Probing the properties of dark matter particles with astrophysical observations

Magalich, A.

Citation

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/82071 holds various files of this Leiden University dissertation.

Author: Magalich, A.
Title: Probing the properties of dark matter particles with astrophysical observations
Issue Date: 2019-12-16

Probing the properties
of dark matter particles
with astrophysical observations

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op maandag 16 december 2019
klokke 10:00 uur

door

Andrii Magalich

geboren te Kiev (Oekraïne)
in 1992

| Promotores: | Dr. Alexey Boyarsky |
| | Prof. dr. Ana Achúcarro |

Promotiecommissie:	Prof. dr. J. Lesgourgues (RWTH, Aachen, Duitsland)
	Dr. O. Ruchayskiy (Niels Bohr Institute, Kopenhagen, Denemarken)
	Prof. dr. E.R. Eliel
	Dr. S. Ploeckinger
	Prof. dr. J. Schaye

The cover shows superimposed distributions of dark matter in Cold Dark Matter and Warm Dark Matter cosmologies (in blue and green colors, correspondingly) obtained in cosmological simulations performed for this work.

Contents

Chapter 1

Introduction

Attempts to phenomenologically explain subatomic physics together with the well-rounded theory of electrodynamics have culminated in the development of the most advanced description of particle physics to date. The Standard Model (SM) of particles unites the models of the electromagnetic, weak and strong interactions in a rigid and elegant theoretical framework.

The SM is a very successful theory that passed a very large number of experimental checks [1]. The Standard Model has also performed very well in the *precision tests* where its predictions were tested not only to the leading order of perturbation theory, but also to next or to next-to-next leading order. The *last prediction* of the SM, the Higgs boson, was discovered at the LHC in 2012 [2, 3]. In the subsequent years it was confirmed that the interactions of the discovered particles are exactly as predicted by the Standard Model [4]. This means that the Standard Model provides a complete and closed description of particle physics as observed at accelerators. It is also a mathematically consistent theory that can be valid up to very high energies, probably up to the Planck scale (see e.g. [5] and references therein).

> Nevertheless, it is established today that the SM has to be extended.

This knowledge is based on several observations that can not be incorporated into the Standard Model. These phenomena include:

- **Neutrino masses**: *Within the SM, neutrinos are massless. However, the observation of neutrino oscillations suggests however that these particles have small masses.*

- **Dark matter**: *Astrophysical observations indicate that the mass of matter in the Universe is dominated by a form of matter that does not interact with light – dark matter. None of the SM particles have the required properties.*

- **Baryon asymmetry of the Universe**: *The laws of particle physics, as described by the SM, are the same for matter and anti-matter. Therefore, the SM fails to explain a*

tiny matter-antimatter imbalance in the Early Universe that results in the present day that the Universe is almost completely missing anti-matter.

Experimental challenges "beyond the Standard Model" suggest that some additional physics beyond the SM exists.

From theoretical point of view, the spectrum of possible resolutions of the Beyond SM (BSM) puzzles are essentially boundless. At the same time, we have no firm knowledge about the masses, interaction strength, spin and charges of the new particles responsible for neutrino masses, dark matter and generation of the baryon asymmetry in the Universe. The main science goal of Tevatron [6, 7] and then the LHC [8] was to search for the Higgs boson. Of course, in this situation, it was motivated to search for new physics that could be found together with the Higgs boson, at the same machine. After the results of the LHC Run I and Run II [1], that did not reveal any confirmed signatures of new physics, it has become even more important to search for new physics in a wider context. After the discovery of the Higgs boson [2, 3] the era of "guaranteed discoveries" of particles with predicted properties has finished.

A special role is played in this respect by the data from cosmology and astrophysics. Laboratory experiments are typically sensitive to some specific models or narrow classes of models. At the same time, cosmological and astrophysical data can provide model independent constraints on the properties of new particles and give insights on the nature beyond Standard Model problems. Of course, the problem of dark matter is the one that is the most related to astrophysics and cosmology.

1.1 Dark Matter

Historically, the name "Dark Matter" (DM) was attributed to different phenomena that seemed to indicate that the mass of galaxies or galaxy clusters, that could be deduced from the velocities of their parts, was many times larger than the mass of matter that can be deduced from the luminosity of these objects (or the absorption of light in them). The first observations of this kind (the study of dynamics of the Coma cluster) were performed by F. Zwicky in 1933 [9]. For a long time, this was considered by many researchers as a problem specific for astronomy and, probably, related to observational uncertainties, the existence of non-luminous objects like cold stars or star remnants, planets or some form of dust.

However, currently, it is widely accepted that the Universe is permeated by a form of matter that does not emit or absorb light and manifests itself so far only through the gravitational field it creates. Together with observations that: a) dark matter is much more abundant than normal "luminous matter" and b) there is no candidate within the Standard Model of particle physics that can make such matter – makes the nature of Dark Matter one of the very few most important mysteries of today's physics and astronomy.

Why do we believe that DM exists? The modern evidence for the existence of dark matter comes from at least 5 observationally independent sources each providing very rich data [10]. Moreover, this evidence comes not only from the study of individual galaxies or clusters but from the data describing the observed Universe as a whole. These cosmological arguments, as well as observations of merging clusters (see below) leave very little room for alternative explanations such a modified gravity or modified Newtonian dynamics as a candidate for a unique mechanism behind all observed "DM phenomenology".

The astrophysical evidence for dark matter can be derived from:

- **Rotational velocities in spiral galaxies**

  For any matter emitting or absorbing light with a clear spectral feature (most importantly, emission and absorption lines), the Doppler effect allows measuring the projection of the velocity of this matter along the line of sight of the observer. To use these velocities to reconstruct gravitational force acting on the source (or the absorber), one needs to know the absolute value of the full 3D velocity. In spiral galaxies, this is in principle possible since stars and interstellar gas in these objects form spiral structures embedded in a relatively thin disk (see Fig. 1.1 for an illustration). Geometry and overall dynamics of this disk allow in many cases to reconstruct the direction of 3D velocity and therefore deduce its absolute value from the measured line-of-sight projection. Velocities measured in such a way can be split into the velocity of the galaxy as a whole and rotational velocities of stars and gas inside the galaxy $v(r)$.

  Then, the total mass can be calculated as

  $$M(r) = \frac{v^2(r)r}{G_N},\tag{1.1.1}$$

  where G_N is the Newton's constant.

  Far enough from the center of the galaxy, where the enclosed mass does not grow with the radius anymore, gravitational force starts to get weaker and we expect (from Newton's gravity law) $v \sim 1/\sqrt{r}$. In particular, if the mass of a galaxy is mainly due to gas and stars, such a picture should be observed in the dark regions, where the emission and absorption of light drops by an order of magnitude and the densities of gas and stars are also expected to be orders of magnitude lower. However, in many galaxies [11–14] the rotation curves $v(r)$ become flat far from the center (e.g. see Fig. 1.2). Such a behavior can be explained if a dominant part of the mass of a galaxy is due to some dark matter that does not interact with light and extends to much larger distances from the center (as compared to the ordinary luminous matter).

- **Velocities of stars in dwarf spheroidal satellites of the Milky Way**

  Dwarf spheroidal satellites (dShps) are small galaxies with masses from $10^7 M_\odot$ to $10^{10} M_\odot$ that are part of the Milky Way (or some other galaxy) halo. These objects

Figure 1.1: A spiral galaxy (NGC 1232) in the constellation Eridanus. *Credit: ESO.*
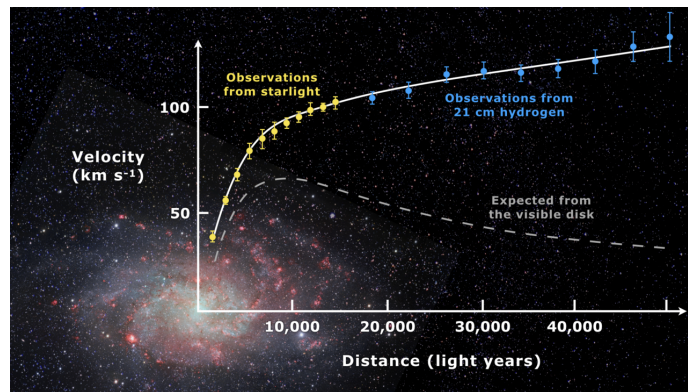
Figure 1.2: The rotation curve for a galaxy M33. *Credit: Wikipedia.*

stopped their independent history a long time ago because they were confined by the Milky Way. As a result, these galaxies do not contain any significant amount of gas and they do not form disks. The absence of the disk does not allow to reconstruct the direction of the total 3D velocity and therefore mass distribution can not be reconstructed using rotation curves. Instead, one can measure the dispersion of the line-of-sight projections of the velocities of stars. Assuming isotropic velocity distribution (having the same dispersion in all directions) one can use Jeans equation to reconstruct the gravitational potential and, therefore, mass (assuming Newtonian gravity). Such an analysis demonstrates that the observed dShps of the Milky Way are extremely dark matter dominated.

DShps have the largest known mass-to-light ratios $(M/L)_\odot \sim 100$ or even more in some cases (e.g. [15, 16] and references therein). It should be noted, however, that

the main uncertainty in the Jeans analysis is that the velocity anisotropy is largely unknown. This uncertainty can significantly affect measurements of the inner mass distributions in dwarf galaxies. However, it is minimized for the mass with the so-called half-light radius [16] and therefore this mass is known with relatively small errors. It may be difficult to reconstruct the total mass of a dSph galaxy or measure the masses around the very center. However, the well-measured mass inside the half-light radius is sufficient to claim that the dynamics of these objects cannot be explained by the masses of its member stars or other luminous matter. Within Newtonian gravity and dynamics, it requires the existence of dark matter dominating these objects.

- **Temperature of gas in galaxy clusters and elliptical galaxies**

A cluster consists of hundreds of galaxies that look more or less like point sources inside the cluster (see right panel of Fig. 1.3). X-ray observations (see left panel of Fig 1.3) show that the intergalactic medium inside a cluster is filled with a diffuse source of thermal X-ray emission with temperatures in the range of 1-10 keV. Modeling shows that the mass of the gas is ~ 15 times larger than the mass of the member galaxies (e.g. [17] and references therein).

In clusters one can, therefore, apply several methods of mass measurements: using the motion of galaxies to reconstruct mass (with some uncertainty); using the temperature of the X-ray emitting gas; using weak and sometimes also strong gravitational lensing.

The most common method for galaxy clusters is based on X-ray observations. X-ray surface brightness and spectrum allow reconstructing the gas temperature. The average temperature (i.e. the average kinetic energy) is roughly related to the gravitational potential energy, i.e. the total mass. As the mean free path of the gas particles is much smaller than the size of the cluster, the thermal equilibrium and the temperature of the gas are local. We can measure the temperature profile $T(r)$ and use it for more detailed mass modeling to reproduce $M(r)$ by solving the hydrostatic equilibrium equation

$$\frac{dp}{dr} = n_{\text{gas}}(r)\frac{dT(r)}{dr} + T(r)\frac{dn_{\text{gas}}(r)}{dr} = -\frac{GM(r)n_{\text{gas}}(r)}{r^2}. \qquad (1.1.2)$$

Mass measurements in clusters reveal the same picture: only 1% of the total mass is given by galaxies, 15% by X-ray gas and 84% by dark matter, see e.g. [18] and references therein. At large enough distances from the center the clusters are very much DM dominated. It was observed long ago [19] that the ratio between the DM density and the density of normal matter in clusters is very close to the average value of this ratio in the whole Universe (see below).

- **Gravitational lensing**
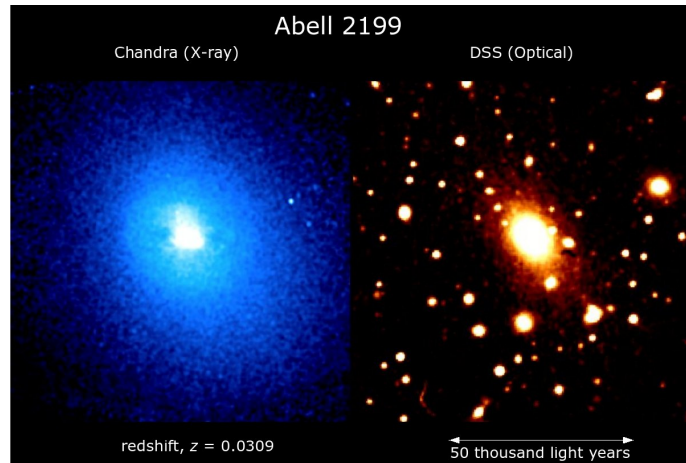
Figure 1.3: The distribution of X-ray gas (left) and galaxies (right) in the cluster of galaxies Abell 2199. *Credit: DSS.*

A prediction of general relativity is that mass can bend light and therefore act as a gravitational lens. The image of a background object, for example a galaxy, therefore gets distorted if its light passes through a massive object, for example a galaxy cluster (see [20] for a review). The degree of the distortion depends critically on the mass density of the gravitational lens, see Fig. 1.4. We might observe a galaxy behind a cluster either as a very thin arc or even as multiple images or a ring ("Einstein ring"), if the cluster is very massive ("strong gravitational lensing") or just slightly elongated ("weak gravitational lensing"). In both cases, the degree of distortion reveals the gravitational mass of the object acting as gravitational lens. Applying this method to galaxies and clusters we see once again that the baryonic mass is not enough to explain the lensing effects.

We can conclude therefore that measurements of gravitational potential in various types of galaxies and clusters of galaxies, with masses from $10^8 M_\odot$ for small galaxies to $\sim 10^{15} M_\odot$ for clusters, using 4 different observational methods (velocities of gravitationally bound objects, X-ray emission of intergalactic gas, weak and strong gravitational lensing) consistently require some additional mass that is not related to absorption and emission of light and dominates the dynamics of these objects.

The analysis of the evolution of the whole Universe at large provides other, **cosmological evidence** for the existence of dark matter and allows us to measure it *average density*.

- **Relic radiation and formation of structures.**

  This line of reasoning start from the observation that we know experimentally that all matter that interacted with light was initially homogeneous to a very large extent.
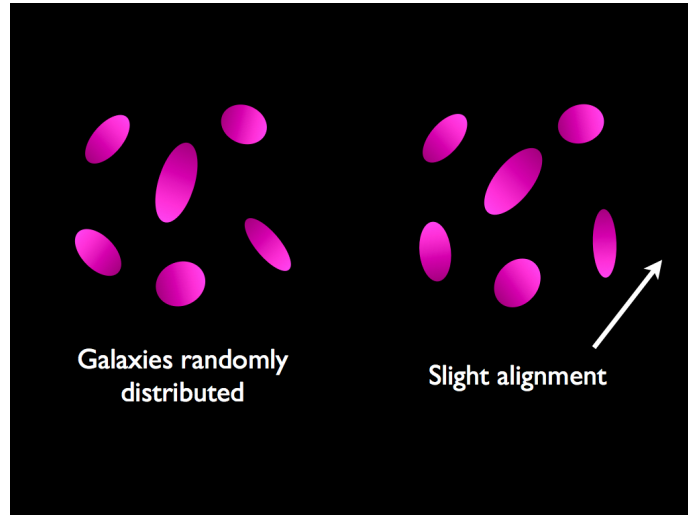
Figure 1.4: An illustration of the effect of weak gravitational lensing. *Credit: E. Grocutt, IfA, Edinburgh* [www.cfhtlens.org].

Indeed, cosmic micro-wave background (CMB), the most ancient light that we can detect, is at the level $\delta T/T \sim 10^{-5}$ [21]. CMB decoupled from matter at recombination of hydrogen, when most of the normal matter in the Universe was combined into a neutral state. This means that already at that moment of recombination (described by the red-shift $z_{CMB} \approx 10^3$) gas of photons was also homogeneous to the same extent [1]. At the same time, photons dominated "normal" matter – the number of photons exceeded the number of baryons by 10 orders of magnitude, as we know both from the properties of CMB and from primordial nucleo-synthesis (see Section 3). Therefore, before recombination, normal matter was also as homogeneous as $\delta\rho/\rho \sim 10^{-5}$ as it interacted strongly with light. Indeed, photons can not be confined in the gravitational field of small overdensities and their scattering on charged particles would stop any clustering of matter. Therefore, these very small (10^{-5}) perturbations of the density of normal matter could start growing only after decoupling of light at recombination.

On the other hand, small density perturbations ($\delta\rho/\rho < 1$) grow linearly with the scale factor in the matter-dominated epoch. Therefore, we would naively expect that today such perturbations are still rather small

$$\delta\rho/\rho = 10^{-5} \cdot \left(\frac{1 + z_{CMB}}{1 + z_0} \right) \approx 10^{-2} \qquad z_{CMB} \approx 10^3. \qquad (1.1.3)$$

This is very different from the Universe that we observe today – densities in galaxies are many orders of magnitude larger than the average matter density in the Universe. This means that some other, "dark" matter that did not interact with light started

---

[1] After recombination interactions of CMB with matter where minimal and could not make it much more homogeneous than it was at recombination

to cluster significantly before recombination. Its over-densities would grow with a scale factor logarithmically in the radiation dominated epoch and linearly in the matter-dominated epoch. After recombination, the dark structures are already quite developed and ordinary matter (in its neutral state) quickly catch up and fall in the gravitational potential formed by DM.

Additionally, the spectrum of anisotropies of CMB (even if their amplitude is so small) is sensitive to the parameters that define the expansion rate of the Universe and the physics of CMB, such as the density of dark matter. The latest results from the Planck collaboration [21] allow to say the the abundance of DM is $\Omega_{\mathrm{DM}} h^2 = 0.120 \pm 0.001$.

We can conclude, that the evidence for dark matter is overwhelming and comes from both theoretically and observationally independent sources. This evidence is based on the data describing dynamics of a large number of individual objects of various types, probed by 4 independent observational techniques as well as on the global picture of structure formation in the Universe as a whole as well as the properties of its Cosmic Microwave background (see e.g. [22–25]).

The Nature of dark matter   There are probably three main possibilities concerning the nature of dark matter. It could consist of some very *special particles* that do not (or almost do not) interact with light; it could be made of some *macroscopic objects* (of some primordial origin?) or it can be a consequence of some *modifications to Newtonian gravity* or dynamics. In fact, all three scenarios require some new physics. The hypothesis of a new dark matter particle can arguably be considered as the simplest and minimal from a theoretical point of view. Moreover, it is very difficult to simultaneously explain *all* the evidence for dark matter with macroscopic objects or modified gravity/dynamics. In addition, massive compact objects (MACHOs) are strongly constrained as dominant DM candidate by microlensing surveys [26, 27].

There is, of course, a possibility that dark matter is not one phenomenon and different observations are explained by different underlying mechanisms. In Physics, however, it is important to study the simplest possibilities first, as required by the *Occam's razor* principle. Indeed, the simplest possibilities are typically also testable to a much larger extent. Therefore, in what follows we will assume that there is only one reason for all the phenomena described above as "evidence for dark matter" and that this reason is microscopic: *all dark matter is made of one particle*. Already this simplest option still requires an enormous and diverse research program, as we have (incomprehensibly) sketched below.

We study below the option that DM is made of particles. These particles should be stable (or be cosmologically long-lived), massive, and electrically neutral. Among

---

known particles, in the Standard Model of particle physics, there are three such particles – three neutrinos of different flavors.

Can SM neutrinos constitute all dark matter in the Universe? As it was shown first in [28], one can put a robust bound on the mass of any fermionic dark matter particle (see [29–32] for more recent discussion). To obtain this so-called *Tremaine-Gunn bound*, one can consider a dwarf galaxy inside the Milky Way halo. The phase space density of the DM particle can be estimated from above (using a lower bound on the mass within a given radius and an upper bound on the velocity of the DM particle that is confined inside this radius). Dividing by the mass of each DM particle, we obtain a lower bound on phase-space number density that should not violate the Pauli exclusion principle and therefore be smaller than the phase space number density of degenerate Fermi gas. This requirement gives a lower bound on the mass of a DM fermion:

$$\frac{M}{\frac{4\pi}{3}r^3}\frac{1}{\frac{4\pi}{3}v^3} \leq \frac{2m_{\mathrm{DM}}^4}{(2\pi\hbar)^3}. \tag{1.1.4}$$

Let us apply this bound to a so-called *classical* dSphs of the Milky Way, where velocities of many stars are measured. For example, for Sculptor dwarf galaxy [16] we can take as a proxy of the object size its half-light-radius $r_h = 283$ pc, the mass inside this radius $M_h = 1.4 \cdot 10^7 M_\odot$ and as a characteristic velocity we take the velocity dispersion $v = \sigma_v = 9.2$ km/s. Substituting these values into Eq. (1.1.4) we get $m_{\mathrm{DM}} > 460$ eV. Other dSphs give similar constraints.

On the other hand, a primordial abundance of relic neutrinos expected in SM of particle physics also depends on the mass of neutrinos and can be easily estimated. Weak interaction keep neutrinos in the equilibrium in the early Universe as long as temperatures are large enough $T > 1$ MeV. Below this temperature weak interactions are too slow (as compared to the expansion of the Universe). As a results, the number density of neutrinos becomes constant in the co-moving frame. Calculating the density of neutrinos at decoupling, we can calculate its present value

$$n_{\nu,0} \sim T_\nu^3(t_0) \simeq 112\,\mathrm{cm}^{-3}, \tag{1.1.5}$$

with $T_\nu(t_0) \approx 1.95$ K being the temperature of neutrino decoupling. This gives

$$\Omega_{\nu\mathrm{DM}}h^2 = \frac{1}{\rho_{c,100}}\sum m_\nu n_{\nu,0} = \frac{\sum m_\nu\,\mathrm{eV}}{94\,\mathrm{eV}}, \tag{1.1.6}$$

where $\rho_{c,100} = \frac{3H_{100}^2}{8\pi G}$ with $H_{100} = 100$ km/s/Mpc. We see that only if

$$\sum m_\nu \simeq 11\,\mathrm{eV} \tag{1.1.7}$$

15

then the SM neutrino can constitute the correct abundance $\Omega_{\mathrm{DM}}h^2 = 0.12$ [21].

> The lowest value of the dark matter particle mass required by the Tremaine-Gunn bound is inconsistent with the cosmological requirement on the neutrino's mass by a factor of ~ 30. This rules out the possibility that SM neutrinos constitute all dark matter. In fact, SM neutrinos give a very sub-dominant contribution to the DM density.

