing the total matter distribution is enormous.

## 1.5 Galaxy clusters

The final cosmological probe of interest to this thesis, is the abundance of massive galaxy clusters. Clusters are located at the nodes of the intricate cosmic web, resulting in a continuous accretion of smaller galaxy groups, gas, and an occasional merger with another massive cluster. The cluster mass growth depends strongly on the amount of matter in the Universe and how clustered it is. The late-time accelerating expansion due to dark energy suppresses their growth and leaves a clear imprint in the cluster abundance as a function of mass and redshift, making clusters powerful probes of the matter distribution of the Universe and also of the history of dark energy (e.g. Allen et al., 2011). However, since the cluster abundance decreases exponentially with increasing mass, any error in the cluster mass calibration will result in highly biased estimates of the cosmology.

Clusters are fascinating laboratories: they contain hundreds to thousands of galaxies within a radius of a few million lightyears, making them easy to detect as localized overdensities of galaxies in large optical surveys. In the 50s and 60s, Abell (1958) and Zwicky et al. (1961, 1963, 1966) released large catalogues of clusters identified exactly this way. The launch of the first X-ray telescopes in the 70s, opened up a whole new view (e.g. Sarazin, 1986). It became clear that clusters are filled with hot, metal-enriched gas that is supported in a state of quasi-equilibrium with the gas pressure opposing full gravitational collapse. The hot gas is a shining beacon in X-rays. Additionally, the hot cluster electrons can boost the photons of the CMB through inverse Compton scattering, leaving behind a clear hole in the temperature fluctuations of the CMB below a frequency of $\approx 200\,\mathrm{GHz}$, known as the Sunyaev–Zel'dovich effect (Sunyaev & Zeldovich, 1972). There are thus many different ways to observationally identify clusters.

On the theoretical side, we can use large-volume, dark matter-only simulations to accurately predict the abundance of clusters for varying cosmologies (e.g. McClintock et al., 2019; Nishimichi et al., 2019; Bocquet et al., 2020). The problem that we face, however, is that a significant fraction of the total mass of actual clusters consists of hot

gas that behaves differently from the dark matter. Velliscig et al. (2014) showed that AGN feedback redistributes the hot cluster gas component to large scales, resulting in lower masses closer to the cluster centre when comparing to clusters consisting purely of collisionless dark matter. Since galaxy clusters dominate the total matter along their line-of-sight, their total mass, including dark and ordinary matter, can be measured from the weak lensing signal that they generate in the shapes of the background galaxies, as described in Section 1.4.1. This effect was first detected by Tyson et al. (1990). The only remaining problem is then linking the inferred mass from observations to the theoretical, DMO cluster mass.

The mass change is small for the most massive clusters ($m_{500c} > 10^{14.5}\,\mathrm{M}_\odot$)[6] since they are able to retain practically all their baryons within the radius $r_{500c}$ where weak lensing observations are able to infer the total mass. Hence, the differing distribution of the baryons is not a big problem for current cluster surveys whose detection limit is $m_{500c,\mathrm{lim}} \approx 10^{14.5}\,\mathrm{M}_\odot$ (e.g. Bocquet et al., 2016). However, future surveys will be able to detect clusters down to masses $m_{500c,\mathrm{lim}} \approx 10^{14}\,\mathrm{M}_\odot$ (e.g. Sartoris et al., 2016), where the mass change compared to the DMO halo is predicted to significantly exceed the sub-percent statistical uncertainty of the cluster mass determination afforded by the massive cluster sample (e.g. Köhlinger et al., 2015). Our only options are then to either link the observed cluster mass to the theoretical mass that the same cluster would have in a universe consisting only of dark matter (e.g. Balaguera-Antolínez & Porciani, 2013; Cusworth et al., 2014), or to predict the cosmology dependence of the cluster abundance in large-volume cosmological hydrodynamical simulations. Unfortunately, the computational cost of the latter option has so far been a barrier to its realization.

## 1.6 This thesis

In this thesis, we study how our imperfect knowledge of galaxy formation processes and their impact on the total matter distribution impacts the cosmological analysis of future all-sky galaxy surveys. Galactic winds generated by supernova explosions or buoyant bubbles lifted by the powerful supermassive black holes at the centres of galaxies significantly alter the matter distribution compared to what it would be if we only accounted for gravitational collapse. This poses a challenging problem for future surveys, since these processes are still not well-understood theoretically.