Independently, if DM would be made of particles as light as 100 eV or less that were in thermal equilibrium once (like SM neutrinos), the structure formation in the Universe would happen in a qualitatively different way as compared to what is observed [33]. Indeed, so light DM particles would have velocities close to the speed of light at their decoupling. These velocities would remain relativistic even in the matter-dominated epoch, homogenizing primordial plasma, erasing the overdensities smaller than the "free streaming length" of DM particles. This would mean that clusters of galaxies would form earlier than galaxies. Observationally, however, the galaxies are seen at much larger red-shifts than clusters [18].

> We have convincing evidence that the dominant part of dark matter is not made of the only available Standard Model candidates – neutrinos. Therefore, some new physics beyond the Standard Model is needed to explain dark matter.

1.1.1 Dark matter candidates and their possible properties

If dark matter particle is not a part of the Standard Model we have to consider hypothetical new particles as DM candidates. Particle physics literature offers a wide range of such candidates, motivated by various logic and different approaches [18]. These candidates can be bosons (scalars or vectors) or fermions, have masses from 10^{-20} eV till many TeV or even more. For a long time the so-called WIMPs (weakly interacting massive particles) [34] – heavy fermions involved in weak interactions and decoupling from the primordial plasma like ordinary SM neutrinos (the so-called "thermal relics") – were considered by many scientists as the "most probable" DM candidate (whatever "the most probable" means for a hypothetical particle). This interest was bases on: (i) the so-called "WIMP miracle" – the fact that a particle with mass above 5 GeV and interaction cross-section close to that of ordinary neutrinos has primordial abundance that is an order of magnitude correct, almost independently of its mass; (ii) on general expectation to find new physics at the LHC, more or less together with the Higgs boson – and a WIMP DM particle could be a natural part of such a new physics; (iii) WIMPs could be efficiently search by the so-called direct detection experiments [18, 35–37] as well as colliders, including the LHC [36, 38–41].

The results of the LHC Run I and Run II [1] as well as many years of (so far) un-successful searches for WIMPs at the direct detection experiments as well as the general situation in particles physics create additional motivation to address the problem of dark

matter in a maximally model-independent way. Possible properties of DM as we know them today are compatible with many very different particle physics models.

> Model independent astrophysical and cosmological constraints on the properties of dark matter particles provide invaluable information for particle physics. These results can potentially disfavor the whole directions in particle physics beyond its Standard Model.

Let us try to describe possible properties of DM particles in a maximally model-independent way by their:

- *Self-interaction*: completely ballistic or having potentially observable self-interacting cross-section;

- *Primordial velocities and free streaming*: cold, warm and hot dark matter (see below for definitions);

- *Life-time*: completely stable particles that could only annihilate or particles that could decay (but with cosmologically long life-time).

Cold, Warm dark matter and Hot dark matter This classification is based on the primordial properties of DM particles that depend on the production mechanism in the early Universe and the mass of the particle.

Probably the simplest option is that DM particles are produced with non-relativistic momenta. The examples of particle physics model include: (i) WIMPs, that are "freeze out" from thermal equilibrium at the temperatures smaller than their mass; (ii) axions that are created via the so-called misalignment mechanism [42] that naturally produces particles with very small momenta.

DM particles that are created non-relativistic form the so-called cold dark matter (CDM). Such dark matter is easy to be confined gravitationally and can form haloes of various sizes, including very small ones (e.g. for WIMPs the smallest halo size is set by the horizon at kinetic decoupling in the early universe, see e.g [43]).

For the cases of warm (WDM) and hot (HDM) dark matter, primordial velocities of DM particles are relativistic. Relativistic particle can not be gravitationally confined by the small density perturbations that exist in the early Universe. An average distance traveled by a dark matter particle before it falls into a potential well is called the *free streaming length*. At the scales smaller than the free streaming length, the density becomes more homogeneous, small density perturbations are washed out by random movements of the particles. Density perturbations that are larger than free-streaming length remain untouched by this process. The proper comoving free-streaming length is given by:

$$\lambda_{\text{FS}}(t) \equiv a(t) \int_{t_i}^{t} d\tau \, \frac{v(\tau)}{a(\tau)} \approx 1 \, \text{Mpc} \left(\frac{1 \text{keV}}{M_{\text{DM}}} \right) \frac{\langle p_{\text{DM}} \rangle}{\langle p_\nu \rangle} \tag{1.1.8}$$
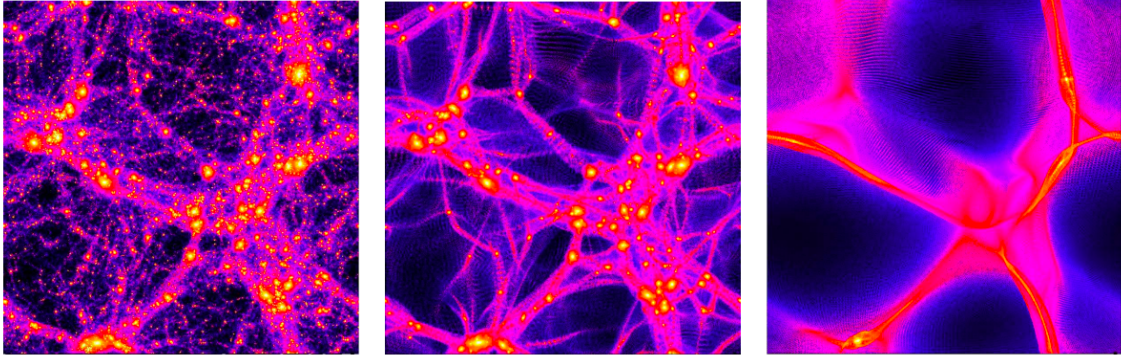
Figure 1.5: Simulation of the large scale structure for the cases of CDM, WDM and HDM (from left to right) cosmologies. *Credit: Maccio et al.* [44].

where $v(t)$ is a typical velocity of DM particles, t_i is the initial time (its particular value plays no role); $a(t)$ is the scale factor as a function of (physical) time. In the last equality, $\langle p_{\mathrm{DM}} \rangle$ and $\langle p_\nu \rangle$ are the average absolute values of momentum of DM particles and active neutrinos. The integral is saturated at early times when DM particles are relativistic ($v(t) \approx 1$).

The resulting picture is presented in Fig. 1.5, i.e. CDM forms structures of almost any size, WDM washes out small structures and HDM creates only large structures.

DM candidates can be classified as:

- *Cold dark matter*: particles that are created non-relativistic;

- *Warm dark matter*: particles that are initially relativistic, but became non-relativistic in the radiation-dominated epoch;

- *Hot dark matter*: particles that are still relativistic at the beginning of the matter-dominated epoch.

As discussed above in the context of SM neutrinos, Hot dark matter predict the top-down structure formation and is therefore excluded. Cold and Warm DM models have different predictions at the small scales only and therefore are *equally successful at large scales*. Both models correctly describe the data on large scale structure (e.g. galaxy-galaxy correlation functions), CMB, the properties of clusters of galaxies etc. [45, 46]. Only at the smallest observable scales one could try to see a difference between this two models [47, 48].

The work described in this thesis is mostly related to the attempts to distinguish between warm and cold dark matter. In the next section we describe a popular candidate for a warm dark matter particle – sterile neutrino (or heavy neutral lepton, HNL) as well as

its particle physics origin, context and motivation. In the Section 3 we describe how one of the pillars of modern cosmology – primordial nucleosynthesis – can be used to constraint properties of new particles and derive the most up to date bounds on the parameters of sterile neutrinos. Then we proceed with the discussion of how to distinguish CDM from WDM observationally. For this, we review the theory of structure formation in the Section 4. In the following Section 5 we introduce on of the promising approaches to distinguish between CDM and WDM – the Lyman-α forest method. After describing the method itself and its limitations, we review the data available by now, discuss in details possible interpretations of these data and main uncertainties related to this. At the end we present the constraints on warm dark matter and sterile neutrinos that can be derived from Ly-alpha forest.

Chapter 2

Sterile neutrinos and
Beyond the Standard Model problems

A simple and interesting extension of the SM is given by the so-called *sterile neutrinos*. All fermions in the SM can have left and right helicities – the projection of the angular momentum on the direction of motion. The only apparent exceptions are neutrinos – it was observed experimentally that only left handed participate in interactions. In the SM left handed neutrinos form a doublet with respect to $SU(2)$ gauge symmetry. A right handed neutrino, if added to the SM, would be neutral with respect to all main forces of the SM – electromagnetic, weak and strong. Naively, such a particle can not be involved in any interaction – it is "sterile" and as such could not be created and can not decay into SM particles. Therefore, right-handed neutrino are not included in the minimal version of the SM.

However, there is an interaction not prohibited by the gauge symmetries of the SM:

$$\Delta\mathcal{L} = F_{\alpha I}(\bar{L}_\alpha \cdot \tilde{H})N_I, \tag{2.0.1}$$

where L_α are SM lepton doublets, $\tilde{H} = i\sigma_2 H^*$ is the Higgs doublet in the conjugated representation and N_I are right-handed neutrinos with $I = 1, \ldots, n$.

If such an interaction exists, sterile neutrino can be involved in any process where normal neutrino interacts, but suppressed by Yukawa, see Fig. 2.1. In fact, this interaction not only mixes N with ν, but also makes neutrinos massive, in the same way as all fermion masses are generated in the SM. Observations of neutrino masses (together with the fact that in the SM they should be equal to zero, see a detailed discussion below) is a very interesting argument in favor of the existence or sterile neutrinos. To explain the smallness of masses of the SM neutrino (also called "active" neutrinos as opposed to the sterile ones), the right handed neutrinos should be relatively heavy, thus another name is "heavy neutral leptons", or HNLs. With three HNLs added, the SM contains left and right handed counterparts of all species of fermions and looks more "complete". Moreover, it appears that HNLs are capable not only to explain neutrino masses, but also give a mechanism of generation of
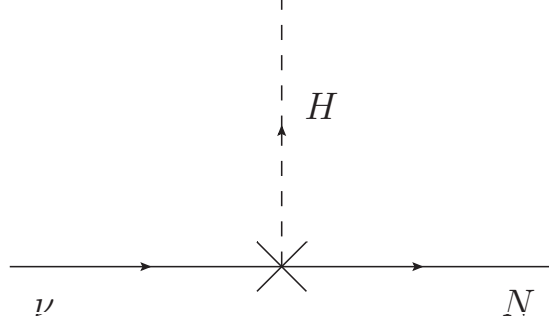
Figure 2.1: The interaction of HNLs N with SM neutrinos ν and the Higgs doublet H. After acquiring the Higgs VEV, the interaction becomes the mass mixing between N and ν.

matter-antimatter asymmetry of the Universe and a can be dark matter candidate.

In what follows we will discuss various ways to detects effects of HNLs on cosmological and astrophysical data (Chapters 4 and 3). Therefore below we present these particles and a minimal extension of the SM describing them in more details. We also discuss main particle processes where sterile neutrinos may be involved, as this will be used in the subsequent sections.

2.1 Heavy neutral leptons as a resolution of the Beyond the Standard Model problems

Lagrangian The fermion operator introduces n additional right-handed fermions – heavy neutral leptons – introducing with the SM through the gauge invariant operator $(\bar{L}_\alpha \cdot \tilde{H})$, where $L_\alpha, \alpha = 1, 2, 3$ is the lepton doublet and $\tilde{H} = i\sigma_2 H^*$ is the Higgs doublet in the conjugated representation. The general Lagrangian is

$$\mathcal{L}^c_{\text{neutrino portal}} = F_{\alpha I}(\bar{L}_\alpha \cdot \tilde{H})N_I + i\bar{N}_I \slashed{\partial} N_I - \frac{M_{N,I}}{2}\bar{N}^c_I N_I + h.c., \qquad (2.1.1)$$

where $I = 1, \ldots, n$.

After gaining the Higgs vacuum expectation value, the coupling $F_{\alpha I}$ provides the mass mixing between HNLs and active neutrinos. As a result of this mixing, HNL couples to the SM fields in the same way as active neutrinos,

$$\mathcal{L}_{int} = \frac{g}{2\sqrt{2}}W^+_\mu \bar{N}^c_I \sum_\alpha \theta^*_{\alpha I}\gamma^\mu(1-\gamma_5)\ell^-_\alpha + \frac{g}{4\cos\theta_W}Z_\mu \bar{N}^c_I \sum_\alpha \theta^*_{\alpha I}\gamma^\mu(1-\gamma_5)\nu_\alpha + \text{h.c.} ,$$
$$(2.1.2)$$

except the coupling is strongly suppressed by the small *mixing angles*

$$\theta_{\alpha I} = M^D_{\alpha I}M^{-1}_{N,I}, \quad (M^D)_{\alpha I} = vF_{\alpha I} \qquad (2.1.3)$$
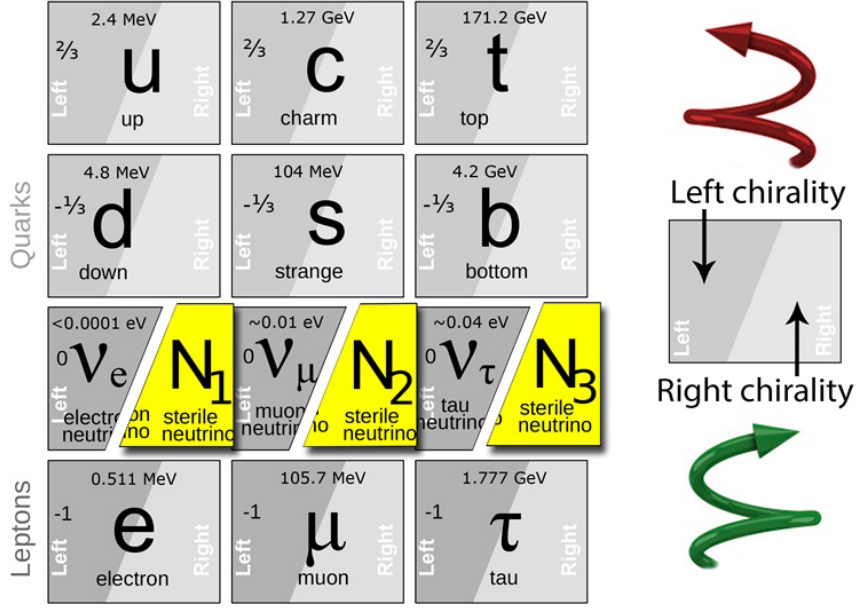
Figure 2.2: The Standard Model particles with three sterile neutrinos N_1, N_2, N_3. *Taken from [49].*

Fermion portal allows not only connecting the BSM to the Standard Model but also explaining the BSM problems without introducing additional fields.

Neutrino masses   SM neutrino masses m_{ν_α} manifest themselves experimentally through neutrino oscillations entering the oscillation length $l_{\text{osc}} = 4\pi p_\nu / \Delta m_{ij}^2$, where $\Delta m_{ij}^2 = m_i^2 - m_j^2$. In studying the oscillations of atmospheric and solar neutrinos there have been measured two mass differences [1]:

$$\Delta m_{\text{atm}}^2 \equiv \Delta m_{31}^2 = 7.55^{+0.2}_{-0.16} \cdot 10^{-5} \text{ eV}^2, \quad \Delta m_{\text{sol}}^2 \equiv \Delta m_{31}^2 \approx 2.5^{+0.3}_{-0.3} \cdot 10^{-3} \text{ eV}^2, \quad (2.1.4)$$

suggesting that at least two SM neutrinos are massive. A simple way to introduce masses of the SM neutrinos is by adding the Weinberg operator

$$\mathcal{L} = c_{\alpha\beta} \frac{(\bar{L}_\alpha^c \cdot \tilde{H})(\tilde{H} \cdot L_\beta)}{\Lambda}, \quad (2.1.5)$$

where Λ is some high-energy scale, L_α is the lepton SM doublet and the superscript c denotes the charge conjugation. The operator gives Majorana masses to SM neutrinos

$$(m_\nu)_{\alpha\beta} = -\frac{c_{\alpha\beta} v^2}{\Lambda}, \quad (2.1.6)$$

where v is the Higgs vacuum expectation value. Within the fermion portal, the Weinberg operator appears is the limit $||M_D|| \ll |M_N|$ of the Lagrangian (2.1.1), with

$$(m_\nu)_{\alpha\beta} = \sum_I M_{\alpha I}^D \frac{1}{M_{N,I}} M_{\beta I}^D \qquad (2.1.7)$$

The smallness of $M_{\alpha I}^D$ comparing to the Majorana masses $M_{N,I}$ in (2.1.1) naturally leads to the smallness of SM neutrino masses comparing to masses of HNLs and the electroweak scale. Such mechanism is called the *see-saw mechanism*. It does not fix both the parameters $F_{\alpha I}$, $M_{N,I}$, but only their combination in the form of neutrino mass (2.1.7).

Adding $\mathcal{N}$ new particles N_I to the Lagrangian $\mathcal{L}_{SM}$ adds

$$N_{\text{parameters}} = 7 \times \mathcal{N} - 3 \qquad (2.1.8)$$

new parameters to the Lagrangian. These parameters can be chosen as follows: $\mathcal{N}$ real Majorana masses M_I, *plus* $3 \times \mathcal{N}$ absolute values of Yukawa couplings $F_{\alpha I}$ *plus* $3 \times \mathcal{N}$ complex Yukawa couplings $F_{\alpha I}$ *minus* 3 phases absorbed in redefinitions of ν_e, ν_μ, ν_τ. The Pontecorvo-Maki-Nakagawa-Sakata matrix[1] plus three mass eigenstates m_1, m_2, m_3 of the active neutrino sector provide 9 parameters that can be determined experimentally. This shows that one needs $\mathcal{N} \geq 2$ to explain the neutrino oscillations by means of heavy neutral leptons.

If all the three SM neutrinos are massive, we need three HNLs.

Two HNLs in the broad mass range can explain the observed mass difference Δm_{atm}^2 and $\Delta m_{\text{solar}}^2$.

Baryogenesis. In the Early Universe the baryogenesis requires three conditions (the so-called Sakharov conditions) [50] to be satisfied. HNLs (2.1.1) are able to satisfy all the conditions:

1. *Baryon number violation*: HNLs violate the conservation of lepton number through the Majorana mass term and gives a possibility to generate non-zero baryon number (see details below).

2. *C and CP violation (C is already violated in the standard model)*: HNLs provide additional CP-violating phases (analogously to the CKM matrix in the quark sector) and thus CP violation in processes of the production and decay of HNLs.

3. *Out of equilibrium*: HNLs allow out-of-equilibrium processes involving HNLs if their coupling to the SM is small.

---

[1] The Pontecorvo-Maki-Nakagawa-Sakata matrix is a matrix of active neutrinos mixing in charged current weak interactions appearing because of the mismatch between the mass eigenstates and interaction eigenstates of neutrinos, similar to the CKM matrix for quarks.

HNLs of a wide mass range – from sub-GeV to 10^{15} GeV – can be responsible for the baryon asymmetry of the Universe due to, different mechanisms (see [51–53] and references therein). In particular, two GeV scale HNLs with nearly degenerate masses can produce the baryon asymmetry through their oscillations.

> The same two HNLs that provide masses to SM neutrinos can be responsible for the baryogenesis.

Dark matter   HNLs are the perfect candidate for the role of the dark matter. Their mass can be significantly large to generate the observed DM energy density of the Universe. From the other side, their interaction is similar to the interaction of SM neutrinos but suppressed by the small mixing angle. If the mixing angle is so small such that HNL DM never reaches the thermal equilibrium, their number density is smaller then for usual neutrinos and they do not violate the Tremaine-Gunn bound [28]. Simultaneously, for sufficiently small mixing angles and masses, HNLs have lifetimes comparable with the age of the Universe. The current constraint on the HNL DM is shown in Fig. 2.3. The HNL DM mass is limited from
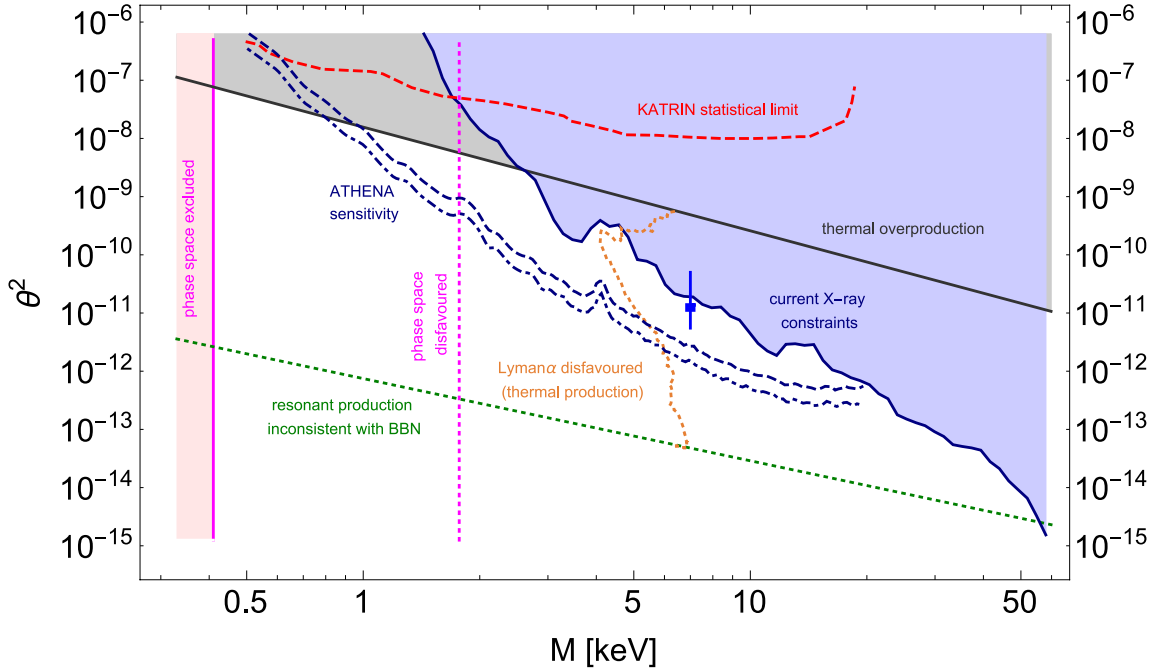
Figure 2.3: Constraints on HNL DM, see [54] for details.

below $M_N \gtrsim 1$ keV from the phase space density arguments [54]. HNL DM lifetime is strongly bounded from X-ray observations, while for small angles there is no production mechanism in this model. So HNL DM is expected to be a keV-scale particle.

Several years ago, an unidentified feature in the X-ray spectra of galaxy clusters [55, 56] as well as Andromeda [56] and the Milky Way galaxies [57] have been observed. The signal

can be interpreted as coming from the decay of a DM particle with the mass ~ 7 keV (in particular of sterile neutrinos with mixing angles in the range $\sin^2 2\theta \simeq (0.2 - 2) \times 10^{-10}$). The signal was confirmed in the spectra of galaxy clusters [58–60] or galaxies [61–64]. Other DM-dominated objects did not reveal the presence of the line [65–72]. This non-detection, however, did not exclude dark matter interpretation of the 3.5 keV line (see discussion in the review [54]).

The DM population of such HNLs can be produced thermally or resonantly, due to the enhancement of the mixing angle in a dense medium of SM plasma (see Sec. 2.2). The HNL dark matter is warm and decaying. We will return to these statements in detail in the next sections.

Long-lived and weakly interacting HNLs can be a perfect DM candidate. The observations indicate that masses of HNL DM candidate must be in keV range. However, such HNLs cannot be responsible for masses of active neutrinos. Indeed, the latter requires large mixing angles, for which there will be produced too much dark matter particles through the Dodelson-Widrow mechanism [73].

It is possible to explain all these three BSM phenomena by introducing three HNLs – one relatively light with the mass range of $m_N \simeq \mathcal{O}(\text{keV})$ responsible for the DM and two other with close masses in the $\mathcal{O}(\text{GeV})$ mass range responsible for the active neutrino masses and the baryogenesis. This model introduces 18 new parameters – 3 neutrino masses and 15 Yukawa couplings parameters. All these quantities can be measured experimentally, so the model is fully testable. This model is called the *Neutrino Minimal Standard Model* (νMSM) [51, 52].

2.2 Phenomenology of heavy neutral leptons

Production mechanisms of HNLs in the Early Universe   The main production channels of HNLs with masses $m_N \lesssim m_W$ in Early Universe are decays of W/Z bosons (if the Universe is hot, $T \gtrsim m_Z$) $W \to lN$, $Z \to \nu_l N$ and $2 \to 2$ processes $l/\nu_l f \to N f'$, where f, f' denote a fermion (either a lepton or a quark).

The production can be resonantly enhanced due to effects of medium. Namely, assuming that HNLs are ultrarelativistic, coherent forward scattering of HNLs on active neutrinos ν_l introduces a correction to the Hamiltonian describing evolution of N and $\bar{\nu}_l$,

$$H \approx \frac{m_N^2}{4E} \begin{pmatrix} -\cos(2U) & \sin(2U) \\ -\sin(2U) & \cos(2U) \end{pmatrix} + \begin{pmatrix} V_l & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.2.1}$$

where $V_l = -\sqrt{2}G_{\mathrm{F}}(-N_n/2 + N_{\nu_l} - N_{\bar{\nu}_l})$. This effect introduces effective mixing angle $\tan(2U_{\mathrm{eff}}) \approx 2U\Delta/(\Delta + V_l)$, where $\Delta = m_N^2/2E$. It is maximized if $V_l = -\Delta$, which leads to resonant enhancement of the HNL production.

Decays of HNLs. The main decay channels of HNLs with masses $m_N \lesssim 2m_e$ are decays into three neutrinos $N \rightarrow 3\nu$ and radiative decay into a photon and a neutrino, $N \rightarrow \gamma\nu$ (including charge conjugated channels), see Fig. 2.4. Heavier HNL decays into three leptons
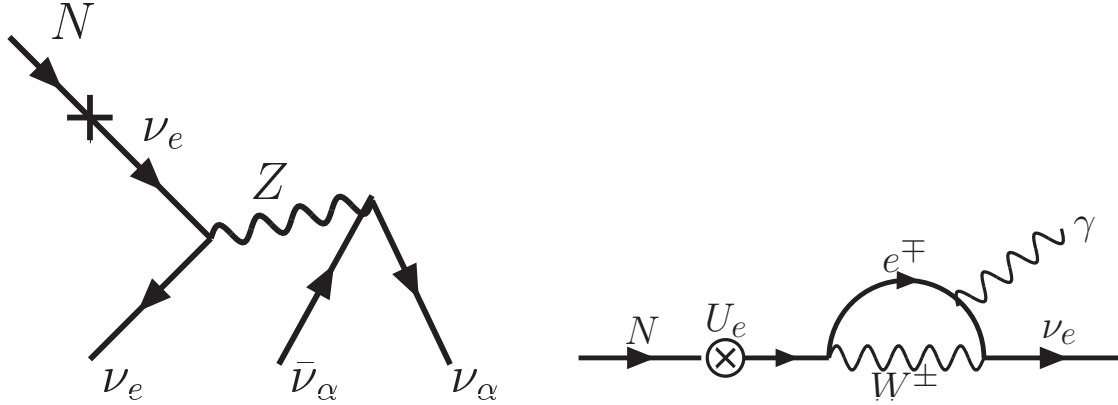
Figure 2.4: Decay channels of sterile neutrinos with masses $m_N \lesssim 2m_e$: decay into three neutrinos and into a neutrino and a photon. The figure is given from [49].

through charged and neutral weak currents, $N \rightarrow \nu_l \bar{l}'l'$ or $N \rightarrow \nu_l \bar{l}l'$. Once $m_N \gtrsim m_\pi$, semileptonic decays open. If $m_N \lesssim 1$ GeV, the decays are two-body, $N \rightarrow hl/\nu_l$, where h is a hadron. For heavy HNLs, $m_N \gtrsim 1$ GeV, decays into hadrons can be described by the three-body channels $N \rightarrow q_1\bar{q}_2l/\nu_l$ [74]. The branching ratios of HNLs in the range $m_N < 5$ GeV are shown in Fig. 2.5.
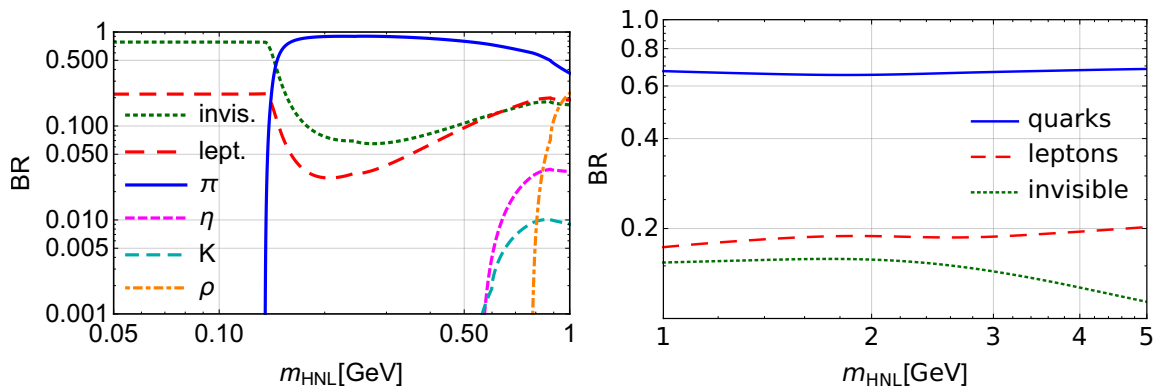
Figure 2.5: Branching ratios of HNL decays vs the HNL mass. Figures are given from [74].

Chapter 3

Big Bang Nucleosynthesis[1]

Big Bang Nucleosynthesis (BBN) is the earliest probe of the Universe that we currently have. It is sensitive to the physics that happened from ~ 1 to ~ 200 seconds after the Big Bang. The observable quantities are abundances of light elements that were fused during nuclear reactions in the Early Universe. They are sensitive to the presence of new particles that exist or decay during BBN.

In this chapter we describe the basic idea of the primordial nucleosynthesis (Section 3.1), discuss the prediction of the Standard Model (SM) with standard cosmology (Section 3.2) and compare it with observations (Section 3.3). Next, we discuss the effects of Beyond the Standard Model physics on BBN (Section 3.4) and specifically in the case of Heavy Neutral Leptons (HNL, see Section 3.5) that we would like to constrain. To do this we developed a numerical code that is described in Section 3.6. In the final section 3.7 we present our result and give conclusions.

3.1 Origin of chemical elements

Discovery of the constituents of atomic nuclei and their ability to combine into chemical elements posed a question: how the observed distribution of elements has been generated?

We know that many elements that are present at the Earth can be created by stellar evolution. However, there are regions in the Universe that are not much influenced by star formation (see detailed discussion in Section 3.3). Measurement of abundances of chemical elements in such regions indicates the existence of large amount of Helium-4 (about 25% mass fraction) and some other light elements like Deuterium, Helium-3, Lithium. The stellar origin of Deuterium is not probable because of its depletion in stars [75, 76]. Even more striking evidence is a measurement of Helium at the time of the decoupling of

---

[1]Results of this chapter are presented in papers [3, 4]. The main contribution of Andrii Magalich is the development of *pyBBN*, the numerical code that models primordial nucleosynthesis in presence of Heavy Neutral Leptons with masses above the pion mass that haven't been modelled previously. Treatment of chains of meson decays has been developed with major contribution by Nashwan Sabti.

cosmic microwave background [21]. The possible explanation of these measurements is the primordial origin of these elements.

> Most of the heavy chemical elements originate from stars. However, we have reasons to believe that some chemical elements are primordial – created in the Early Universe when it was hot and homogeneous.
>
> This poses a paradox: usually we imagine the Early Universe as being in equilibrium, but then the most abundant element would be Iron, because it has the largest binding energy per nucleon (Fig. 3.2).

We observe only light primordial elements, but not Iron. This means that nuclear reactions took place in the Early Universe, but were not in the thermodynamic equilibrium. It is possible if the rate of nuclear reaction Γ_{nuc} has the same order as the rate of expansion of the Universe H at the time of BBN and the conditions for nuclear reactions to be effective existed only for a relatively short period of time.

For the beginning of nuclear reactions the electric repulsion between protons has to be overcome. Strong interactions are extremely short-range, therefore either a very high density is needed, such that the distance between nucleons is of the order of characteristic scale of strong interactions (as it happens e.g. in neutron stars) or the temperature should be large enough, such that protons have enough kinetic energy to overcome electrostatic barrier and approach each other closely enough.

For the second scenario the temperature about $T \sim 10^9$ K ~ 100 keV is needed (this is a typical temperature in the cores of stars [77]). At such a temperature, the energy density of photons is very large. To estimate the concentration of baryons, Gamov assumed that a) it should be such that the reaction rate of nuclear reactions is of the order of the expansion rate of the Universe (see the discussion above) and b) the Universe is radiation dominated, as the density of baryons will be much lower. The latter assumption can be verified by the calculation (see below).

Assuming that the Universe is radiation-dominated, the Hubble rate can be estimated as

$$H \sim \frac{T^2}{M_{\text{pl}}} \tag{3.1.1}$$

The rate of nuclear reactions is given by

$$\Gamma_{\text{nuc}} = \langle \sigma_{\text{nuc}} v \rangle n_b, \tag{3.1.2}$$

where $\sigma_{\text{nuc}} \sim 10^{-29}$ cm^2 is a capture cross section of fast neutrons in hydrogen [77], $v \sim \sqrt{T/m_p}$ and n_b is a number density of baryons. From this one can easily estimate the baryon-to-photon ratio

$$\eta_b = \frac{n_b}{n_\gamma} \sim 10^{-10} \tag{3.1.3}$$

The Early Universe was radiation-dominated with baryon-to-photon ratio $\eta_B \sim 10^{-10}$.

3.2 Standard Model Nucleosynthesis

In this section, we will discuss the Standard Model BBN. We will start with a simplified qualitative picture, that is still able to predict helium abundance with precision $\sim 10\%$ (Section 3.2.1). Next, we will discuss a method on how to predict abundances of other light elements and how to estimate helium abundance more accurately (Section 3.2.2).

3.2.1 Simplified picture of Big Bang Nucleosynthesis

As we discussed in the previous section, the Early Universe was radiation-dominated ($\eta_b \sim 10^{-10}$). At the temperature $\sim 100 \, \text{keV}$ that is relevant for nucleosynthesis there are only e^-, γ, ν, p, n in the plasma. At some point protons and neutrons start to fuse into nuclei.

In the Early Universe, only nuclear reactions with 2 initial particles are effective, as the probability of 3-particle reactions is low. Also, we know that stable elements with atomic numbers $A = 5$ and $A = 8$ are absent, see Fig. 3.1. This results in the absence of elements with a mass number larger than 7 during BBN.
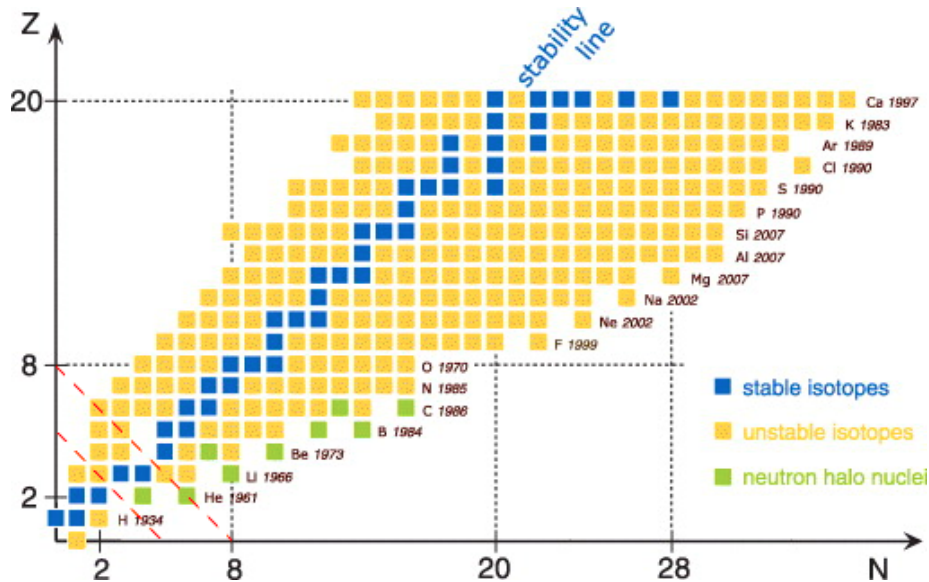
Figure 3.1: Chart of the nuclei stability. Here N is a number of neutrons, Z is a number of protons. Red dashed lines indicate the absence of a stable elements with atomic number $A = 5, 8$. Adopted from [78].

Production of the heavy elements is limited by the absence of stable nuclei with atomic numbers $A = 5$ and $A = 8$ – all elements heavier than Li^7 and B^7 are created through the 3-particle interactions in the dense medium of stars.

Among the lightest nuclei Helium-4 has the maximal binding energy per nucleon among light elements, see Fig 3.2. This means that the most probable reaction with light nuclei will be the formation of Helium-4, while reactions of helium dissociation have a much lower probability.
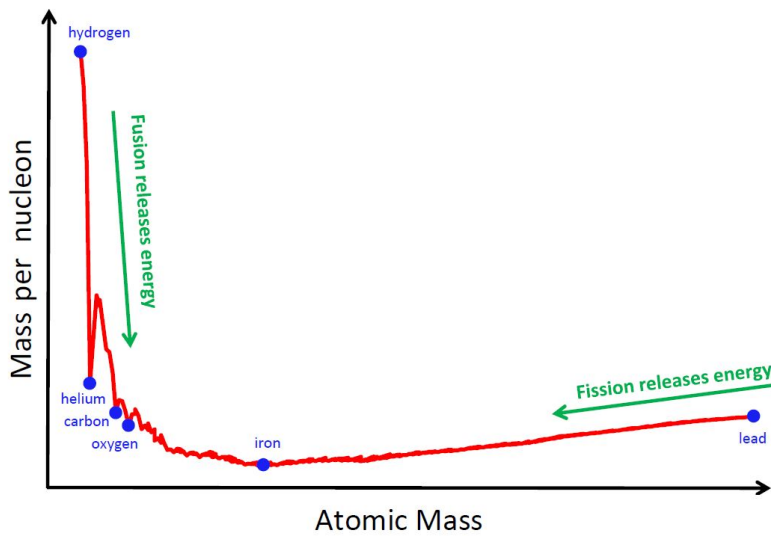
Figure 3.2: Mass per nucleon of stable nuclei. *Credit: oa-abruzzo.inaf.it*

Almost all neutrons and protons will fuse into Helium. The resulting amount of Helium is defined by the neutron-to-proton ratio at the start of the nuclear reactions.

Assuming that (almost) all available neutrons will be used to create helium, the mass abundance of Helium-4 can be estimated as

$$Y_{^4\text{He}} \approx \frac{4 m_p n_n / 2}{m_p n_p + m_p n_n} = \frac{2 \dfrac{n_n}{n_p}(t_{\text{BBN}})}{1 + \dfrac{n_n}{n_p}(t_{\text{BBN}})}, \tag{3.2.1}$$

where n_n and n_p are number densities of neutrons and protons, and t_{BBN} is a time of start of nucleosynthesis. Therefore, it is important to understand what is the value of neutron-to-proton ratio.

At large temperatures neutrons and protons were in thermal equilibrium because of weak reactions. In order to estimate $n_n/n_p(t_{\text{BBN}})$, we need to establish whether the neutrons are in equilibrium at t_{BBN}. Let us consider the effectiveness of weak $p \leftrightarrow n$ processes

$$e + p \leftrightarrow n + \nu, \quad \bar{\nu} + p \to e^+ + n \tag{3.2.2}$$

Before the decoupling density of neutrons or protons are given by thermal distribution,

$$n_A = g_A \left(\frac{m_A T}{2\pi}\right)^{3/2} e^{(\mu_A - m_A)/T}, \tag{3.2.3}$$

where

$$\mu_p + \mu_e = \mu_n + \mu_{\nu_e} \tag{3.2.4}$$

Electron chemical potential is negligibly small, since

$$\frac{\mu_e}{T} \sim \frac{\Delta n_e}{T^3} \sim \frac{n_p}{n_\gamma} \sim \eta_b \sim 10^{-9} \tag{3.2.5}$$

So neutron-to-proton ratio at the temperature of freeze out T_n is

$$\frac{n_n}{n_p}(T_n) \approx \exp\left(\frac{\mu_n - m_n}{T_n} - \frac{\mu_p - m_p}{T_n}\right) = \exp\left(-\frac{\Delta m}{T_n} - \frac{\mu_{\nu_e}}{T_n}\right), \tag{3.2.6}$$

where $\Delta m = m_n - m_p \approx 1.293$ MeV. In SM there is no reason to consider that $\mu_{\nu_e} \gg \mu_e$, so we take $\mu_{\nu_e} \approx 0$ and get

$$\frac{n_n}{n_p}(T_n) = \exp\left(-\frac{\Delta m}{T_n}\right) \tag{3.2.7}$$

The simple estimate of T_n is given by

$$\langle \Gamma_{n \to p} \rangle (T_n) \simeq H(T_n) = \frac{T_n^2}{M_{\text{Pl}}^*}, \quad M_{\text{Pl}}^* = \frac{M_{\text{Pl}}}{1.66\sqrt{g_*}} \tag{3.2.8}$$

where $g_* \approx 10.75$ is the number of ultra-relativistic DoF at $T_n \simeq$ MeV and $\langle \Gamma_{n \to p} \rangle$ is thermally averaged neutron conversion rate,

$$\langle \Gamma_{n \to p} \rangle = \langle \sigma_{ne^+ \to p\bar{\nu}_e} v_{ne^+} \rangle n_e + \langle \sigma_{n\nu_e \to pe} v_{n\nu_e} \rangle n_{\nu_e}, \tag{3.2.9}$$

where v_{ne^+} and $v_{n\nu_e}$ are relative velocities. Also we neglect neutron decays. Very rough simple estimate obtained assuming $T \gg \Delta m = m_n - m_p, m_e$ gives $\langle \Gamma_{n \to p} \rangle \simeq G_F^2 T^5$ and provides

$$T_n \simeq 1/(m_{\text{Pl}}^* G_F^2)^{1/3} \simeq 1.4 \text{ MeV} \tag{3.2.10}$$

More accurate estimate for the interaction rates [79], that keep $m_e, \Delta m$ and include elec-

troweak corrections, gives

$$T_n \approx 0.72 \text{ MeV} \tag{3.2.11}$$

This temperature is larger than $T_{\text{BBN}} \simeq 100$ keV, so we need to trace changing of the n-p ratio for $T \lesssim T_{\text{BBN}}$. The neutron-to-proton ratio at this temperature T_n is

$$\left.\frac{n_n}{n_p}\right|_{T_n} \approx \frac{1}{6}. \tag{3.2.12}$$

At high temperatures, neutrons and protons are in thermal equilibrium. Because of the expansion of the Universe, the weak interactions decouple from equilibrium. This interrupts the nucleon conversion and fixes the initial neutron-to-proton ratio.

After freeze-out the ratio (3.2.7) changes only due to neutrons decays and comoving number densities change as

$$n_n(t) = n_n(T_n)e^{-t/\tau_n}, \qquad n_p(t) = n_p(T_n) + n_n(T_n)\left(1 - e^{-t/\tau_n}\right), \tag{3.2.13}$$

where τ_n is neutron lifetime and t is a time after freeze out T_n. So, the neutron-to-proton ratio at the beginning of BBN is

$$\frac{n_n}{n_p}(t_{\text{BBN}}) = \frac{\exp\left(-\frac{\Delta m}{T_n}\right)e^{-t_{\text{BBN}}/\tau_n}}{1 + \exp\left(-\frac{\Delta m}{T_n}\right)\left(1 - e^{-t_{\text{BBN}}/\tau_n}\right)} \tag{3.2.14}$$

Because of the neutron decay, the neutron-to-proton ratio at the beginning of BBN is very sensitive to the time t_{BBN} when nuclear reactions become effective. This is called the *cosmic chronometer*.

3.2.1.1 Start of nuclear reactions

The chain of 2-body nuclear reaction for the light elements is shown in Fig. 3.3. We see that to start nuclear reaction Deuterium should be formed in the first place. However, it has low binding energy and cannot be abundantly generated in the Universe full of high energy photons that can disintegrate it. The relative number of baryons to photons is very low ($\eta_b \sim 10^{-10}$), so typically there are many photons with energy $E > \Delta_D$ for each nucleus even when $T < \Delta_D$ because of the high-energy tail of photon distribution.