In **Chapter 2**, we shed light on the issue of the total matter distribution uncertainty for cosmic shear surveys. We resort to a phenomenological halo model that can incorporate observational information to constrain the total matter distribution, while including enough freedom to quantify how the unobserved baryonic matter can impact the final answer. We show that it is crucially important to constrain the amount of hot gas in groups and clusters of galaxies, since these objects dominate the total matter clustering signal on intermediate scales that dominate the cosmic shear signal. We also reiterate the important point that the dominant baryonic effect is not changing the clustering properties of

[6] The mass $m_{500c}$ is the total mass within the region $r_{500c}$ where the average density of the cluster is $\langle\rho\rangle(< r_{500c}) = 500\rho_{\mathrm{crit}}(z)$.

haloes, but rather the redistribution and systematic lowering of the halo mass compared to a universe that only includes dark matter.

We switch our focus to cluster abundance studies in **Chapter 3**, instigated by how successfully the halo model of Chapter 2 reproduces the impact of baryons on the matter distribution seen in simulations, while only using observational data for the baryon distribution. We use X-ray observations of the hot gas in clusters to constrain the total galaxy cluster density profile and we establish a link between observed haloes and the theoretical haloes in a universe containing only dark matter whose abundance we can predict with simulations. Inferring the cluster mass from mock weak lensing observations then allows us to quantify the mass bias incurred when not properly accounting for the differing density profiles between the ordinary and the dark matter. We show that future surveys with such inaccurate mass calibrations will infer significantly biased values for the amount of matter, its clustering, and the dark energy equation of state.

In **Chapter 4**, we tackle the mass calibration of galaxy clusters in a different way. We suggest that, currently, there is an unfair onus on weak lensing mass calibrations to constrain 3D cluster masses that are easy to measure in simulations but difficult to infer observationally. We propose to calibrate the cluster abundance as a function of the excess projected mass of a cluster: a quantity that is directly probed observationally and easily measured in simulations. We show that these projected masses have a significantly lower observational uncertainty than 3D cluster masses. Moreover, using a large suite of dark matter-only simulations, we find that the cluster abundance as a function of the projected mass is at least as sensitive to changes in cosmology as the abundance measured for 3D masses. As a result, this projected mass determination can significantly reduce the systematic uncertainty in the mass calibration for cluster abundance studies.

Finally, in **Chapter 5**, we use cosmological hydrodynamical simulations to quantify the impact of baryons on the excess projected mass measurements proposed in Chapter 4 by comparing the change in mass between the same haloes in universes with and without baryons. We find that the projected masses are slightly less affected by feedback processes than 3D halo masses, since the projected mass receives contributions from the halo outskirts along the line-of-sight where the baryons eventually trace the dark matter. While the reduction is not dramatic, together with the decreased systematic uncertainty in the mass calibration it provides another benefit to projected mass calibrations.

# Bibliography


Aarseth S. J., 1963, Monthly Notices of the Royal Astronomical Society, 126, 223 → p. 16

Abbott T. M. C., et al., 2021, ApJS, 255, 20 → p. 4

Abdalla E., et al., 2022, Journal of High Energy Astrophysics, 34, 49 → p. 7

Abell G. O., 1958, The Astrophysical Journal Supplement Series, 3, 211 → p. 26

Addison G. E., Huang Y., Watts D. J., Bennett C. L., Halpern M., Hinshaw G., Weiland J. L., 2016, ApJ, 818, 132 → p. 7

Aihara H., et al., 2018, Publ. Astron. Soc. Japan, 70, 1 → p. 4

Allen S. W., Evrard A. E., Mantz A. B., 2011, Annu. Rev. Astron. Astrophys., 49, 409 → p. 26

Amon A., et al., 2022a, arXiv:2202.07440 [astro-ph] → p. 8

Amon A., et al., 2022b, Phys. Rev. D, 105, 023514 → p. 8, 25

Anand G. S., Tully R. B., Rizzi L., Riess A. G., Yuan W., 2022, Comparing Tip of the Red Giant Branch Distance Scales: An Independent Reduction of the Carnegie-Chicago Hubble Program and the Value of the Hubble Constant (arXiv:2108.00007), doi:10.48550/arXiv.2108.00007 → p. 7

Asgari M., et al., 2021, A&A, 645, A104 → p. 8

Bacon D. J., Refregier A. R., Ellis R. S., 2000, Monthly Notices of the Royal Astronomical Society, 318, 625 → p. 6