One can estimate the time when the abundance of Deuterium becomes large as a time when the number of the high-energy photon approximately equal to the number density of baryons,

$$n_\gamma(E > \Delta_D) \sim n_b. \tag{3.2.15}$$
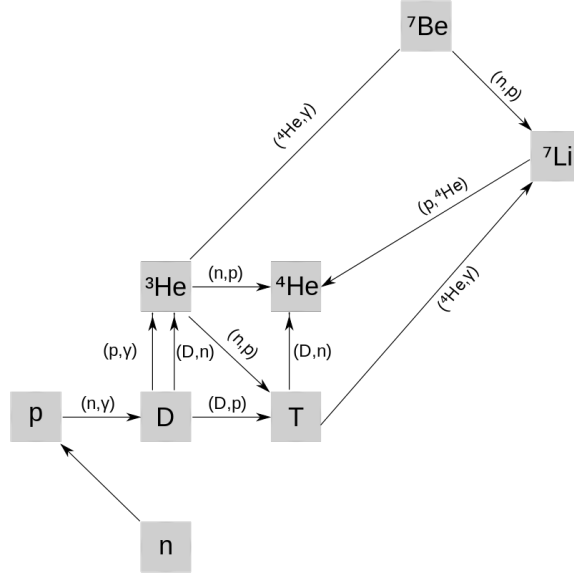
Figure 3.3: Nuclear reaction framework of nucleons and light elements. *Credit: www.aldebaran.cz*

Using Bose-Einstein distribution one can estimate number density of high-energy photons as

$$n_\gamma(E > \Delta_D) = 4\pi \int_{\Delta_D}^{\infty} \frac{p^2 dp}{e^{\frac{p}{T}} - 1} \approx 4\pi T^3 \int_{\Delta_D/T}^{\infty} e^{-x} x^2 dx = \tag{3.2.16}$$

$$= 4\pi T^3 \left(2 + 2\frac{\Delta_D}{T} + \frac{\Delta_D^2}{T^2} \right) e^{-\frac{\Delta_D}{T}} \approx 4\pi T \Delta_D^2 e^{-\frac{\Delta_D}{T}} \tag{3.2.17}$$

Hence, Deuterium becomes abundant starting from the temperature T_D satisfying

$$\eta_b \cdot \frac{\zeta(3)}{2\pi^3} \approx \frac{\Delta_D^2}{T_D^2} e^{-\frac{\Delta_D}{T_D}} \tag{3.2.18}$$

For $\eta_b = 10^{-10}$ this results in $T_D \approx 65\,\text{keV}$ (which is consistent with assumption $T_D \ll \Delta_D$).

To produce Helium (or any other element) through 2-particle interactions, nucleons need to go through Deuterium "bottleneck". Naively, its production should become effective at temperatures around the binding energy $\Delta_D = 2.2\,\text{MeV}$. However, Deuterium generation is challenged by a large amount of high-energy radiation. This delays nucleosynthesis until $T \sim 65\,\text{keV} \ll \Delta_D$

We are now ready to make a simple prediction for He abundance. The neutron-to-proton

ratio (3.2.14) is decreased to

$$n_n/n_p(t_{\text{BBN}}) \approx \frac{1}{6} \times e^{-(t(t_{\text{BBN}})-t(T_n))/\tau_n} \approx \frac{1}{7}. \qquad (3.2.19)$$

So, the primordial abundance of Helium-4 we can be estimated from Eq. (3.2.1) as

$$\boxed{Y_{\text{He}} \approx 0.227.} \qquad (3.2.20)$$

This simple estimate has uncertainty of only $\simeq 10\%$.

A simple analytic prediction of He abundance based on two numbers: neutron freeze-out temperature T_n and beginning of deuterium nuclei T_D gives Helium abundance $Y_{\text{He}} \approx 23\%$.

3.2.1.2 Neutron decay during Big Bang Nucleosynthesis

The intuitive picture that was used above for estimation of Helium abundance was that all available neutron at the beginning of BBN combines with protons to form Helium-4. This is correct if one can neglect neutron decay during BBN. Let us check that the rate of nuclear reaction is high in comparison to the rate of neutron decay, so the neutrons quickly become part of nuclei and do not decay.

At the start of the BBN the neutron decay competes with the neutron burning. Let us compare the decay rate

$$\Gamma_{n,\text{decay}} = 1/\tau_n \approx 1.1 \cdot 10^{-3} \text{ s}^{-1} \qquad (3.2.21)$$

with the rate of the nuclear reaction $\Gamma_{pn \to \gamma D}$. A simple estimate of the nuclear rate is given by

$$\Gamma_{pn \to \gamma D} = n_n \cdot \langle \sigma v \rangle \simeq \frac{2}{\pi^2} \frac{\eta_b}{7} \xi(3) T^3 \cdot \frac{\alpha}{m_\pi^2} \sqrt{\frac{T}{m_n}} \approx 6 \cdot 10^{-2} \left(\frac{T}{T_D} \right)^{7/2} \text{ s}^{-1}, \qquad (3.2.22)$$

where we considered

$$\langle \sigma v \rangle \simeq \alpha_{\text{EM}}/m_\pi^2 \sqrt{T/m_n} \approx 1.5 \cdot 10^{-25} \text{ m}^2/s. \qquad (3.2.23)$$

Here m_π^{-1} corresponds to the spatial range of strong interactions and α is the fine structure constant related to the creation of a photon, and used the neutron-to-proton ratio (3.2.19) and baryon-to-photon ratio to estimate $n_n = \frac{1}{7}\eta_b n_\gamma$.

Comparing (3.2.21) and (3.2.22), we conclude that $\Gamma_{pn \to \gamma D} \gg \Gamma_{n,\text{decay}}$, and therefore the most of neutrons are captured in D before decaying.

As soon as BBN starts, we do not need to consider the decay of neutrons since most of them are incorporated in nuclei.

3.2.1.3 End of nuclear reactions

In this section, we will discuss when nuclear reactions stop. First, we will consider the general mechanism of freeze-out of nuclear reactions because the rate of reaction becomes smaller than the Hubble rate. However, as nuclei are positively charged their fusion can be stopped because of the Coulomb repulsion, which is known as the Coulomb barrier, so we will consider this effect below as well.

Freeze-out of nuclear reactions Let us make of conservative estimate of the nuclear reaction decoupling considering some schematic nuclear reaction $A + B \rightarrow C + D$. Similarly to the $p + n \rightarrow D + \gamma$ process, the reaction rate can be roughly approximated as

$$\Gamma_{\text{nuclear}} = n_A \cdot (\sigma v) \simeq \frac{2}{\pi^2} \eta_b \xi(3) T^3 \cdot \frac{1}{m_\pi^2} \sqrt{\frac{T}{m_C}} \approx 10^{-16} \left(\frac{T}{1 \text{ MeV}} \right)^{7/2}, \qquad (3.2.24)$$

where

$$\langle \sigma v \rangle \approx \frac{1}{m_\pi^2} \sqrt{\frac{T}{m_A}} \approx 1.7 \cdot 10^{-6} \text{ MeV}^{-2} \sqrt{\frac{T}{1 \text{ MeV}}}, \qquad (3.2.25)$$

and we considered $n_A \simeq n_n$, $m_A \simeq m_n$.

Comparing the interaction rate (3.2.24) with the expansion rate of the Universe, we can determine the decoupling temperature:

$$\frac{T^2}{m_{\text{Pl}*}} = H \simeq \Gamma_{\text{nuclear}} \qquad (3.2.26)$$

The solution is given by

$$T_{\text{dec}} \simeq 1.9 \cdot 10^{-1} \text{ keV} \qquad (3.2.27)$$

So we expect that at lower temperature nuclear reactions stop.

Coulomb barrier Let us now estimate the temperature at which the nuclei reaction is not possible because of the Coulomb barrier. Due to Coulomb repulsion, a typical cross-section contains an exponential factor depending on nuclear charges [80, 81]

$$\langle \sigma v \rangle \propto e^{-\eta}, \quad \eta = \frac{Z_1 Z_2 \alpha_{\text{EM}}}{v(T)} \approx \frac{Z_1 Z_2 \alpha_{\text{EM}}}{\sqrt{T}} \cdot \frac{\sqrt{A_1 A_2}}{\sqrt{A_1} + \sqrt{A_2}} \qquad (3.2.28)$$

where Z is atomic number and A is mass expressed in atomic units.

As soon as this factor becomes small, reaction inevitably will freeze out. Then, for each pair of interacting nuclei we can compute a tentative interaction freeze-out temperature:

$$\eta \simeq 1 \Rightarrow T_{\text{Coulomb}} \simeq \frac{A_1 A_2 Z_1^2 Z_2^2}{(\sqrt{A_1} + \sqrt{A_2})^2} \text{ keV} \qquad (3.2.29)$$

Coulomb barrier kick-in temperatures for some reactions are given in Fig. 3.1. They are

$^{A_1}_{Z_1}\text{X}_1$	$^{A_2}_{Z_2}\text{X}_2$	T, keV
^1_1p	^2_1D	0.34
^2_1D	^2_1D	0.5
^2_1D	^3_1T	0.6
^3_1T	^4_2He	3.4
^7_3Li	^1_1p	4.7
^3_2He	^4_2He	13.8

Table 3.1: Coulomb barrier suppression temperature T for 2-body nuclear reactions with initial nuclei $^{A_1}_{Z_1}\text{X}_1$ and $^{A_2}_{Z_2}\text{X}_2$.

close to the scale of decoupling of nuclear reactions, but Coulomb barrier is especially important in the generation of heavier nuclei of Lithium and Beryllium.

For a complete understanding of when nuclear reactions fall out of equilibrium and stop, one has to consider a full cross-section for each reaction.

Nuclear reactions between charged nuclei cease because of the freeze-out of nuclear reactions at the temperature of about 0.2 keV. Some nuclear reaction freeze-out earlier because of the Coulomb barrier.

3.2.2 Predictions beyond ^4He

Above we have shown how to predict the abundance of Helium. What about other elements?

Before the start of BBN, the concentration of Deuterium is very low because of the high amount of radiation. Because of this, the rate of reactions involving Deuterium as the reactant is also low ($\Gamma \ll H$). The chain of nuclear reactions starts when Deuterium is generated in significant numbers ($n_D \sim n_n$). This system is not in thermal equilibrium and its dynamics can be described by the kinetic approach.

The abundance of Helium is easily estimated because of its high binding energy (using a simple estimate (3.2.20) one can get it with $\sim 10\%$ precision). To predict the abundances of other elements we need to use **kinetic Boltzmann equations** because rates of nuclear reactions do not exceed Hubble rate and nuclei are not in equilibrium. We write these equations for nucleons and nuclei using the laboratory information about the nuclear cross-sections.

Boltzmann equations for different particles. Naively, to describe the system kinetically, one would write all possible Boltzmann equations [82] including all possible interactions. However, electromagnetic interactions are sufficiently fast at all times, so all electrically charged particles are considered to be in dynamical (but not necessarily chemical) equilibrium. These particles have a thermal distribution that is fully characterized by temperature (equal to photon temperature) and chemical potential. Fast interactions equilibrate individual temperatures of the species with the temperature of radiation (photons). This fact greatly reduces the system of kinetic equations and the number of quantities to find solutions for.

Nuclei are heavy and charged. Because of EM interaction, they are in kinetic equilibrium, so their velocity distribution is given by Boltzmann distribution. We can integrate the Boltzmann equations for nuclei to get equations on number density n.

Then, the following system of equation arises:

$$\frac{\partial f_{\nu_i}(t, y)}{\partial t} - Hp\frac{\partial f_{\nu_i}}{\partial p} = I^f_{coll}\{f, n\} \tag{3.2.30}$$

$$\frac{\partial n_n(t)}{\partial t} + 3Hn_n = I^n_{coll}\{f, n\} \tag{3.2.31}$$

$$\frac{\partial n_X(t)}{\partial t} + 3Hn_X = I^n_{coll}\{f, n\} \tag{3.2.32}$$

These equations describe the evolution of neutrinos ν_i, neutrons n and nuclei $X = D, T, He, Li, \ldots$

Electrons and photons are in thermal equilibrium and we know their distribution functions. Nucleons and nuclei are in kinetic equilibrium with the medium, so it is enough to use for them an integrated Boltzmann equation for number density. As neutrinos freeze-out, we need to use the full Boltzmann equation for them.

Because of the tiny baryon-to-photon ratio, nuclei negligibly influence the cosmological expansion. This allows separating the nuclear reactions as a subsystem on the cosmological background.

To close the system of equations, one has to supply the Friedmann equation and condition of energy conservation:

$$H^2 = \frac{8\pi G\rho}{3} \tag{3.2.33}$$

$$\frac{d\rho}{dt} = -3H(\rho + P) \tag{3.2.34}$$

These equations contain the following independent variables: distribution functions of the particles that departed from equilibrium f_ν, number densities of the neutrons and nuclei n_n, X, temperature T and scale factor a.

3.2.2.1 Big Bang Nucleosynthesis codes

Kinetic equations described above together with an expansion of the Universe give us a closed system of equations that can be solved numerically. There are plenty of codes that compute the abundances of the chemical elements:

- KAWANO [83] (also known as NUC123; 1992) – modified version of the Fortran code by Wagoner (1972). Solves the nuclear reactions network with simple neutron-to-proton conversion rates in presence of lepton asymmetry and various cosmological parameters

- FASTBBN [84] (1999) and JAVA calculator [85] (1999) – simple codes designed primarily to impose bounds on N_{eff}.

- AlterBBN [86] (2011) – public rewrite of KAWANO. Designed to compute abundances in alternative cosmologies (extra dof, quintessence, etc).

- PArthENoPE [87] (2018) – Fortran code that builds on KAWANO and adds the following effects. Improved calculations for the neutron-to-proton reactions are implemented via new fits, not as effective corrections added *a posteriori*. They also include effects of finite nucleon mass and non-thermal neutrino spectral distortions.

- PRIMAT [79] (2018) – Mathematica code that claims to provide 10^{-4} Helium accuracy. Comes with precomputed neutron-to-proton rates in various assumptions: radiative, zerotemperature, corrections, finite nucleon mass corrections, finite temperature radiative corrections, weak-magnetism, and QED plasma effects, which are for the first time all included and calculated in a self-consistent way.

The codes all agree in the predictions of the SM BBN and give the following results for the mass abundances of Helium-4 and Deuterium:

$$Y_{\text{He}} \approx 0.247, \qquad \text{D/H} \approx 2.68 \cdot 10^{-5}. \tag{3.2.35}$$

3.2.3 Summary of Standard Model Nucleosynthesis

BBN predicts the existence of light primordial nuclei: mainly Deuterium, Helium-3/4, Lithium, and Boron. The abundance of Helium-4 depends only on the neutron-to-proton ratio at the time of the start of nuclear reactions. Simple estimates allow us to obtain this value with accuracy of $\sim 10\%$ (Sec. 3.2.1.1). To get better accuracy and predict the abundances of other elements, we need to use the kinetic approach (Sec. 3.2.2).

Standard Model predictions depend on the baryon-to-photon ratio which can be measured in some independent experiment – e.g. from CMB. Additional uncertainty comes from the contradictory measurements of the neutron lifetime [88]. However, the effect of this uncertainty on Helium abundance is quite small.

Standard nucleosynthesis theory gives specific predictions for abundances of elements with high precision. The only free parameter of the Standard BBN is the *baryon-to-photon* ratio η_b. It can be measured in some independent experiments – e.g. from CMB.

3.3 Observations

In this section, we want to discuss which observables can be deduced from the experiments and compare our theoretical predictions with measured values. Moreover, since any measurement carries also an error, available data instructs us to what precision it is required to do theoretical calculations.

3.3.1 Helium-4

Primordial helium abundance has to be measured in regions with no star formation. Below, we discuss 3 methods that utilize metal-poor extragalactic regions, intergalactic medium and CMB.

Why these targets are used to infer He abundance? At the time of formation of CMB, the Universe was homogeneous and did not contain any stars – hence no star formation by definition. Later the Helium abundance will be the closest to the primordial in metal-poor environments (by "metals", astronomers refer to all elements heavier than He), like HII regions (zones of ionized Hydrogen around hot stars that emit a lot of UV radiation) of metal-poor star-forming galaxies, and gas clouds observed in absorption against a distant quasar. These environments have a composition very close to primordial because they were virtually not affected by star formation. This is confirmed by the observation of low metallicity in these objects (100 times lower than in Sun), which is known by the extreme weakness of emission/absorption lines of metals in their spectra (dominated by H and He lines). If there was vigorous star formation occurring at some epoch in these objects, the metals lines observed in spectra, will be significantly stronger, which is not the case. Thus, extremely low metallicity indicates that these targets didn't undergo processes responsible for change of chemical composition, and that's why they are believed to have nearly original (primordial) composition.

3.3.1.1 Low-metallicity extragalactic method

Using the measurements in low-metallicity HeII regions, one can build the $Y - O/H$ relation. Extrapolating it to 0 metallicity one is able to compute the primordial Helium abundance [89, 90], see Fig. 3.4. This method gives [90] at 95% confidence

$$Y = 0.2551 \pm 0.0022 \tag{3.3.1}$$
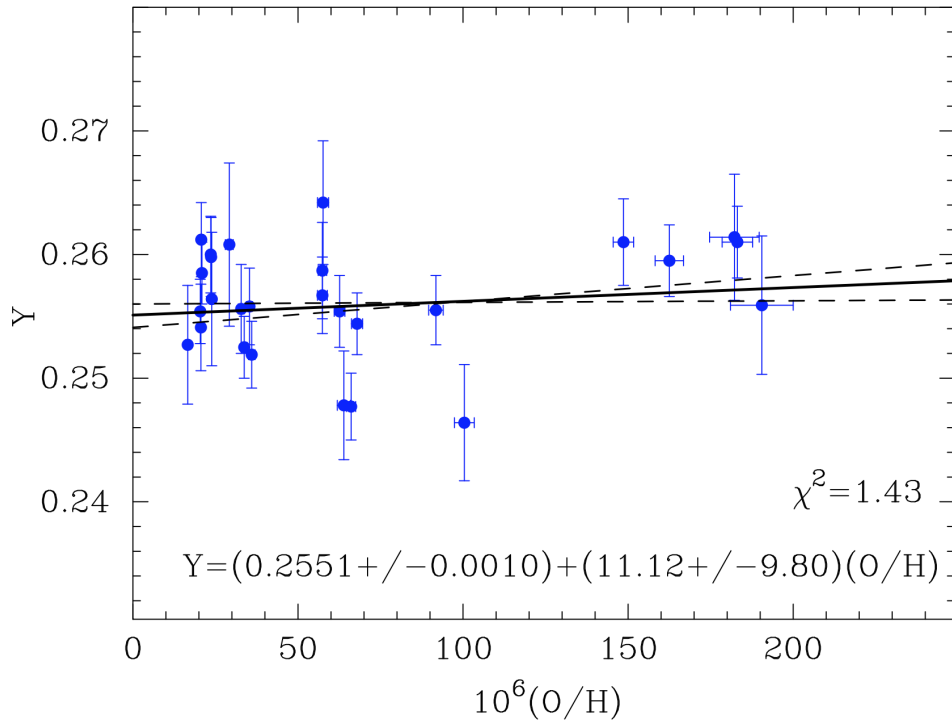
Figure 3.4: Helium abundance in 28 HII low metallicity regions as a function of Oxygen-to-Hydrogen ratio. The lines show the results of the linear regression to the data. The primordial value is measured by extrapolation of O/H to zero. *Credit: Izotov et al. [90]*

Although formal statistical error bar of this method are at sub-% level, the measurement is known to be dominated by the systematic uncertainties, see [91, Section 4.3] for a critical overview in this method.

The method of measuring Y_p from low-metallicity extragalactic regions formally gives very precise results ($\Delta Y/Y < 1\%$), but is expected to have large systematic errors.

3.3.1.2 Intergalactic medium method

The work [92] uses observational data of gas clouds with low metal content, which are observed against a quasar, located far behind the cloud (not in its vicinity). These clouds are composed mostly of Hydrogen and Helium, and very little of other elements. They are highly suitable for robust determination of the Helium abundance because the concentration of Hydrogen is high enough to produce strong absorption that can be detected, but at the same time not too high to distort the absorption line profile and make it impossible to extract exactly the H and He abundances. If the density is too high in the cloud, the atoms absorbing photons are colliding with other atoms before they de-excite. The lifetime of the electron in the excited state is reduced, and this yields a broader line profile due to the Heisenberg uncertainty principle, eventually saturating and distorting the absorption line

profile. Reconstruction of the column densities for saturated spectral lines is not reliable. On the other hand, at moderate densities, the atom is de-excited before colliding with another atom, and the resulting line profile has a pure Lorentzian shape determined only by the lifetime of the atom in the excited state ($\sim 10^{-8}$ s.) and the column density of the atom. Using pure Lorentzian profiles one can unambiguously and reliably find the column density of an element.

The neutral Hydrogen (HI) column density is inferred from the flux decrease beyond 91.2 nm wavelength (13.6 eV, the threshold energy for ionization of Hydrogen). The photons having shorter wavelengths are able to ionize Hydrogen and thus less of these photons will reach us. The flux decreased N_{obs}/N_{emit} is proportional to the column density of neutral Hydrogen n_{HI}:

$$N_{obs} = e^{-\tau} N_{emit}, \qquad \tau = L/\lambda, \qquad \lambda = \frac{1}{\sigma_{ion} n_{HI}}, \tag{3.3.2}$$

where τ is the *optical depth*, L is the distance and λ is the mean free path given by ionization cross section σ_{ion}. The Helium column density is retrieved from neutral Helium (HeI) absorption lines. The equivalent width of an absorption line (the area of the line divided by the continuum emission level) is a measure of how much flux was absorbed, and thus is proportional to the column density.

Important remark: since only neutral elements are seen in these spectra, it is necessary to estimate the ionization level of the cloud. The authors of [92] use simulations to illuminate gas with UV radiation imitating conditions (exposure due to quasars, galaxies, etc) in the observed cloud and extracting the corresponding ionization level.

The same method has been earlier applied to the determination of Deuterium [93].

$$Y = 0.250^{+0.033}_{-0.025}, \qquad D/H = (2.527 \pm 0.030) \cdot 10^{-5} \qquad \text{(68\% confidence)} \tag{3.3.3}$$

> The IGM method studies the regions of lower metallicity than the extragalactic method (by $\sim 30\%$). These regions are supposedly closer to the primordial composition. However, this method has a much larger statistical error of $\Delta Y/Y \sim 10\%$.

3.3.1.3 Cosmic Microwave Background method

Planck collaboration [21] provide a Helium measurement based on the free electrons density between Helium and Hydrogen recombination. The damping tail of CMB anisotropies is sensitive to the electron density and it is possible to measure this effect when Helium is already recombined ($z \sim 2000$), but before Hydrogen recombination ($z \sim 1100$).

This effect of Helium abundance is however partially degenerate with the effective

number of relativistic degrees of freedom N_{eff}, requiring a simultaneous fit:

$$Y = 0.246 \pm 0.035 \qquad (3.3.4)$$

$$N_{eff} = 2.97^{+0.58}_{-0.54} \qquad (3.3.5)$$

at 95% confidence interval, using Planck TT,TE,EE+lowE+lensing+BAO [21, Section 7.6.2].

Measurement through CMB guarantees that the Helium abundance is not affected by stellar nucleosynthesis. The error of this method is about $\Delta Y/Y \sim 14\%$. However, Planck's collaboration emphasizes those systematics of the polarization spectra have not been accurately characterized.

3.3.2 Deuterium

Because of its small binding energy, Deuterium is not created by other sources and is destroyed in stars [94]. This means that any measurement of Deuterium is guaranteed to be not higher than the primordial value. The local interstellar value of D/H ratio is $D/H = (1.56 \pm 0.40) \cdot 10^{-5}$ [95].

High-resolution absorption spectra of quasars allow measuring Deuterium in Intergalactic Medium environments with metallicities $\sim 10^3$ times smaller than Solar. Since the Lyman-series transitions of different elements are shifted, this is in principle can be used to identify absorption lines of two elements like Hydrogen and Deuterium (or Helium) and to deduce the abundance from the relative intensities. In reality, this, however, is difficult because of contamination by Hydrogen Lyman-α forest lines and requires accurate knowledge of the HI column density corresponding to absorbers.

Damped Lyman-α systems corresponding to dense regions in high-redshift galaxies allow to measure the HI density independently of the cloud model and there exists a number of systems where Deuterium line was identified and D/H abundance was measured:

$$D/H = (2.569 \pm 0.027) \cdot 10^{-5} \qquad (3.3.6)$$

This method is also used as an independent probe for ^{4}He.

Any measurement of Deuterium provides an upper bound on primordial abundance.

3.3.3 Helium-3

There is no reliable data on the primordial abundance of Helium-3. This isotope is only measured in the Solar system and HII regions of the Milky Way (that also have high metallicity) [96]. At this point, it is not possible to make conclusions about the effect of

stellar nucleosynthesis on the Helium-3 abundances since theoretical models are inconsistent with observations [97].

At the moment there is no ^{3}He data available to constrain BBN.

3.3.4 Lithium

Lithium is best measured in the Population II metal-poor stars of the Milky Way. It is notable that Lithium abundance exhibits the *Spite plateau* – is nearly constant in stars with metallicity $\lesssim 1/30$ of the Solar value [98]. Considering this abundance as a primordial one gets

$$\mathrm{Li/H} = (1.6 \pm 0.3) \cdot 10^{-10} \tag{3.3.7}$$

However, in extremely metal-poor stars with $\mathrm{Fe/H}$ ratio $\lesssim 10^{-3}$ than Solar, no Lithium is detected. This suggests some mechanism destroying ^{7}Li as well as that the abundance at the Spite plateau might also be modified from the primordial value. This, in fact, makes the method of 0-metallicity extrapolation inapplicable to Lithium. Instead, the measured abundance (3.3.7) should be considered a lower bound on the primordial value [99].

Recent observations indicate that Lithium might be destroyed in low metallicity stars and hence the measurements constitute the lower bound on primordial ratio ^{7}Li/H.

3.3.5 Comparison of measurements with Standard Model Nucleosynthesis predictions

Since Standard BBN predictions depend only on measurements of the baryon-to-photon ratio and neutron lifetime, abundances of the elements can be used to check the validity of the theory.

Baryon-to-photon ratio can be independently measured from CMB [100]:

$$\eta_B = (6.118 \pm 0.041) \cdot 10^{-10} \tag{3.3.8}$$

Fig. 3.5 summarizes the available measurements and predictions. This figure shows long-standing Lithium Problem: Standard BBN theory predicts Lithium abundance ~ 3 times larger than measured in metal-poor stars. However, new observational data of Lithium show that the previous measurements should be considered as a lower bound on the primordial abundance, so the Standard BBN theory is consistent with observations.
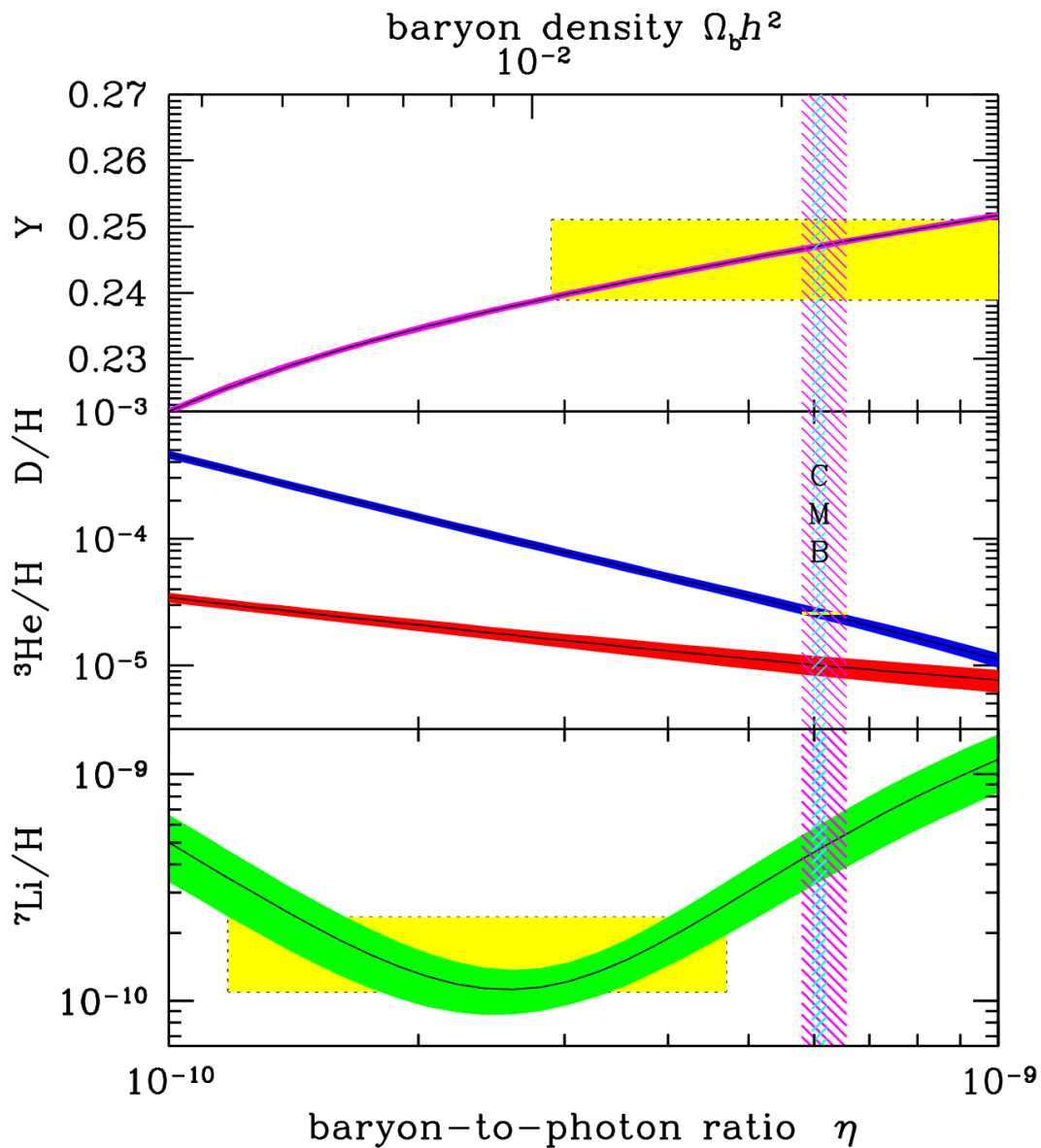
Figure 3.5: Predictions of the Standard BBN theory (with uncertainty due to error in neutron lifetime τ_n; black curves with coloured bands) compared to the measurements of primordial elements (yellow rectangles) and determination of the baryon-to-photon ratio η_B from CMB (vertical crosshatched band). *Credit: Cyburt et al. [101]*

Theoretical predictions of Standard BBN are in good agreement with the modern measurements of baryon-to-photon ratio and element abundances.

The only significant deviation is with the measurement of Lithium which is lower than predicted value by a factor of ~ 3. However, recent observations indicate that Lithium might be destroyed in low metallicity stars and hence the measurements

constitute the lower bound on primordial ratio $^7\mathrm{Li}/\mathrm{H}$.

Any Beyond the Standard Model physics should not break this agreement.

3.4   New physics and Big Bang Nucleosynthesis

As we have seen above, the predictions of the SM nucleosynthesis are consistent with observations. This puts a requirement on any extension of the Standard Model – its predictions for nuclear abundances should be different from the SM ones only within observational error bars. This makes BBN a powerful tool to constrain new physics.

Possible effects of new physics on BBN include:

- Change of Hubble expansion rate during BBN

- Change of neutron-to-proton ratio at and after freeze-out of weak interactions

- Disintegration of nuclei due to collisions with the new particle or its decay products

- Entropy production: the creation of additional photons or electrons between BBN and CMB and, therefore a change of the baryon-to-photon ration at BBN time (as compared to the value observed e.g. by CMB).

3.4.1   Excluded domain of the parameter space

Typically a new particle is characterized by its mass M and coupling constant θ. Then the lifetime of the particle is inversely proportional to it: $\tau \propto \theta^{-2}$. This naturally gives us a bound when $\tau \ll 1sec$ – these particles vanish long before the BBN. On the other hand, there also exists a lower bound on the particle coupling, since the interaction rate becomes negligible.

New physics particles (NPPs) can contribute to each of the effects of new physics on BBN considered above. As lifetime is related to the coupling constant as $\tau \propto \theta^{-2}$, it is convenient to separate the effects of independence on the lifetime τ_X of NPP.

3.4.1.1 $\tau_X \gtrsim 0.1\,\mathrm{s}$

Sufficiently long-lived non-relativistic NPPs can contribute a major part to the energy density, thus changing the Universe from radiation dominated to matter-dominated during the BBN.

Ultrarelativistic decay products of NPPs can contribute to the energy density, thus increasing N_{eff}, neutron decoupling temperature and the neutron-to-proton ratio.

Weakly interacting particles (leptons) produced in decays of NPPs participate in *weak interactions* involving neutron and are able to decrease the neutron decoupling temperature, thus decreasing the neutron-to-proton ratio.

Heavy NPPs can decay into mesons, which can convert $p \to n$ in *strong interactions* before decaying themselves, e.g.

$$\pi(K) + p \to n + \pi(K)/\gamma \tag{3.4.1}$$

and therefore increase the neutron-to-proton ratio.

3.4.1.2 $\tau_X \gtrsim t_{\text{BBN}}$

Muons and mesons from decays of heavy NPPs, $m_X \gtrsim m_\mu$, are able to dissociate primordial nuclei, e.g.

$$\pi +^4 \text{He} \to D + 2n, \quad \pi + D \to 2n \tag{3.4.2}$$

directly changing the primordial abundances.

3.4.1.3 $\tau_X \gtrsim 10^4$ s

If NPPs survive down to small temperatures $T_{\text{photodiss}} \simeq$ few keV, the photons produced in its decay or in decays of its daughter particles are able to dissociate primordial nuclei.

$T_{\text{photodiss}}$ can be estimated from the requirement that the maximally possible energy of photons in SM plasma, $E_{\gamma,\text{max}} \approx m_e^2/22T$ [102],[2] becomes comparable with the binding energy of the nuclei. For example, for D and ^{4}He we have

$$E_{\gamma,\text{max}} \simeq \begin{cases} \Delta_D \approx 2.2\,\text{MeV} \to T \simeq 5\,\text{keV}, \\ \Delta_{^4\text{He}} \approx 20\,\text{MeV} \to T \simeq 0.6\,\text{keV}, \end{cases} \tag{3.4.3}$$

This translates for the requirement $\tau_X \gtrsim 5 \cdot 10^4$ s for D dissociation and $\tau_X \gtrsim 10^6$ s for ^{4}He dissociation.

Different effects of new particles on BBN can be classified by the lifetime of a new particle. Namely, short-lived particles that decay before decoupling of weak interactions do not produce observable effects. Vice versa, particles produced in small numbers and very long-lived influence BBN negligibly. Therefore constraints from BBN usually have an upper and lower bound by particle's lifetime.

Below we provide several examples of modifications to the BBN by new physics.

Example 1: additional relativistic particles. Assume that there exists additional neutrino species. Then the number of effective degrees of freedom change as $N_{\text{eff}} \to N_{\text{eff}} + \Delta N_{\text{eff}}$. This effect increases a Hubble rate $H = T^2/M_{\text{Pl}}^* \propto g_*$, where g_* depends on

---

[2]The cut-off is determined by the process $\gamma + \gamma_{\text{SM}} \to e^+ + e^-$.

N_{eff} as

$$g_* = 2 \cdot N_\gamma + \frac{7}{8} \cdot 2 \cdot 2 \cdot N_e + \frac{7}{8} \cdot 2 \cdot \left(\frac{T_\nu}{T_\gamma}\right)^4 \cdot N_{\text{eff}} \approx \begin{cases} 10.83 + 1.75\Delta N_{\text{eff}}, & \text{before annihilation} \\ 3.38 + 0.45\Delta N_{\text{eff}}, & \text{after annihilation} \end{cases}$$
(3.4.4)

Here number of photons $N_\gamma = 1$, effective number of neutrinos in SM $N_{\text{eff}} = 3.046$ [103]. Before electron-positron annihilation ($T_\gamma \simeq m_e$) $N_e = 1$, $T_\nu = T_\gamma$. After electron-positron annihilation $N_e = 0$, $T_\nu = \left(\frac{4}{11}\right)^{1/3} T_\gamma$.

Let us estimate effect of faster Universe expansion on T_n. From Eq. (3.2.8) we have a condition

$$\langle\Gamma_{n\to p}\rangle(T_n) = H(T_n) \approx 1.66\sqrt{g_*}\frac{T_n^2}{M_{\text{Pl}}}.$$
(3.4.5)

Using approximate temperature dependence $\langle\Gamma_{n\to p}\rangle(T_n) \sim T_n^5$ we get

$$T_n \sim (g_*)^{1/6} \quad \Rightarrow \quad \frac{\Delta T_n}{T_n} \approx \frac{\Delta g_*}{6g_*}$$
(3.4.6)

Using (3.4.6) and simple analytic estimates (3.2.14), (3.2.1) for $n_n/n_p(T_{\text{BBN}})$ and $Y_{^4\text{He}}$, we can constrain the value ΔN_{eff} from the Helium abundance. Indeed, a shift ΔT_n changes $n_n/n_p(T_{\text{BBN}})$ and, correspondingly, the $Y_{^4\text{He}}$. Corresponding corrections can be found expanding the expressions (3.2.14), (3.2.1) in series on $\Delta T_n/T_n$:

$$\Delta Y_{\text{He}} \approx \Delta\left(\frac{n_n}{n_p}\right)\frac{Y_{\text{He,SM}}^2}{2(n_n/n_p)_{\text{SM}}^2}, \quad \Delta\left(\frac{n_n}{n_p}\right) \approx e^{t_{\text{BBN}}/\tau_n + \Delta m/T_n}\left(\frac{n_n}{n_p}\right)_{SM}^2 \cdot \frac{\Delta m \Delta T_n}{T_n^2}$$
(3.4.7)

Plugging in the numbers $T_n \approx 0.716$ MeV, $(n_n/n_p)_{\text{SM}} \approx 1/7$, $Y_{\text{He,SM}} \approx 0.227$, $t_{\text{BBN}} \approx 200$ s corresponding to analytic estimates, we obtain

$$\Delta(n_n/n_p) \approx 0.28\frac{\Delta T_n}{T_n}, \quad \frac{\Delta Y_{\text{He}}}{Y_{\text{He}}} \approx 1.54\frac{\Delta T_n}{T_n} = 0.024\Delta g_* = 0.041\Delta N_{\text{eff}}$$
(3.4.8)

Requiring $\Delta Y_{\text{He}}/Y_{\text{He}} < 0.05$, we get

$$\Delta g_* < 2.08, \quad \Delta N_{\text{eff}} \lesssim 0.82$$
(3.4.9)

Assuming the relative error of Helium to be $\Delta Y_{He}/Y_{He} \sim 5\%$, the maximal allowed $\Delta N_{\text{eff}} = 0.82$

Example 2: additional energy density. Consider the NPP that was in thermal equilibrium, decoupled being relativistic but being non-relativistic during the BBN (i.e.,

$m_X \gg T_n$). We estimate the corresponding energy density as

$$\rho_X \approx m_X Y_X s \approx 2 \cdot 10^{-4} \frac{Y}{4 \cdot 10^{-4}} g_*(T) e^{-t/\tau_X} T^3 m_X \qquad (3.4.10)$$

The ratio of ρ_X to the energy density of the SM plasma is

$$\rho_N/\rho_{\text{SM}} \approx 3.1 \cdot 10^{-4} \frac{Y_X}{4 \cdot 10^{-4}} e^{-t/\tau_X} m_X/T(t) \qquad (3.4.11)$$

with the maximal value

$$\rho_X/\rho_{\text{SM}}\big|_{\text{max}} \simeq \rho_X/\rho_{\text{SM}}\big|_{t \simeq \tau_X} \approx 0.4 \frac{Y_X}{4 \cdot 10^{-4}} (m_X/1\,\text{GeV}) \sqrt{\tau_X/1\,\text{s}}, \qquad (3.4.12)$$

where we assumed radiation dominated time-temperature relation. Having the parametric dependence of Y, τ_X on the mass of NPPs and its coupling to the SM, we can impose a constraint using the requirement

$$\rho_X/\rho_{\text{SM}} \ll 1 \qquad (3.4.13)$$

3.5 Big Bang Nucleosynthesis in the presence of Heavy Neutral Leptons

We are interested in HNLs that can generate the masses of the SM neutrinos – in particular, the observed mass difference $\sqrt{\Delta m_{\text{atm}}^2} \approx 0.05$ eV. No HNLs heavier than $m_N \gtrsim 1$ GeV with lifetimes $\tau_N \gtrsim 0.1$ s are able to provide such difference.

Indeed, to provide the observed mass difference, the mixing angle must be larger than the see-saw bound [49]:

$$\theta^2 \gtrsim \theta_{\text{see-saw}}^2 \simeq \sqrt{\Delta m_{\text{atm}}^2}/m_N \approx 5 \cdot 10^{-11} \left(\frac{1\,\text{GeV}}{m_N}\right) \qquad (3.5.1)$$

The comparison of the parameter space of HNLs with the lifetimes $\tau_N = 0.1$ s, $\tau_N = t_{\text{BBN}} \simeq 200$ s with the bound (3.5.1) is shown in Fig. 3.6. We see that the see-saw bound intersects the line $\tau_N = 0.1$ s at $m_N \simeq 1$ GeV. Based on the figure, we conclude that for heavy enough HNLs, $m_N \gtrsim m_\pi$, the only possible effects above the see-saw bound are:

- change of the Hubble expansion rate

- decrease of neutron decoupling temperature due to injection of weakly interacting particles

- increase of n-p ratio due to injection of muons or mesons
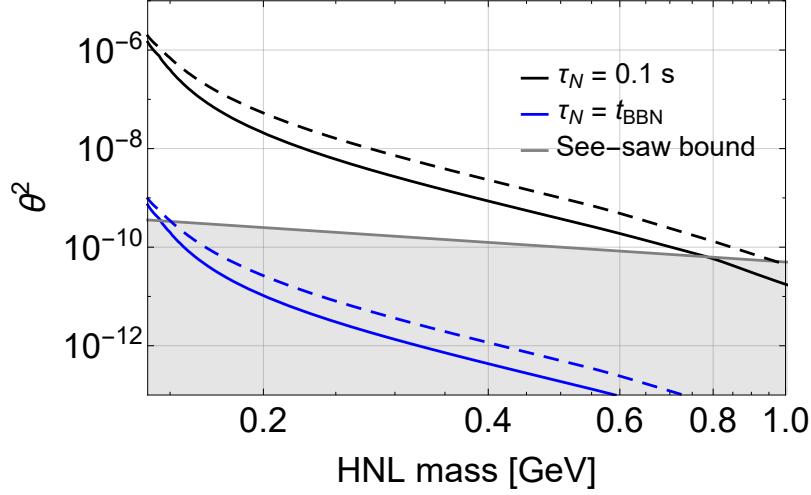
Figure 3.6: Parameter space of HNLs with lifetimes $\tau_N = 0.1$ s (black line) and $\tau_N = t_{\text{BBN}}$ (blue line); solid lines corresponds to the mixing with ν_e, while dashed lines – to the mixing with ν_τ. The shaded gray region corresponds to the parameter space excluded if we consider HNLs that are able to provide observable mass difference of SM neutrinos $\sqrt{\Delta m_{\text{atm}}^2} \approx 0.05$ eV.

HNLs lighter than pion mass can also affect the BBN at the later stages.

To account for the influence of HNLs we make the following changes in the equations of BBN:

- HNLs introduce additional Boltzmann equations (on f_{HNL} and their decay products – muons, pions and heavier mesons) and corresponding matrix elements

- Processes with HNLs or their decay products should be added to other Boltzmann equations

- HNLs and their decay products have pressure and energy density that modify the law of expansion

    At this point, we do not consider the effect of the increase of n-p ratio due to the injection of muons and mesons.

3.6 Description of numerical methods and code

The physical system of BBN consists of the particles, some of which are in thermodynamical equilibrium with each other and some of which are not. All properties of the equilibrium plasma are given by the evolution of temperature. Non-equilibrium particles (neutrinos,

nucleons, and nuclei) evolve with the expansion of the Universe (according to Friedmann equations) and particle collisions (Boltzmann equations).

Hence the complete set of unknowns in our system of equations is:

- plasma temperature $T(t)$

- distribution functions of neutrinos $f_{\nu_i}(t, p)$

- densities of nucleons and nuclei $n_X(t)$

- distribution function of sterile neutrino and its decay products $f_N(t, p)$, $f_{\text{product,i}}(t, p)$

In the cosmological setup it is convenient to use the scale factor $a(t)$ as a parametrization of time.