Bahcall N. A., Cen R., 1993, The Astrophysical Journal, 407, L49 → p. 6

Balaguera-Antolínez A., Porciani C., 2013, J. Cosmol. Astropart. Phys., 2013 → p. 27

Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, Phys. Rev. D, 28, 679 → p. 13

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, Astrophys. J., 304, 15 → p. 15

Barnes J., Hut P., 1986, Nature, 324, 446 → p. 16

Bartelmann M., Schneider P., 2001, Physics Reports, 340, 291 → p. 20, 21, 24

Birrer S., et al., 2020, A&A, 643, A165 → p. 7

Blandford R. D., 1999, 182, 87 → p. 18

Blandford R. D., Saust A. B., Brainerd T. G., Villumsen J. V., 1991, Monthly Notices of the Royal Astronomical Society, 251, 600 → p. 21, 23, 25

Bocquet S., Saro A., Dolag K., Mohr J. J., 2016, Mon. Not. R. Astron. Soc., 456, 2361 → p. 27

Bocquet S., et al., 2019, Astrophys. J., 878, 55 → p. 7

Bocquet S., Heitmann K., Habib S., Lawrence E., Uram T., Frontiere N., Pope A., Finkel H., 2020, Astrophys. J., 901, 5 → p. 17, 26

Booth C. M., Schaye J., 2009, Mon. Not. R. Astron. Soc., 398, 53 → p. 19

Chiba T., De Felice A., Tsujikawa S., 2013, Phys. Rev. D, 87, 083505 → p. 6

Clowe D., Bradač M., Gonzalez A. H., Markevitch M., Randall S. W., Jones C., Zaritsky D., 2006, The Astrophysical Journal, 648, L109 → p. 6

Cole S., et al., 2005, Monthly Notices of the Royal Astronomical Society, 362, 505 → p. 6, 12

Croft R. A. C., Metzler C. A., 2000, The Astrophysical Journal, 545, 561 → p. 8, 25

Curtis H. D., 1915, Publications of the Astronomical Society of the Pacific, 27, 214 → p. 1

Cusworth S. J., Kay S. T., Battye R. A., Thomas P. A., 2014, Mon. Not. R. Astron. Soc., 439, 2485 → p. 27

Dalla Vecchia C., Schaye J., 2008, Mon. Not. R. Astron. Soc., 387, 1431 → p. 18

Dalla Vecchia C., Schaye J., 2012, Mon. Not. R. Astron. Soc., 426, 140 → p. 18

Dyson F. W., Eddington A. S., Davidson C., 1920, Philosophical Transactions of the Royal Society of London, 220, 291 → p. 19