Equilibrium density is defined by the temperature, while non-equilibrium density is given by distribution functions. Their evolution is given by Friedmann and Boltzmann's equations. Below we will discuss our method of their numerical solution.

Temperature evolution equation. The energy conservation equation can be used to derive the temperature evolution of the plasma. We can split the total energy density into equilibrium and non-equilibrium parts $\rho = \rho_{\text{eq}} + \rho_{\text{noneq}}$. Then the derivative of ρ_{eq} is given by $\dot{T}$ while the time evolution of non-equilibrium density depends on the Boltzmann collision integrals:

$$\dot{\rho} = \left(\frac{d\rho_{\text{eq}}}{dT} \frac{dT}{dt} + \frac{d\rho_{\text{noneq}}}{dt} \right) = -3H \left(\rho + P \right)$$

$$\frac{dT}{dt} = -\frac{3H(\rho + P) + \frac{d\rho_{\text{noneq}}}{dt}}{\frac{d\rho_{\text{eq}}}{dT}} \tag{3.6.1}$$

where P is the total pressure and H is the Hubble rate.

Expressions for ρ_{eq} and ρ_{noneq} can be substituted in this equation to obtain an explicit formula for the temperature evolution. This is done in detail in Appendix 3.A.

Comoving coordinates Because of the expansion of the Universe, even non-interacting particles evolve non-trivially. This significantly complicates the computations and analysis.

According to Friedmann equations, particle positions and momenta behave as $\vec{r} \propto a(t)$ and $\vec{p} \propto a^{-1}(t)$. From this it is easy to conclude that particle densities $n \propto a^{-3}(t)$. Combining the latter two facts, we also find that the energy density of relativistic particles necessarily $\rho_{rel} \propto a^{-4}(t)$ while for non-relativistic ones $\rho_{nonrel} \propto m_{nonrel} a^{-3}(t)$.

It is convenient to switch to the *comoving* frame of reference that is expanding along with the Universe:

$$\vec{r} \rightarrow \vec{r}/a \tag{3.6.2}$$

$$\vec{p} \rightarrow \vec{y} = \vec{p} \cdot a \tag{3.6.3}$$

In this frame, many *comoving* quantities (denoted by ˜) become constant with respect to the expansion (e.g. photon energy density $\tilde{\rho}_\gamma \equiv \rho_\gamma a^4 = \text{const}$).

A huge additional advantage is a fact that in comoving coordinates the Boltzmann equations lose the term with the Hubble rate:

$$\frac{df(t, p(t))}{dt} = \frac{\partial f}{\partial t} - Hp\frac{\partial f}{\partial p} = aH\frac{\partial f(a, y)}{\partial a} = I_{coll} \tag{3.6.4}$$

$$\frac{\partial f(a, y)}{\partial a} = \frac{I_{coll}}{aH} \tag{3.6.5}$$

This change of variables also suggests rewriting the temperature evolution equation for the quantity $\tilde{T} \equiv a \cdot T$. For relativistic particles, temperature evolves with expansion as $T \propto a^{-1}$. Then $\dot{\tilde{T}}$ describes the change of entropy of the system due to transitions in the plasma (like electron-positron annihilation) or influence of non-equilibrium particles. Finally, we introduce the comoving mass $\tilde{m} = m \cdot a$ for particles.

BBN physics formulated in terms of comoving positions, momenta and masses look in a sense very similar to physics in non-expanding Universe, but with interaction rates decreasing with time.

The evolution of physical quantities even in the simplest cases is non-trivial because of the expansion of the Universe. However, if we consider the *comoving* frame (expanding with the Universe), we can significantly simplify the description for numerics.

3.6.1 pyBBN: code for non-standard nucleosynthesis

Our goal is to put constraints on the Heavy Neutral Leptons from the primordial nucleosynthesis. This task has been to some extent done by [104–106]. Papers [107, 108] also consider relevant aspects of the influence of new physics on the BBN (like entropy injection). This is, however, not sufficient to describe all effects of HNLs (e.g., HNL decays produce non-equilibrium neutrinos that can directly influence the weak reactions of nucleons).

Code by [106] is based on [105] and implements the physics of HNLs with masses up to $M < m_\pi$. We extend this analysis to larger masses $M < m_\phi$ in a new code *pyBBN*[3] [3]. Our main contribution is the treatment of hadronic decay channels of HNLs and incremental improvements and fixes to the approach of [106].

---

[3]https://github.com/ckald/pyBBN

General structure. Simulations are done in two stages:

1. The cosmological background physics and the rates of the reactions

$$n + e^+ \leftrightarrow p + \bar{\nu}_e, \quad n + \nu \leftrightarrow p + e, \quad n \rightarrow e + \bar{\nu}_e + p \qquad (3.6.6)$$

   are computed in *pyBBN*. This involves solving the system of equations for the evolution of temperature, scale factor and distribution functions of decoupled species like neutrinos, HNLs and relevant decays products (see Sec. 3.2.2).

2. The cosmological quantities together with the aforementioned rates are tabulated and passed to an external code, the modified KAWANO code (also known as NUC123 [83]), that takes care of the nuclear part of the simulation and outputs the light element abundances. This is done to avoid reimplementing the complicated system of nuclear physics with experimental fits to the cross-sections. KAWANO is a relatively outdated code that does not implement a number of corrections introduced in PArthENoPE [87] or PRIMAT [79], but these corrections are small in SBBN and KAWANO is the easiest to modify for custom cosmological evolution and neutrino spectral distortions.

Time evolution. The system of equations step-by-step for values of scale factor evenly spaced in logarithmic scale (as opposed to linearly-spaced values). This is done to naturally increase the timestep at late times when interaction rates are low. The numerical schemes used for time-evolution equations are the *Adams-Bashforth explicit linear multistep method* for temperature evolution equation and *Adams-Moulton implicit linear multistep method* for Boltzmann equations. We decided to use these particular schemes (as opposed to the Runge-Kutta family) since they allow us to reuse the previously computed values (we use up to 5 previous steps) and proved to be stable enough for our purposes.

The implicit methods are known to help mitigate the problem of *stiffness* in Boltzmann equations when it is possible to apply them. However, the temperature evolution equation in our system can be solved only explicitly.

Boltzmann integrals for distribution functions are in general of high dimension. Typically this would suggest the usage of Monte Carlo methods. But in the particular case of the Fermi-like theory of HNLs, it is possible to reduce the number of integrations to 1 or 2 (this is described in detail in App. 3.C). The accuracy of this computation is critical to the evolution of the neutrinos which constitute a large fraction of the total energy density as well as directly influence the neutron-to-proton ratio. Hence we elected to use the adaptive integration algorithm QAG from GNU Scientific Library [109]. Integration is done using the Gaussian quadrature and the error is estimated by dividing the integration region into subintervals and repeating the procedure until the required global tolerance is met. Note that the integrand of the problem is piecewise smooth while the Gaussian quadrature is designed for the polynomials. Because of this the Gaussian quadrature might require arbitrarily large number of points to converge. However, we find that the integration using

the Gaussian quadrature of the order 60 provides the same accuracy as the adaptive method for four-particle interactions. For three-particle interactions we employ adaptive integration with Gauss-Kronrod quadrature of the order 31.

Unstable HNL decay products. Sterile neutrinos of high masses decay into short-lived mesons – e.g. π_0 lifetime is $\sim 10^{-17}$ sec and $\pi_\pm$ lifetime is $\sim 10^{-8}$ sec. These timescales are orders of magnitude smaller than any practical computational timestep for a simulation spanning 10^4 sec.

To the first approximation, short-lived particles can be immediately destroyed without populating their distributions. However, some of them are charged and thermalize with the plasma and transfer some energy.

For illustration, we will consider the case of a muon. HNLs with masses $M_N < 105$ MeV will decay into stable particles. HNLs with higher masses will have decay products that are unstable. Some of these unstable decay products will interact with the plasma before they decay. The analysis here will be done for muons but can be applied to all other particles as well.

There are three important events to consider:

1. $\mu^\pm$ is created from HNL decay
   The distribution function of these muons is a non-thermal distribution of f_{noneq}.

2. $\mu^\pm$ thermalizes
   The muon-photon scattering rate is higher than the muon decay rate: $\Gamma_{\gamma\mu} \sim \frac{\alpha^2}{m_\mu E_\gamma} T_\gamma^3 \sim 10^{-9} \, \text{MeV}$ vs. $\Gamma_{\mu,\text{decay}} \sim 10^{-16} \, \text{MeV}$. This means that the muons will release their energy into the plasma and equilibrate before they decay. This process increases the temperature of the plasma and makes the muons non-relativistic. After thermalization, the muons will share the same temperature as the plasma and will have a thermal distribution $f_{\text{thermal}} = e^{-\frac{m_\mu - \mu}{T}} e^{-\frac{p^2}{2m_\mu T}}$, where μ is determined by the condition that the number density before and after thermalization must be equal. The collision term corresponding to this process is then estimated as

$$I_{\text{thermalization}} \approx \frac{f_{\text{thermal}} - f_{\text{noneq}}}{\Delta t} \, ,$$

   with Δt the timestep of the simulation. The same procedure is followed for charged pions and charged kaons.

3. $\mu^\pm$ decays
   The main decay channel of muons is $\mu^- \to e^- + \overline{\nu_e} + \nu_\mu$. The muon has a lifetime, $\tau_\mu \sim 10^{-6}$ sec., that is much smaller than the timestep of the simulation. This poses a problem right away: when the evolution of the distribution function for the muon and active neutrinos is computed as $\Delta f = I_{\text{coll}} \Delta t$, the behavior of I_{coll} is not resolved. It is assumed to be constant during the whole timestep Δt, which is not true; the created

53

muons have already decayed well within this timestep. What therefore happens is that the number of muons that have decayed and the number of neutrinos that are created are overestimated. This issue can be solved by using a dynamical equilibrium. Consider the chain The timestep Δt is much smaller than the lifetime of the HNL,
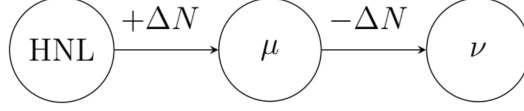
which means that there is approximately a constant inflow of muons *during* each timestep. Since the number of muons created ΔN decays almost instantaneously, the same number of active neutrinos is created: for each muon that decays, one electron neutrino and one muon neutrino is created. Now a scaling α can be introduced in $\Delta f = I_{\text{coll}} \Delta t \alpha$ such that $\int \mathrm{d}^3 p \Delta f / (2\pi)^3 = \Delta N$.

3.6.2 Code testing

pyBBN had been tested in multiple situations modeling both Standard BBN theory as well as in the presence of HNLs. Below we will list only the most representative selection of them.

The main areas of interest are thermodynamical properties (i.e., evolution of temperature) and computation of Boltzmann equations (collision integrals and integration of the equations themselves).

According to scope, tests can be divided into *low-level* (i.e., checking that formulas or numerical schemes are correct) and *integration tests* (i.e., those defining a toy model for which we can produce analytical results or can make a comparison with other codes).

3.6.2.1 Temperature of the Cosmic Neutrino Background

Decoupling neutrinos to the first approximation preserve their thermal distribution function with the same temperature as photons.

However, at the temperatures about $T \sim m_e = 0.511 keV$, electrons and positrons annihilate into additional photons. Since neutrinos decouple earlier, do not get heated up as well. This results in a difference in the temperature of Cosmic Microwave and Neutrino Backgrounds.

This effect is easily quantified through the comoving entropy conservation law ($sa^3 = const$):

$$\frac{T_\gamma}{T_\nu} = \frac{aT_\gamma}{aT_\nu} = \left(\frac{g_*(T_{\text{before}})}{g_*(T_{\text{after}})} \right)^{\frac{1}{3}} = \left(\frac{\frac{7}{8} \cdot 4 + 2}{2} \right)^{\frac{1}{3}} \approx 1.401 \qquad (3.6.7)$$

This law is not explicitly enforced in the code, but comoving entropy is an integral of motion. Check of the Cosmic Neutrino Background temperature allows to test the thermodynamics in the code as a whole: Fig.3.7. The neutrinos are assumed to decouple instantly with any non-equilibrium dynamics beside free propagation.
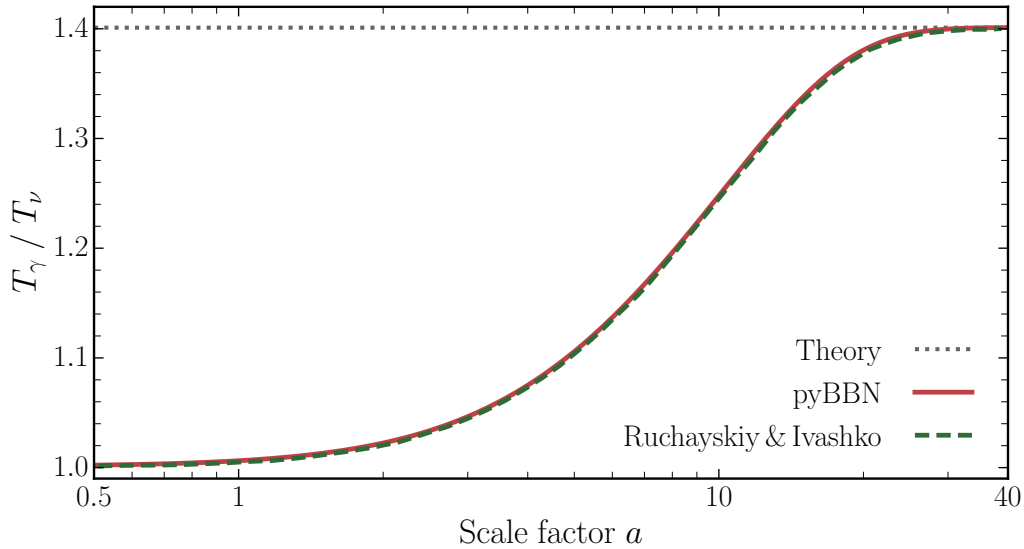
Figure 3.7: The photon temperature divided by active neutrino temperature. The increase here is due to electron-positron annihilation into photons. Dashed curve is from [106].

> In the Universe filled with electrons, positron, photons and decoupled neutrino – photons and neutrinos have different temperatures after electrons-positron annihilation. The ratio of their temperatures can be predicted from the conservation of comoving entropy. Fig. 3.7 shows that the thermodynamic evolution in the code predicts this ratio with accuracy $\lesssim 0.1\%$.

3.6.2.2 Increased Hubble rate

pyBBN predicts Helium-4 abundance for the SM BBN (SBBN) to be equal to 24.8%. This value agrees with a simple analytic estimate (3.2.1) within 10%. In this test we will check that increase of a Hubble rate during BBN result in the correct qualitative behavior of $Y_{4\text{He}}$.

An increase in the Hubble parameter leads to an earlier decoupling of neutron-proton weak reactions and therefore a higher neutron-to-proton ratio. This is shown in the scenario where the Hubble rate in SBBN is increased artificially by a factor of 2, see Fig. 3.8. We check the prediction of the code using analytic estimate (3.2.1), (3.2.14) of the ^{4}He abundance. The neutron decoupling temperature obtained using $\Gamma_{n\to p}(T) = H(T)$ gives $T_n \simeq 0.92 \, \text{MeV}$, which translates into $n_n/n_p(T_{\text{BBN}}) \approx 0.195$ and $Y_{4\text{He}} \approx 0.37$, in good agreement with the result shown in Fig. 3.8.
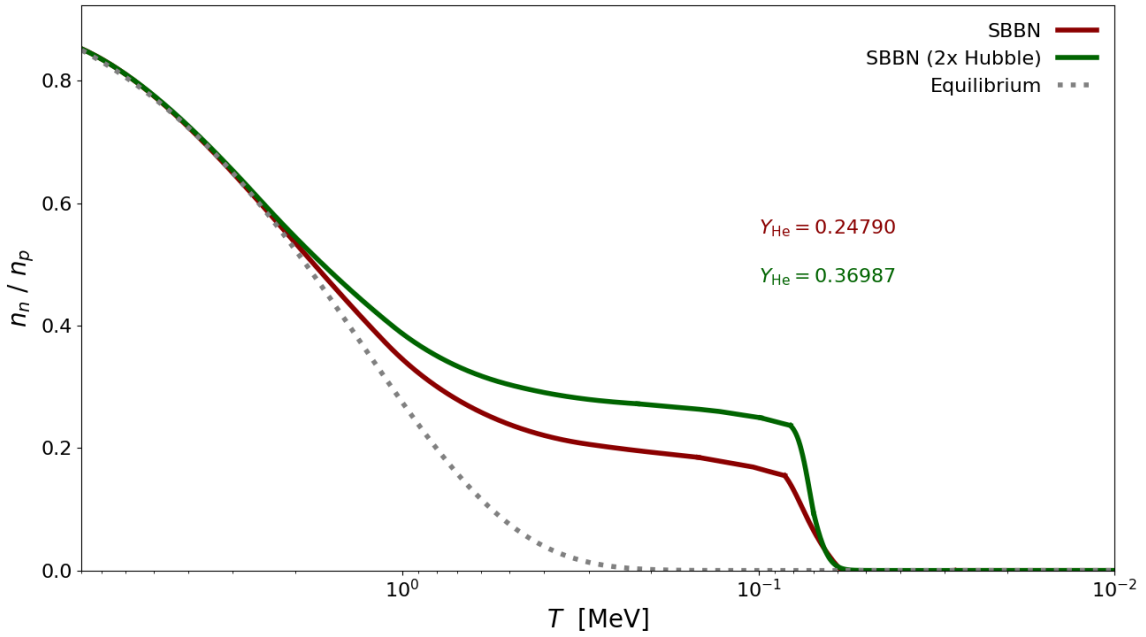
Figure 3.8: Evolution of the neutron-to-proton ratio in SBBN (red line) and SBBN + twice as large Hubble parameter (green line). The dotted line indicates the equilibrium evolution.

> The prediction of the code for the value of $Y_{^4\text{He}}$ in SM BBN agrees with analytic estimate. The result of the simulation with increase Hubble rate also agrees with our expectations.

3.6.2.3 Neutrino spectral distortions

In Fermi theory the cross section increases with momentum as $\sigma \propto G_F^2 p^2$, which means that neutrinos with higher momenta stay longer in equilibrium. Since these neutrinos decouple later, they will in fact briefly experience the heat-up of the plasma due to electron-positron annihilation, shown in Figure 3.7, and the corresponding increase in aT. On the other hand, oscillations of neutrinos are mixing the distributions of different flavours, bringing the spectra closer to each other.

At temperatures of $\mathcal{O}(1)$ MeV electron neutrinos interact through both charged and neutral currents, while muon and tau neutrinos only interact through the neutral current. The temperature is too low for muons and tau leptons to be present in the plasma or to be created from muon and tau neutrinos through the charged current channel. The cross section of electron neutrinos is therefore larger and they get larger spectral corrections. The results are shown in Fig. 3.9.

56
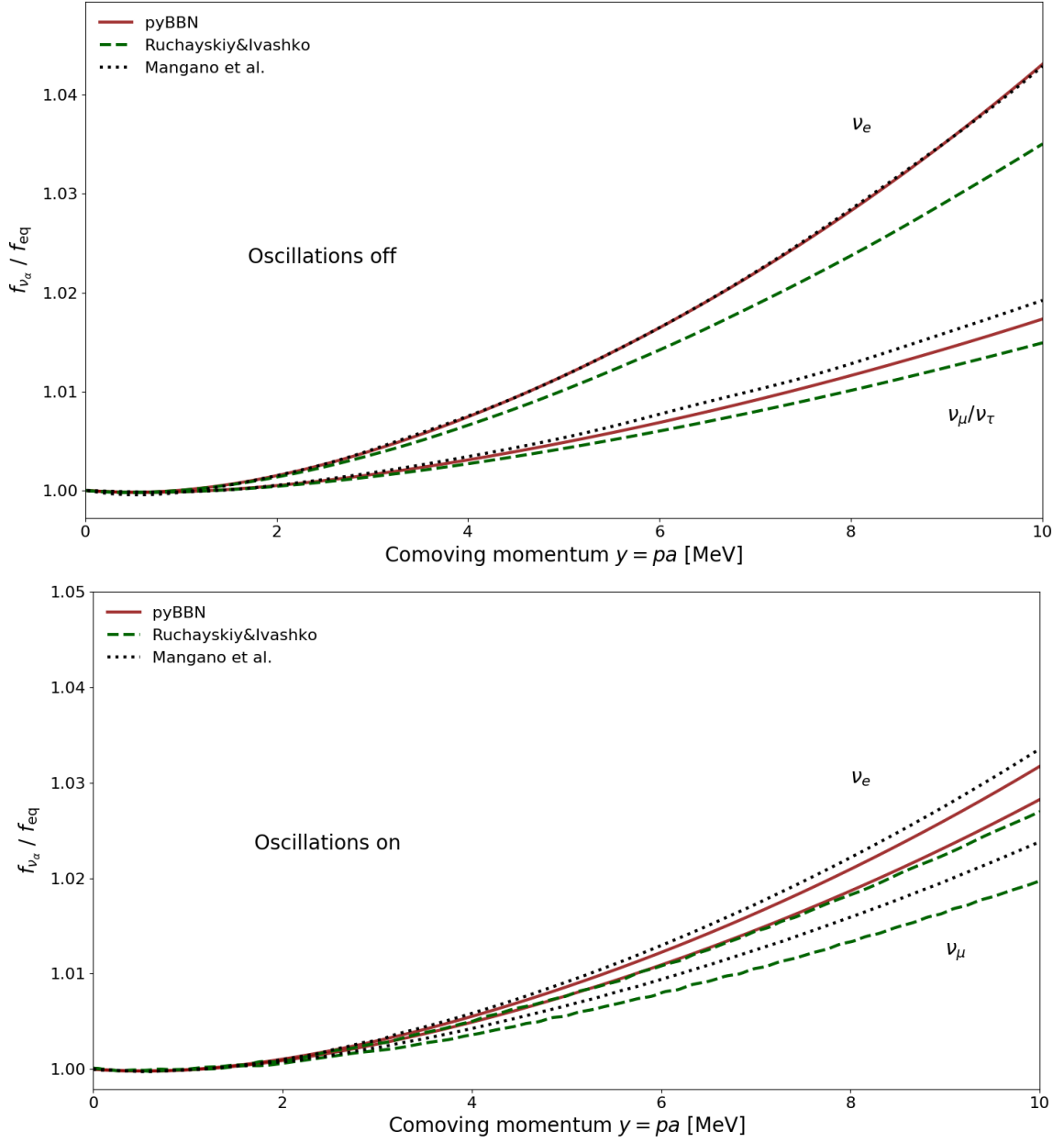
Figure 3.9: Ratio of active neutrino decoupled spectra to their equilibrium distribution before the onset of BBN. The upper curves show the distortion of the electron neutrino spectrum and the lower of muon and tau neutrinos. Dashed curves are from [106], dotted from [103].