Eastwood J. W., 1975, Journal of Computational Physics, 18, 1 → p. 16

Eddington A. S., 1919, The Observatory, 42, 119 → p. 19

Efstathiou G., 2020, A Lockdown Perspective on the Hubble Tension (with Comments from the SH0ES Team) (arXiv:2007.10716) → p. 7
Efstathiou G., Eastwood J. W., 1981, Monthly Notices of the Royal Astronomical Society, 194, 503 → p. 16
Efstathiou G., Lemos P., 2018, Mon. Not. R. Astron. Soc. → p. 8, 25
Efstathiou G., Frenk C. S., White S. D. M., Davis M., 1988, Mon. Not. R. Astron. Soc., 235, 715 → p. 15
Eifler T., Krause E., Dodelson S., Zentner A. R., Hearin A. P., Gnedin N. Y., 2015, Mon. Not. R. Astron. Soc., 454, 2451 → p. 26
Einstein A., 1916, Annalen der Physik, 354, 769 → p. 9
Eisenstein D. J., et al., 2005, The Astrophysical Journal, 633, 560 → p. 6, 12
Euclid Collaboration et al., 2018, 000 → p. 17
Faber S. M., et al., 1997, The Astronomical Journal, 114, 1771 → p. 18
Fabian A. C., 1999, Monthly Notices of the Royal Astronomical Society, 308, L39 → p. 18
Fabjan D., Borgani S., Tornatore L., Saro A., Murante G., Dolag K., 2010, Monthly Notices of the Royal Astronomical Society, 401, 1670 → p. 19
Fall S. M., Efstathiou G., 1980, Monthly Notices of the Royal Astronomical Society, 193, 189 → p. 18
Feeney S. M., Peiris H. V., Williamson A. R., Nissanke S. M., Mortlock D. J., Alsing J., Scolnic D., 2019, Phys. Rev. Lett., 122, 061105 → p. 8
Fortuna M. C., et al., 2021, A&A, 654, A76 → p. 8, 25
Freedman W. L., 2021, ApJ, 919, 16 → p. 7
Freedman W. L., et al., 2019, The Astrophysical Journal, 882, 34 → p. 7
Friedmann A., 1922, Zeitschrift fur Physik, 10, 377 → p. 10
Frieman J. A., Turner M. S., Huterer D., 2008, Annu. Rev. Astron. Astrophys., 46, 385 → p. 6
Gerardi F., Martinelli M., Silvestri A., 2019, J. Cosmol. Astropart. Phys., 2019, 042 → p. 6
Giblin B., et al., 2021, Astronomy and Astrophysics, 645, A105 → p. 4
Guth A. H., 1981, Phys. Rev. D, 23, 347 → p. 13
Habib S., Heitmann K., Higdon D., Nakhleh C., Williams B., 2007, Physical Review D, 76, 083503 → p. 17
Hand E., 2012, Nature, 490, 16 → p. 3
Hearin A. P., Zentner A. R., Ma Z., 2012, J. Cosmol. Astropart. Phys., 2012, 034 → p. 4
Heavens A., Refregier A., Heymans C., 2000, Monthly Notices of the Royal Astronomical Society, 319, 649 → p. 8, 25
Heitmann K., Higdon D., Nakhleh C., Habib S., 2006, The Astrophysical Journal, 646, L1 → p. 17
Heitmann K., Higdon D., White M., Habib S., Williams B. J., Lawrence E., Wagner C., 2009, Astrophys. J., 705, 156 → p. 17
Heymans C., et al., 2021, Astronomy and Astrophysics, 646, A140 → p. 7
Hirata C. M., Seljak U., 2004, Phys. Rev. D - Part. Fields, Gravit. Cosmol., 70, 1 → p. 8, 25
Hoekstra H., Yee H. K. C., Gladders M. D., 2002, The Astrophysical Journal, 577, 595 → p. 6
Hoekstra H., Viola M., Herbonnet R., 2017, Mon. Not. R. Astron. Soc., 468, 3295 → p. 25
Hoekstra H., Kannawadi A., Kitching T. D., 2021, Astronomy and Astrophysics, 646, A124 → p. 25
Hogg D. W., Eisenstein D. J., Blanton M. R., Bahcall N. A., Brinkmann J., Gunn J. E., Schneider D. P., 2005, The Astrophysical Journal, Volume 624, Issue 1, pp. 54-58., 624, 54 → p. 8
Hu W., Dodelson S., 2002, Annu. Rev. Astron. Astrophys., 40, 171 → p. 6
Hu W., Sugiyama N., Silk J., 1997, Nature, 386, 37 → p. 12
Huang H.-j., Eifler T., Mandelbaum R., Dodelson S., 2019, Mon. Not. R. Astron. Soc., 488, 1652 → p. 26

Hubble E. P., 1925, The Observatory, 48, 139 → p. 1
Hubble E., 1929, Proceedings of the National Academy of Sciences, 15, 168 → p. 1
Jain B., Seljak U., 1997, Astrophys. J., 484, 560 → p. 25
Joachimi B., et al., 2015, Space Sci Rev, 193, 1 → p. 8
Joachimi B., et al., 2021, A&A, 646, A129 → p. 8
Joudaki S., et al., 2020, A&A, 638, L1 → p. 8
Joyce A., Lombriser L., Schmidt F., 2016, Annu. Rev. Nucl. Part. Sci., 66, 95 → p. 6
Kaiser N., 1984, Astrophys. J., 284, L9 → p. 15
Köhlinger F., Hoekstra H., Eriksen M., 2015, Mon. Not. R. Astron. Soc., 453, 3107 → p. 27
Kuijken K., et al., 2019, A&A, 625, A2 → p. 4
LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201 → p. 2
Laureijs R., et al., 2011, preprint → p. 2
Laurent P., et al., 2016, J. Cosmol. Astropart. Phys., 2016, 060 → p. 8
Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49 → p. 1
Linde A. D., 1982, PHYSICS LETTERS, 108, 5 → p. 13
Magorrian J., et al., 1998, Astron. J., 115, 2285 → p. 18
Mandelbaum R., et al., 2018, Publ. Astron. Soc. Japan, 70, 1 → p. 4
McCarthy I. G., et al., 2010, Monthly Notices of the Royal Astronomical Society, pp 822–839 → p. 19
McCarthy I. G., Schaye J., Bird S., Le Brun A. M., 2017, Mon. Not. R. Astron. Soc., 465, 2936 → p. 19, 26
McClintock T., et al., 2019, Astrophys. J., 872, 53 → p. 17, 26
Mead A. J., Peacock J. A., Heymans C., Joudaki S., Heavens A. F., 2015, Mon. Not. R. Astron. Soc., 454, 1958 → p. 26
Mellier Y., 1999, Annu. Rev. Astron. Astrophys., 37, 127 → p. 23
Motloch P., Hu W., 2018, Phys. Rev. D, 97, 103536 → p. 7
Motloch P., Hu W., 2020, Phys. Rev. D, 101, 083515 → p. 7
Navarro J. F., Frenk C. S., White S. D. M., 1996, The Astrophysical Journal, 462, 563 → p. 16
Navarro J. F., Frenk C. S., White S. D. M., 1997, Astrophys. J., 490, 493 → p. 16
Nishimichi T., et al., 2019, Astrophys. J., 884, 29 → p. 17, 26
Ntelis P., et al., 2017, J. Cosmol. Astropart. Phys., 2017, 019 → p. 8
Obied G., Dvorkin C., Heinrich C., Hu W., Miranda V., 2017, Phys. Rev. D, 96, 083526 → p. 7
Paczynski B., 1987, Nature, 325, 572 → p. 21
Peacock J. A., et al., 2001, Nature, 410, 169 → p. 6
Peebles P. J. E., 1965, ApJ, 142, 1317 → p. 15
Peebles P. J. E., 1967, The Astrophysical Journal, 147, 859 → p. 15
Peebles P. J., 2012, Annu. Rev. Astron. Astrophys., 50, 1 → p. 6
Peebles P. J. E., 2017, Nat Astron, 1, 0057 → p. 6
Peebles P. J. E., Yu J. T., 1970, Astrophys. J., 162, 815 → p. 12
Percival W. J., et al., 2001, Monthly Notices of the Royal Astronomical Society, 327, 1297 → p. 6
Perlmutter S., et al., 1999, Astrophys. J., 517, 565 → p. 6, 13
Planck Collaboration et al., 2017, A&A, 607, A95 → p. 7
Planck Collaboration et al., 2020a, A&A, 641, A1 → p. 14
Planck Collaboration et al., 2020b, Astron. Astrophys., 641, A6 → p. 4, 6, 7, 9, 11, 12, 14
Press W. H., Schechter P., 1974, Astrophys. J., 187, 425 → p. 15
Puchwein E., Springel V., 2013, Monthly Notices of the Royal Astronomical Society, 428, 2966 → p. 18, 19
Rees M. J., Ostriker J. P., 1977, Monthly Notices of the Royal Astronomical Society, 179, 541 → p. 18