Upper panel: no neutrino flavour oscillations. *Lower panel:* neutrino flavour oscillations with $\sin^2 \theta_{12} = 0.3$, $\sin^2 \theta_{13} = 0.5$ and $\theta_{23} = 0$ (rotated $\nu_\mu - \nu_\tau$ basis).

In this test we considered the spectral distortions of neutrinos during decoupling of

weak reactions. We reproduce the results of codes by [106] and [103] up to $\lesssim 1\%$ accuracy even for very sparsely populated momentum values of neutrinos ($f(p, T) \sim \exp(-p/T) \sim 10^{-3}$).

3.6.2.4 Heavy Neutral Leptons decay width

To check the computation of collision integrals and integration of Boltzmann equations, we restore the decay width of HNLs for the computed evolution of their number density using that

$$\frac{n_i - n_{i-1}}{\Delta t} + 3Hn_i \approx \Gamma_N n_i \tag{3.6.8}$$

We consider an HNL of mass $M_N < M_\mu$ that mixes only with electron neutrino through mixing angle $|\theta_e|$. There are four decay channels:

$$N \to \nu_e + \nu_e + \overline{\nu_e} \qquad\qquad N \to \nu_e + \nu_\mu + \overline{\nu_\mu}$$
$$N \to \nu_e + \nu_\tau + \overline{\nu_\tau} \qquad\qquad N \to \nu_e + e^+ + e^- \ , \tag{3.6.9}$$

The theoretical vacuum decay width is

$$\Gamma_N = \frac{G_F^2 \, |\theta_e|^2 \, M_N^5 \left(\frac{1}{4}\left(1 + 4\sin^2\theta_W + 8\sin^4\theta_W\right) + 1\right)}{192\pi^3}. \tag{3.6.10}$$

We check that our code correctly reproduce this value. The result is shown in Fig. 3.10.

We have confirmed that decay width of HNL computed using collision integrals in our code recovers the theoretical decay width with accuracy $\sim 0.2\%$

3.6.2.5 Reheating due to neutral pion decay

As a check of energy conservation in 3-body interactions, we check the energy injected by decays of $M_N = 135$ MeV sterile neutrino.

$$N \to \pi^0 + \nu_e \tag{3.6.11}$$
$$\pi^0 \to \gamma + \gamma \tag{3.6.12}$$

For each HNL that decays, one neutral pion is created. If the HNL has a mass very close to that of the neutral pion, then the neutral pion created will be very non-relativistic. The energy injected in the plasma due to neutral pion decay during each step is therefore
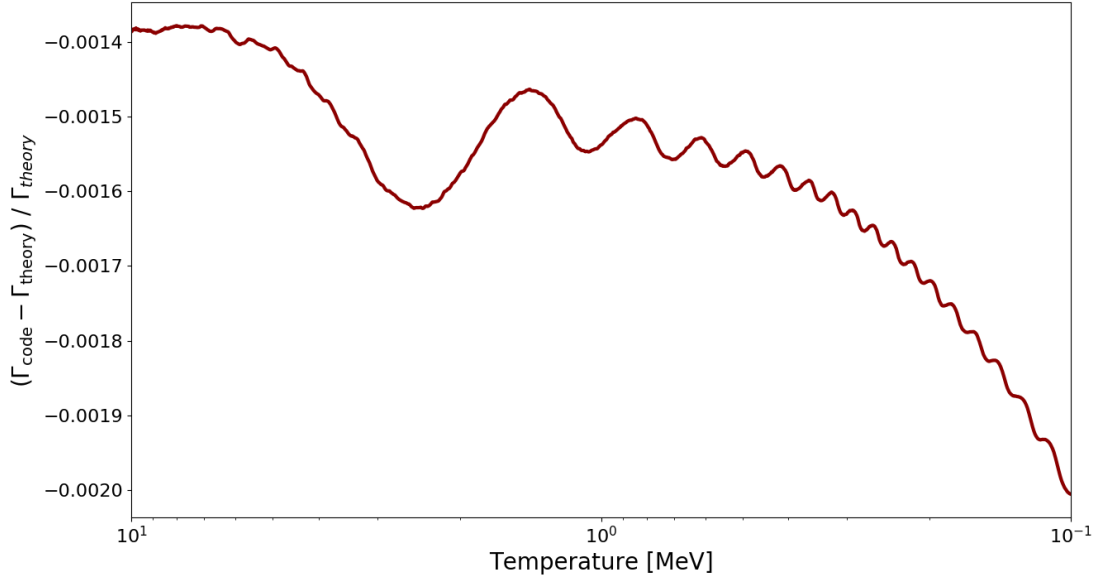
Figure 3.10: Relative difference between averaged HNL decay width calculated in pyBBN and its theoretical value (Eq. 3.6.10) of reactions Eq. 3.6.9. Here an HNL of mass $M_N = 30$ MeV and mixing angle $|\theta_e|^2 = 10^{-4}$ is chosen. The initial temperature of the simulation is $T_{\mathrm{ini}} = 10$ MeV.

approximately

$$\Delta E \approx 2m_{\pi^0}\Delta N \ , \tag{3.6.13}$$

with ΔN the number of HNLs that have decayed during the step. The factor of 2 comes from the fact that the charge conjugated channel also creates a neutral pion. Comoving photon energy density becomes

$$g_\gamma \frac{\pi^2}{30}(aT)^4 = \rho_{\gamma,\mathrm{c}} = \rho_{\gamma,\mathrm{c,old}} + a^4\Delta E = \rho_{\gamma,\mathrm{old}} + 2m_{\pi^0}a^4\Delta N \tag{3.6.14}$$

and

$$(aT)_{\mathrm{new}} = \left(\frac{30}{g_\gamma\pi^2}\rho_{\gamma,\mathrm{c}}\right)^{1/4} . \tag{3.6.15}$$

The result of this test is shown in Fig. 3.11.

Sterile neutrinos with masses $m_N > m_\pi$ have qualitatively new decay channels into 2 particles. This involves a different expression of the collision integral that haven't been tested before. We have confirmed that 3-body interactions conserve particle number, energy and comoving entropy up to $\lesssim 0.1\%$.

59
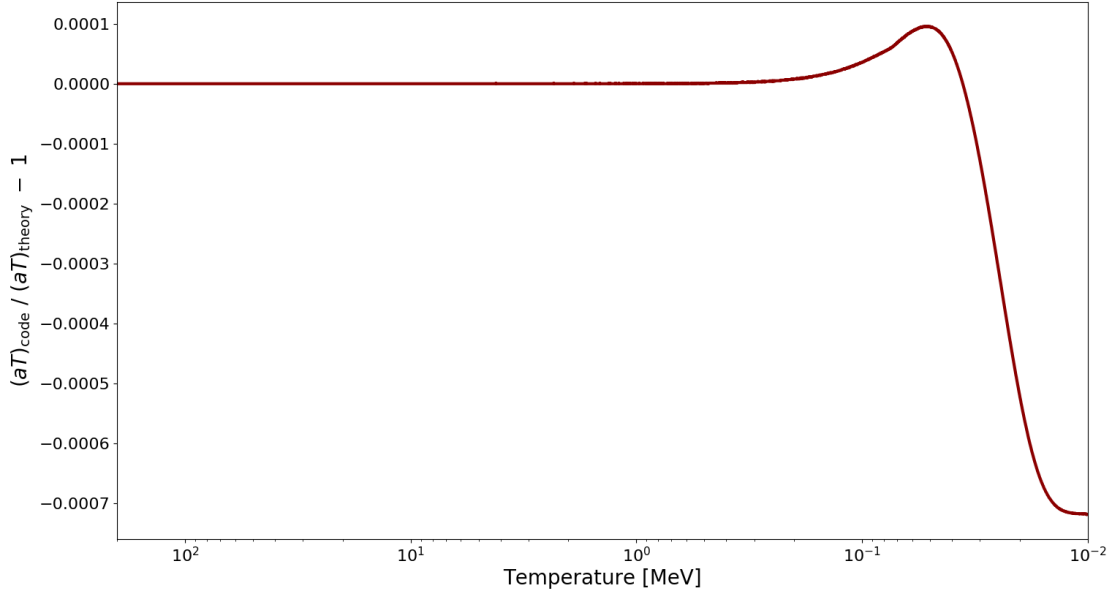
Figure 3.11: Reheating of plasma due to decays 3.6.11 and 3.6.12. Theoretical value aT_{theory} is given by Eq. 3.6.15. Here an HNL of mass $M_N = 135$ MeV with $|\theta_e|^2 = 10^{-4}$ is taken.

3.6.2.6 Chain of decays with short-lived particles

Sterile neutrinos of high masses can decay into heavy mesons that themselves trigger the decay chains. Since mesons have very short lifetimes compared to the timescales of the simulation, their decays are treated by a different numerical scheme than neutrinos or nucleons. Eventually, all mesons decay into photons, leptons and neutrinos. For a given mass of HNL we can compute the total number of decay products and compare it with the code.

Consider the chain of decays

$$N \to \nu_e + \phi \tag{3.6.16}$$

$$\phi \to \pi^0 + \rho^0 \tag{3.6.17}$$

$$\pi^0 \to \gamma + \gamma \; ; \; \rho^0 \to \pi^+ + \pi^- \tag{3.6.18}$$

$$\pi^- \to \mu^- + \overline{\nu_\mu} \tag{3.6.19}$$

$$\mu^- \to e^- + \overline{\nu_e} + \nu_\mu \tag{3.6.20}$$

The total comoving number density of electron and muon neutrinos created is

$$n_{\nu_e,\text{c}} = (1 + 2\text{BR}_{\phi \to \pi^0 \rho^0})n_{\text{ini}} \tag{3.6.21}$$

$$n_{\nu_\mu,\text{c}} = 4\text{BR}_{\phi \to \pi^0 \rho^0} n_{\text{ini}} \tag{3.6.22}$$

with $\mathrm{BR}_{\phi \to \pi^0 \rho^0}$ the branching ratio of the corresponding decay and n_{ini} the initial comoving number density of HNLs. The factor of 2 comes from the fact that the charge conjugated channel $\overline{N} \to \overline{\nu_e} + \phi$ also creates a phi meson. Results are shown in Fig. 3.12.
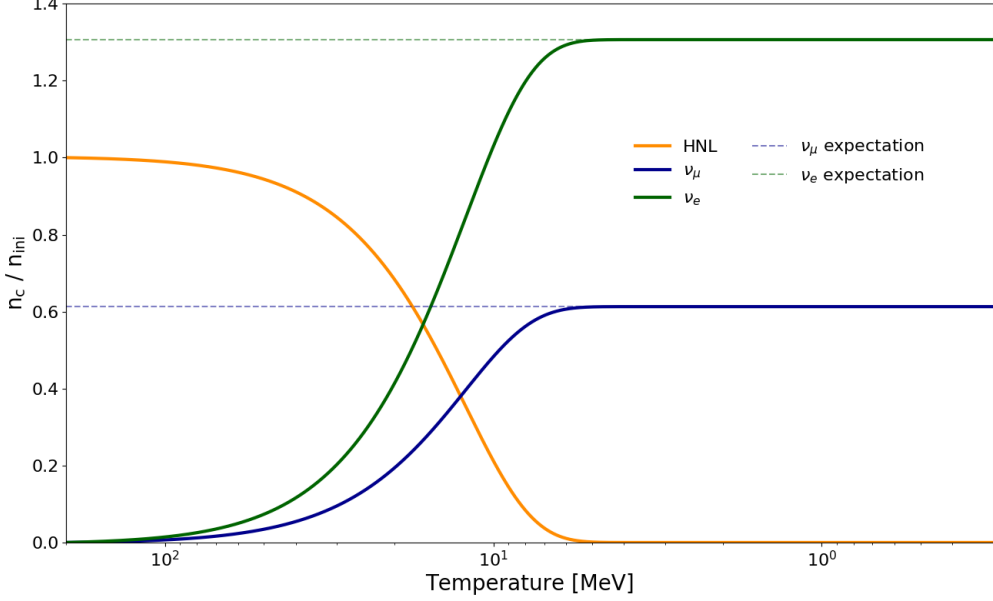
Figure 3.12: Comoving number density n_{c} of HNL, electron neutrino and muon neutrino normalized by the initial HNL comoving number density n_{ini} when the decays 3.6.16 - 3.6.20 are considered. The mass of the HNL here is $M_N = 1200$ MeV and $|\theta_e|^2 = 10^{-8}$. The initial densities of all particles except for HNLs, electrons and photons are taken 0 for convenience. The dynamical equilibrium mechanism discussed in Subsection 3.6.1 is used here.

Even for particles with lifetimes much smaller than the timestep of the simulation $\tau \ll \Delta t$, we correctly reproduce the expected concentrations of the decay products.

3.6.2.7 Helium-4 and Deuterium abundance in Standard Model Nucleosynthesis

We present the predictions for Helium and Deuterium in Standard BBN as computed by *pyBBN* in Fig. 3.13. Our result is consistent with state-of-the art precision BBN code PArthENoPE [87]. In both cases we used the neutron lifetime $\tau_n = 880.2$ sec from [1].

3.6.2.8 Non-equilibrium neutrinos vs thermal neutrinos

The current generation of high-precision BBN codes (e.g. PArthENoPE [87] and PRIMAT [79]) include many corrections not included in KAWANO. However, they

---

Code	$Y_{^4\text{He}}$	D/H
PArthENoPE2.0 [87]	0.24676	$0.26800 \cdot 10^{-4}$
pyBBN	0.24678	$0.26776 \cdot 10^{-4}$
Relative difference	-0.008%	0.09%

Figure 3.13: Comparison of *pyBBN* with PArthENoPE

were designed for the SBBN and treat neutrinos as thermal-like particles, ignoring possible spectral distortions. We show that in presence of HNLs the spectral distortions can have the leading effect on the neutron-to-proton ratio.

In modern BBN codes (e.g. PArthENoPE [87] and PRIMAT [79]), neutrinos are assumed to have a thermal-like Fermi-Dirac distribution at all times. The neutrino temperature T_ν is determined such that

$$g_\nu \int \mathrm{d}^3 p_\nu \frac{E_\nu}{e^{\frac{E_\nu}{T_\nu}} + 1} = \rho_\nu \tag{3.6.23}$$

with ρ_ν the energy density.

The energy density is computed by the evolution equation

$$\frac{\partial \rho_\nu}{\partial t} + 4H\rho_\nu = C \tag{3.6.24}$$

with C the source function (integrated collision integral). In this way they account for the gravitational effect of incomplete neutrino decoupling.

Spectral distortions are ignored. In order to track spectral distortions, it is necessary to use the full machinery of the Boltzmann equation. On the other hand, the authors of [110] have compared their results with the authors of [111] (where spectral distortions are taken into account) and found a correction of only order 10^{-3}.

Since we expect production of neutrinos after decoupling, we expect a larger influence on the neutron-to-proton rates. Therefore, we have performed paired simulations: one using the non-equilibrium neutrino distribution in the calculation of the neutron-to-proton rates, while the other uses a thermal distribution with temperature that gives the same energy density as the non-equilibrium distribution. Note: this is only done at the level of the computation of the neutron-proton weak rates, in order to understand whether spectral distortions can be neglected. The evolution of the neutron-to-proton ratio is shown in Figure 3.14 and the final $Y_{^4\text{He}}$ abundances are displayed in the table below.

	Non-equilibrium	Thermal	Rel. diff.
SBBN	0.24790	0.24743	0.19%
HNL ($M_N = 30\,\text{MeV}, \tau = 0.075\,\text{s}$)	0.24824	0.24714	0.44%
HNL ($M_N = 105\,\text{MeV}, \tau = 0.11\,\text{s}$)	0.29266	0.24405	16.6%
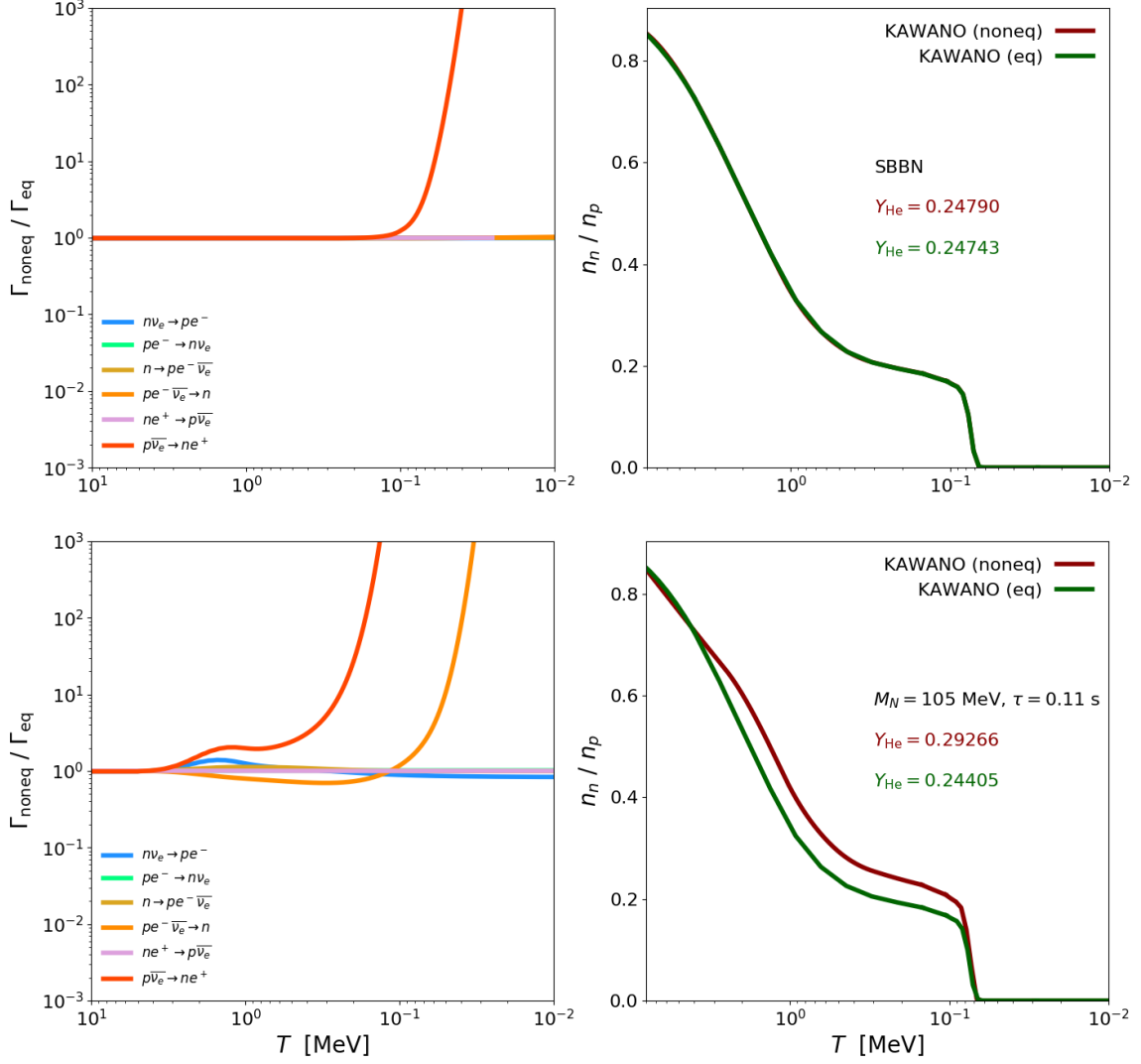
Figure 3.14: Comparison of nucleon reaction rates and neutron-to-proton ratio in SBBN (upper row) and in a model including HNL of mass $105\,\text{MeV}$ and lifetime $0.11\,\text{s}$ (lower row). *"noneq"* quantities are computed using the electron neutrino distribution function as obtained from the Boltzmann equation , while *"eq"* quantities are given under assumption of a thermal-like electron neutrino distribution (with the same energy density).

Fig. 3.14 indicates that spectral distortions can be the dominant effect that change the neutron-to-proton ratio in presence of sterile neutrinos.

3.6.2.9 Summary on tests

We have performed a large number tests, verifying the correctness of the equations, numerical schemes and conservation of basic integrals of motion. We payed special attention to the most computationally intensive and difficult part of the code – the evaluation of Boltzmann integrals. We checked

- integration of the expansion laws

- behaviour of the thermodynamical quantities under expansion

- energy transfer between non-equilibrium particles and plasma

- non-equilibrium corrections to particle species

We have implemented most of the tests previously done in literature. However, since physics of sterile neutrinos with masses $m_N > m_\pi$ have not been implemented elsewhere, we have created a number of sanity checks which we believe to give us reasonable control.

3.7 Results and conclusions

We have extended the constraints on sterile neutrinos from pion mass up to phi-meson mass. To achieve this, we have developed a numerical code *pyBBN* that we use to predict abundances of light chemical elements.

We implemented hadronic decay modes of HNLs, subsequent hadronic decays and their influence on the BBN. We validated the code comparing it with other codes and analytic estimates, and making consistency checks.

The comparison between bounds on the HNL parameter space provided by our code and codes from [105, 106] for HNLs in the mass range $m_N < m_\mu \approx 105$ MeV is shown in Fig. 3.15. Our results are in good agreement with previous studies in the mass range $m_N < 105$ MeV considered there. In the Figures 3.16 and 3.17 we show the predictions of our code for HNLs with masses up to $m_N \simeq 1$ GeV, comparing it with the simple constraint $\tau_N \gtrsim 0.1$ s from [49].