Rhodes J., et al., 2017, ApJS, 233, 21 → p. 3
Rhodes J., et al., 2019, Bulletin of the American Astronomical Society, 51, 201 → p. 3
Riess A. G., et al., 1998, Astron. J., 116, 1009 → p. 6, 13
Riess A. G., et al., 2021, A Comprehensive Measurement of the Local Value of the Hubble Constant with 1 Km/s/Mpc Uncertainty from the Hubble Space Telescope and the SH0ES Team → p. 7
Rubin V. C., Ford Jr. W. K., 1970, The Astrophysical Journal, 159, 379 → p. 6
Rubin V. C., Thonnard N., Ford Jr. W. K., 1980, ApJ, 238, 471 → p. 6
Rudd D. H., Zentner A. R., Kravtsov A. V., 2008, Astrophys. J., 672, 19 → p. 25
Salvato M., Ilbert O., Hoyle B., 2019, Nat Astron, 3, 212 → p. 25
Sarazin C. L., 1986, Rev. Mod. Phys., 58, 1 → p. 26
Sartoris B., et al., 2016, Mon. Not. R. Astron. Soc., 459, 1764 → p. 27
Schaye J., 2004, Astrophys. J., 609, 667 → p. 18
Schaye J., Dalla Vecchia C., 2008, Mon. Not. R. Astron. Soc., 383, 1210 → p. 18
Schaye J., et al., 2015, Mon. Not. R. Astron. Soc., 446, 521 → p. 19
Schneider P., Van Waerbeke L., Jain B., Kruse G., 1998, Mon. Not. R. Astron. Soc., 296, 873 → p. 25
Schneider A., Teyssier R., Stadel J., Chisari N. E., Brun A. M. L., Amara A., Refregier A., 2019, J. Cosmol. Astropart. Phys., 2019, 020 → p. 26
Semboloni E., Hoekstra H., Schaye J., Van Daalen M. P., McCarthy I. G., 2011, Mon. Not. R. Astron. Soc., 417, 2020 → p. 25
Sevilla-Noarbe I., et al., 2021, ApJS, 254, 24 → p. 4
Shapley H., Curtis H. D., 1921, The Scale of the Universe. Washington, D.C., Published by the National research council of the National academy of sciences → p. 1
Silk J., 1977, The Astrophysical Journal, 211, 638 → p. 18
Silk J., Rees M. J., 1998, Astron. Astrophys., 331, L1 → p. 18
Slipher V. M., 1921, Popular Astronomy, 29, 128 → p. 1
Slipher V. M., 1922, 4, 284 → p. 1
Soares-Santos M., et al., 2019, ApJ, 876, L7 → p. 8
Somerville R. S., Davé R., 2015, Annu. Rev. Astron. Astrophys, 53, 51 → p. 18
Spergel D. N., et al., 2003, Astrophys. J. Suppl. Ser., 148, 175 → p. 6
Spergel D., et al., 2015, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report (arXiv:1503.03757) → p. 2
Springel V., et al., 2005, Nature, 435, 629 → p. 16, 17
Stromberg G., 1925, The Astrophysical Journal, 61, 353 → p. 1
Sunyaev R. A., Zeldovich Y. B., 1972, The Observation of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. Vol. 4 → p. 26
Taylor P. L., Kitching T. D., McEwen J. D., 2018, Phys. Rev. D, 98, 043532 → p. 4, 25
The LIGO Scientific Collaboration and The Virgo Collaboration et al., 2017, Nature, 551, 85 → p. 7
Tyson J. A., Valdes F., Jarvis J. F., Mills Jr. A. P., 1984, The Astrophysical Journal, 281, L59 → p. 19, 25
Tyson J. A., Valdes F., Wenk R. A., 1990, The Astrophysical Journal, 349, L1 → p. 27
Undagoitia T. M., Rauch L., 2016, J. Phys. G: Nucl. Part. Phys., 43, 013001 → p. 6
Van Waerbeke L., et al., 2000, Astronomy and Astrophysics, v.358, p.30-44 (2000), 358, 30 → p. 6
Velliscig M., van Daalen M. P., Schaye J., McCarthy I. G., Cacciato M., Le Brun A. M., Vecchia C. D., 2014, Mon. Not. R. Astron. Soc., 442, 2641 → p. 27
Wang Y., Pogosian L., Zhao G.-B., Zucca A., 2018, ApJ, 869, L8 → p. 6
Weinberg S., 1972, Gravitation And Cosmology: Principles And Applications Of The General Theory Of Relativity. John Wiley & Sons, Inc. → p. 10

Weinberg S., 1989, Rev. Mod. Phys., 61, 1 → p. 6
White S. D. M., Rees M. J., 1978, Mon. Not. R. Astron. Soc., 183, 341 → p. 18
Wiersma R. P. C., Schaye J., Smith B. D., 2009a, Mon. Not. R. Astron. Soc., 393, 99 → p. 18
Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, Mon. Not. R. Astron. Soc., 399, 574 → p. 18
Wilson R. E., 1915, Popular Astronomy, 23, 553 → p. 1
Wittman D. M., Tyson J. A., Kirkman D., Dell'Antonio I., Bernstein G., 2000, Nature, 405, 143 → p. 6
Xu G., 1995, The Astrophysical Journal Supplement Series, 98, 355 → p. 16
Zwicky F., 1933, Helvetica Physica Acta, 6, 110 → p. 6
Zwicky F., Herzog E., Wild P., Karpowicz M., Kowal C. T., 1961, Catalogue of Galaxies and of Clusters of Galaxies, Vol. I → p. 26
Zwicky F., Herzog E., Wild P., 1963, Catalogue of Galaxies and of Clusters of Galaxies, Vol. 2 → p. 26
Zwicky F., Herzog E., Wild P., 1966, Catalogue of Galaxies and of Clusters of Galaxies, Vol. 3 → p. 26
van Daalen M. P., Schaye J., Booth C. M., Dalla Vecchia C., 2011, Monthly Notices of the Royal Astronomical Society, 415, 3649 → p. 25
van Daalen M. P., McCarthy I. G., Schaye J., 2020, Mon. Not. R. Astron. Soc., 491, 2424 → p. 25