So far we have not implemented the effect of the interaction of massive decay products of HNLs – mesons and muons – with nucleons. Even if the lifetimes of these particles are much shorter than characteristic times of BBN, their production rate can be large enough to give a significant effect on n-p ratio [81]. To take into account this effect, we need to add in
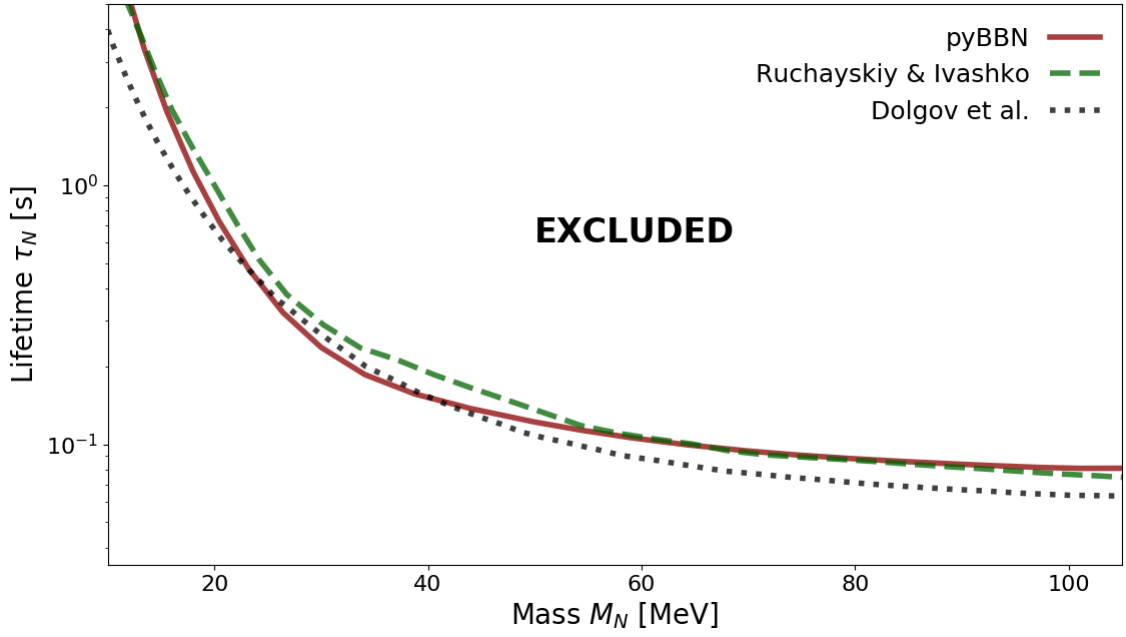
Figure 3.15: Constraints from BBN on the lifetime of HNLs up to the muon mass $M \approx 105$ MeV. Mixing with electron neutrino only is assumed here. Dashed curve is from [106], dotted from [105].

Boltzmann equations for nucleon number densities additional terms describing not decay of mesons, but their scattering on nucleons. This effect increases the neutron-to-proton ratio and hence the Helium abundance. Therefore, our constraints obtained without this effect are conservative. We will investigate this effect in the future.
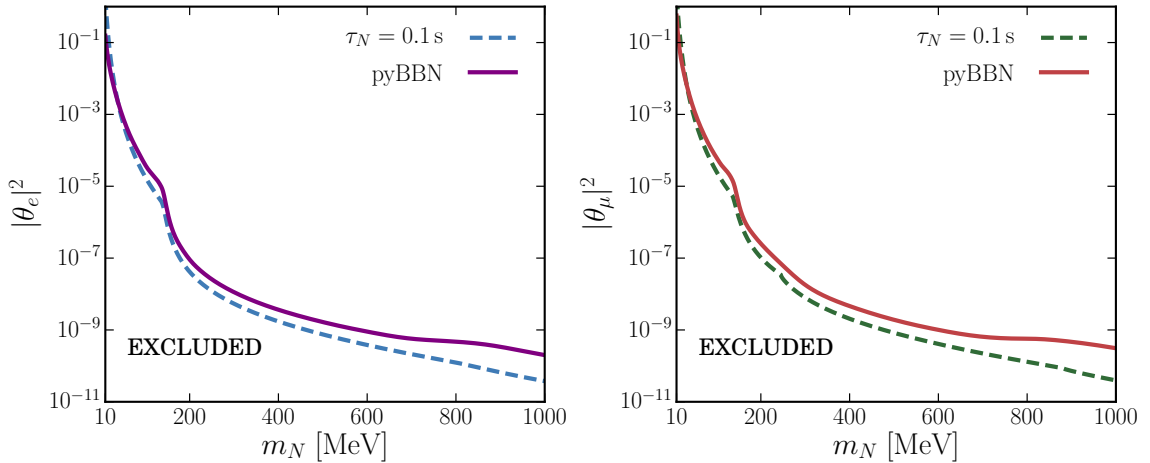
Figure 3.16: Bounds on HNL mixing angles. *Left:* mixing only with electron neutrinos. *Right:* mixing only with muon neutrinos.
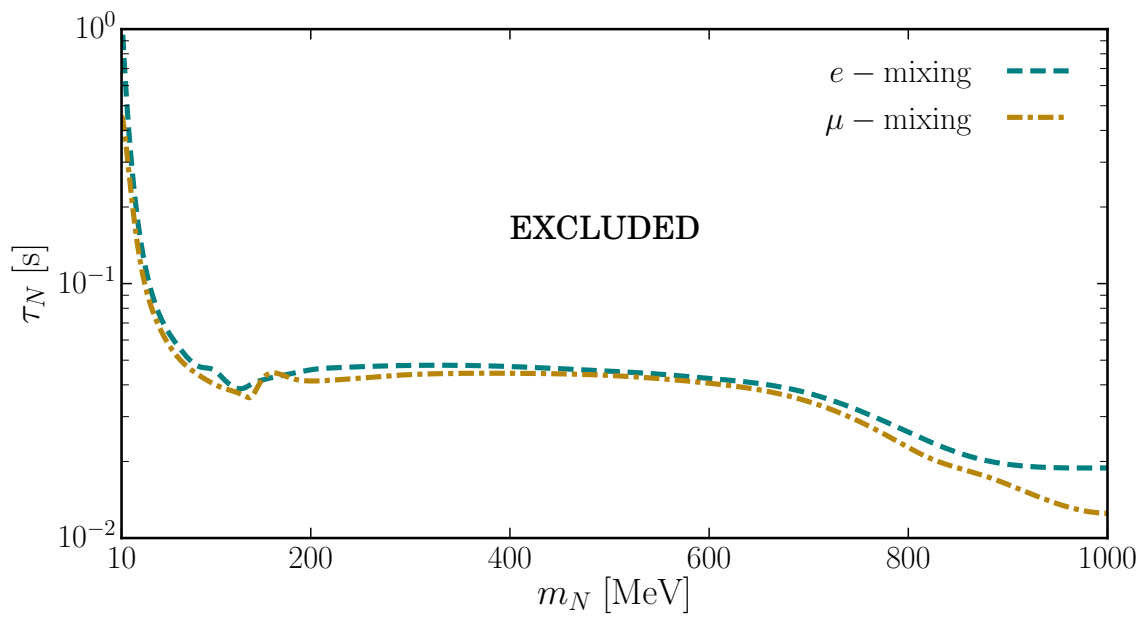
Figure 3.17: Constraints from BBN on the lifetime of HNLs up to the phi-meson mass $M \approx 1$ GeV (for electron and muon mixing).

Appendix

3.A   Temperature evolution

Consider a plasma consisting of four particle species representing the contents of the Universe. For example, the four species here are photons, electrons, active neutrinos and HNLs. The addition of other species (e.g. muons) will then follow a similar procedure. In what follows: $\widetilde{T} = aT$, $\widetilde{E} = aE$. The total energy density and total pressure are given by

$$\rho_{\text{tot}} = \rho_\gamma + \rho_e + \rho_\nu + \rho_N \qquad\qquad P_{\text{tot}} = P_\gamma + P_e + P_\nu + P_N$$

$$\rho_\gamma = g_\gamma \frac{\pi^2}{30} \frac{1}{a^4} \widetilde{T}^4 \qquad\qquad P_\gamma = \tfrac{1}{3}\rho_\gamma$$

$$\rho_e = \frac{g_e}{2\pi^2} \frac{1}{a^4} \int \mathrm{d}y\, y^2 \frac{\widetilde{E}_e}{e^{\frac{1}{\widetilde{T}}\widetilde{E}_e}+1} \qquad\qquad P_e = \frac{g_e}{6\pi^2} \frac{1}{a^4} \int \mathrm{d}y\, \frac{y^4}{\widetilde{E}_e} \frac{1}{e^{\frac{1}{\widetilde{T}}\widetilde{E}_e}+1}$$

$$\rho_\nu = \frac{g_\nu}{2\pi^2} \frac{1}{a^4} \int \mathrm{d}y\, y^3 f_\nu \qquad\qquad P_\nu = \tfrac{1}{3}\rho_\nu$$

$$\rho_N = \frac{g_N}{2\pi^2} \frac{1}{a^4} \int \mathrm{d}y\, y^2 \sqrt{y^2 + a^2 m_N^2}\, f_N \quad P_N = \frac{g_N}{6\pi^2} \frac{1}{a^4} \int \mathrm{d}y\, \frac{y^4}{\widetilde{E}_N} f_N$$

The energy conservation law is

$$\frac{\mathrm{d}\rho_{\text{tot}}}{\mathrm{d}\ln a} \frac{\mathrm{d}\ln a}{\mathrm{d}t} + 3H \left(\rho_{\text{tot}} + P_{\text{tot}}\right) = 0 \implies \frac{\mathrm{d}\rho_{\text{tot}}}{\mathrm{d}\ln a} + 3 \left(\rho_{\text{tot}} + P_{\text{tot}}\right) = 0$$

Taking derivatives:

$$\frac{\mathrm{d}\rho_\gamma}{\mathrm{d}\ln a} = g_\gamma \frac{\pi^2}{30}\left(-4\frac{1}{a^4}\widetilde{T}^4 + 4\frac{1}{a^4}\widetilde{T}^3 \frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a}\right) = -4\rho_\gamma + 4\frac{\rho_\gamma}{\widetilde{T}}\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a}$$

$$\frac{\mathrm{d}\rho_e}{\mathrm{d}\ln a} = -4\rho_e + \frac{g_e}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^2 \left[\frac{\left(e^{\frac{1}{\widetilde{T}}\widetilde{E_e}} + 1\right)\frac{a^2 m_e^2}{\widetilde{E_e}} - \widetilde{E_e}e^{\frac{1}{\widetilde{T}}\widetilde{E_e}}\left(\frac{a^2 m_e^2}{\widetilde{T}\widetilde{E_e}} - \frac{\widetilde{E_e}}{\widetilde{T}^2}\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a}\right)}{\left(e^{\frac{1}{\widetilde{T}}\widetilde{E_e}} + 1\right)^2}\right]$$

$$= -4\rho_e + \frac{g_e}{2\pi}\frac{1}{a^4}\int \mathrm{d}y y^2 \left[\frac{a^2 m_e^2}{\widetilde{E_e}}\frac{1}{e^{\frac{1}{\widetilde{T}}\widetilde{E_e}} + 1} - \left(\frac{a^2 m_e^2}{\widetilde{T}} - \frac{\left(\widetilde{E_e}\right)^2}{\widetilde{T}^2}\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a}\right)\frac{e^{\frac{1}{\widetilde{T}}\widetilde{E_e}}}{\left(e^{\frac{1}{\widetilde{T}}\widetilde{E_e}} + 1\right)^2}\right]$$

$$\frac{\mathrm{d}\rho_\nu}{\mathrm{d}\ln a} = -4\rho_\nu + \frac{g_\nu}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^3 \frac{\mathrm{d}f_\nu}{\mathrm{d}\ln a} = -4\rho_\nu + \frac{g_\nu}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^3 \frac{1}{H}I_\nu$$

$$\frac{\mathrm{d}\rho_N}{\mathrm{d}\ln a} = -4\rho_N + \frac{g_N}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^2 \left[\frac{a^2 m_N^2}{\widetilde{E_N}} + \widetilde{E_N}\frac{\mathrm{d}f_N}{\mathrm{d}\ln a}\right]$$

$$= -4\rho_N + \frac{g_N}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^2 \left[\frac{a^2 m_N^2}{\widetilde{E_N}} + \widetilde{E_N}\frac{1}{H}I_N\right]$$

The equation is also valid for individual species:

$$\frac{\mathrm{d}\rho_\gamma}{\mathrm{d}\ln a} + 3\left(\rho_\gamma + P_\gamma\right) = 4\frac{\rho_\gamma}{\widetilde{T}}\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a}$$

$$\frac{\mathrm{d}\rho_e}{\mathrm{d}\ln a} + 3\left(\rho_e + P_e\right) = \frac{1}{a^4}\left[-\frac{a^2 m_e^2}{\widetilde{T}}R_1 + \frac{1}{\widetilde{T}^2}\left\{R_2 + a^2 m_e^2 R_1\right\}\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a}\right]$$

$$\frac{\mathrm{d}\rho_\nu}{\mathrm{d}\ln a} + 3\left(\rho_\nu + P_\nu\right) = \frac{g_\nu}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^3 \frac{1}{H}I_\nu$$

$$\frac{\mathrm{d}\rho_N}{\mathrm{d}\ln a} + 3\left(\rho_N + P_N\right) = \frac{g_N}{2\pi^2}\frac{1}{a^4}\int \mathrm{d}y y^2 \widetilde{E_N}\frac{1}{H}I_N \ ,$$

with

$$R_1 = \frac{g_e}{2\pi^2}\int \mathrm{d}y y^2 \frac{e^{\frac{\widetilde{E_e}}{\widetilde{T}}}}{\left(e^{\frac{\widetilde{E_e}}{\widetilde{T}}} + 1\right)^2}$$

$$R_2 = \frac{g_e}{2\pi^2}\int \mathrm{d}y y^4 \frac{e^{\frac{\widetilde{E_e}}{\widetilde{T}}}}{\left(e^{\frac{\widetilde{E_e}}{\widetilde{T}}} + 1\right)^2}$$

Adding all these terms together and solving for the temperature derivative gives:

$$\frac{\mathrm{d}\widetilde{T}}{\mathrm{d}\ln a} = \frac{\frac{a^2 m_e^2}{\widetilde{T}}R_1 - \frac{g_\nu}{2\pi^2}\int \mathrm{d}y y^3 \frac{1}{H}I_\nu - \frac{g_N}{2\pi^2}\int \mathrm{d}y y^2 \widetilde{E_N}\frac{1}{H}I_N}{\frac{2\pi^2 g_\gamma}{15}\widetilde{T}^3 + \frac{1}{\widetilde{T}^2}R_2 + \frac{a^2 m_e^2}{\widetilde{T}^2}R_1} \tag{3.A.1}$$

3.B   Relevant matrix elements

The matrix elements listed here are not averaged over any helicities. Subsection 3.B.1 contains the reactions involving SM particles only, Subsection 3.B.2 the reactions involving HNLs above QCD-scale and Subsection 3.B.3 the reactions involving HNLs below QCD-scale. HNL decay channels with a branching ratio of at least 1% for some mass below ~ 1 GeV are considered in this work (see Figure 3.18). The results for HNLs do not take into account charge conjugated channels, which are possible if they are Majorana particles.

The explicit determination of matrix elements involving multiple mesons can be extremely challenging. Therefore, an approximation has been used by assuming the matrix element to be constant and using the definition of decay width,

$$\Gamma = \frac{1}{2gM}\int \left(\prod_i \frac{\mathrm{d}^3 y_i}{(2\pi)^3 2E_i}\right)|\mathcal{M}|^2 (2\pi)^4 \delta^4(P - \sum_i P_i),$$

together with its measured value (from e.g. [88]) to solve for $|\mathcal{M}|^2$. For three-particle reactions this method gives the exact matrix element.

The values of the meson decay constants used in Subsection 3.B.3 are from [74] and summarized below.

f_{π^0}	$f_{\pi^\pm}$	f_η	f_{ρ^0}	$f_{\rho^\pm}$	f_ω	$f_{\eta'}$	f_ϕ	
130.2	130.2	81.7	208.9	208.9	195.5	-94.7	229.5	MeV

69
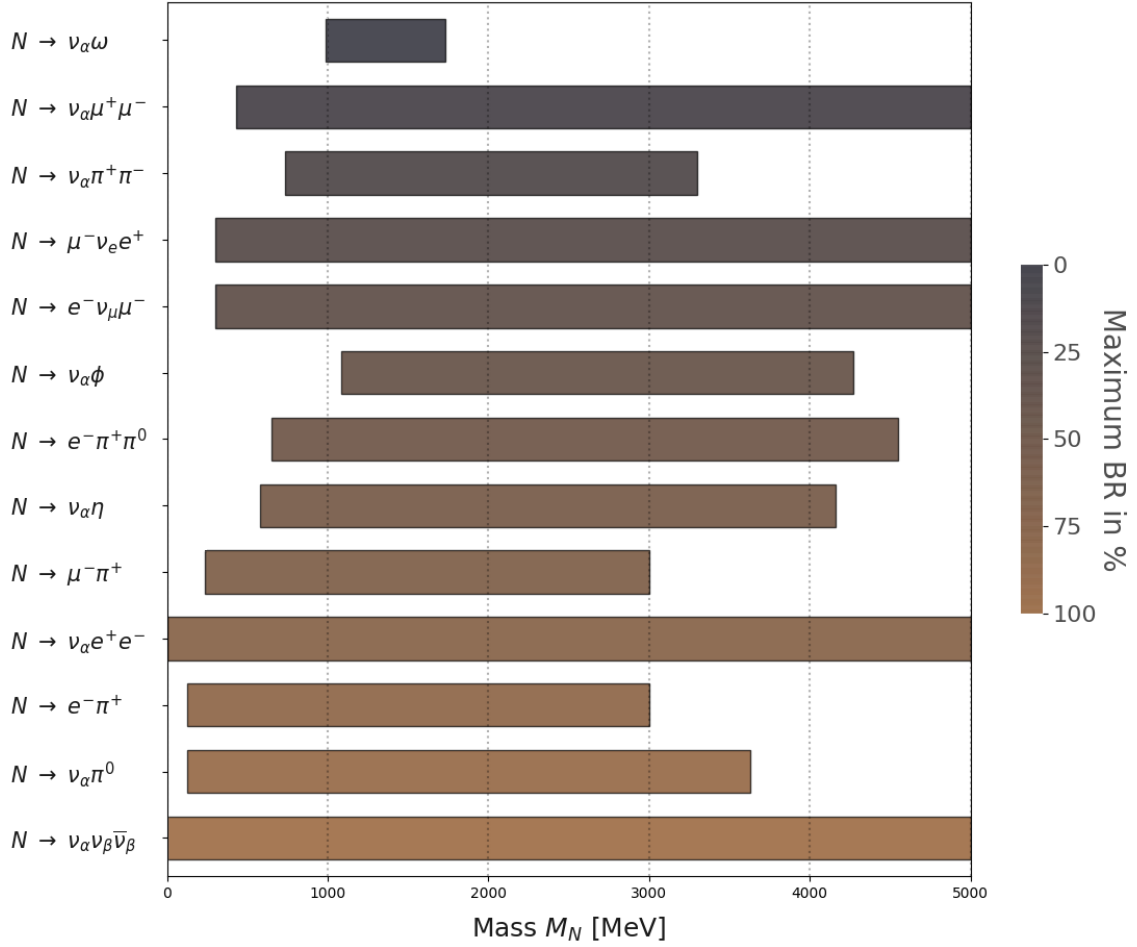
Figure 3.18: List of HNL decay channels with branching ratios more than 1% for some HNL mass below ~ 1 GeV. The left border indicates the HNL mass where the branching ratio exceeds 1%, the right border when it falls below the 1 % threshold. In this plot a model was assumed where all three mixing angles are equal to each other.

3.B.1 Matrix elements in the Standard Model

3.B.1.1 Four-particle processes with leptons

Process $(1 + 2 \to 3 + 4)$	S	$SG_F^2 a^{-4} \left| \mathcal{M} \right|^2$
$\nu_\alpha + \nu_\beta \to \nu_\alpha + \nu_\beta$	1	$32 \left(Y_1 \cdot Y_2 \right) \left(Y_3 \cdot Y_4 \right)$
$\nu_\alpha + \overline{\nu_\beta} \to \nu_\alpha + \overline{\nu_\beta}$	1	$32 \left(Y_1 \cdot Y_4 \right) \left(Y_2 \cdot Y_3 \right)$
$\nu_\alpha + \nu_\alpha \to \nu_\alpha + \nu_\alpha$	$\frac{1}{2}$	$64 \left(Y_1 \cdot Y_2 \right) \left(Y_3 \cdot Y_4 \right)$
$\nu_\alpha + \overline{\nu_\alpha} \to \nu_\alpha + \overline{\nu_\alpha}$	1	$128 \left(Y_1 \cdot Y_4 \right) \left(Y_2 \cdot Y_3 \right)$
$\nu_\alpha + \overline{\nu_\alpha} \to \nu_\beta + \overline{\nu_\beta}$	1	$32 \left(Y_1 \cdot Y_4 \right) \left(Y_3 \cdot Y_2 \right)$
$\nu_e + \overline{\nu_e} \to e^+ + e^-$	1	$128 \left[g_L^2 \left(Y_1 \cdot Y_3 \right) \left(Y_2 \cdot Y_4 \right) + g_R^2 \left(Y_1 \cdot Y_4 \right) \left(Y_2 \cdot Y_3 \right) \right.$ $\left. + g_L g_R a^2 m_e^2 \left(Y_1 \cdot Y_2 \right) \right]$
$\nu_e + e^- \to \nu_e + e^-$	1	$128 \left[g_L^2 \left(Y_1 \cdot Y_2 \right) \left(Y_3 \cdot Y_4 \right) + g_R^2 \left(Y_1 \cdot Y_4 \right) \left(Y_3 \cdot Y_2 \right) \right.$ $\left. - g_L g_R a^2 m_e^2 \left(Y_1 \cdot Y_3 \right) \right]$
$\nu_e + e^+ \to \nu_e + e^+$	1	$128 \left[g_L^2 \left(Y_1 \cdot Y_4 \right) \left(Y_3 \cdot Y_2 \right) + g_R^2 \left(Y_1 \cdot Y_2 \right) \left(Y_3 \cdot Y_4 \right) \right.$ $\left. - g_L g_R a^2 m_e^2 \left(Y_1 \cdot Y_3 \right) \right]$
$\nu_{\mu/\tau} + \overline{\nu_{\mu/\tau}} \to e^+ + e^-$	1	$128 \left[\widetilde{g_L}^2 \left(Y_1 \cdot Y_3 \right) \left(Y_2 \cdot Y_4 \right) + g_R^2 \left(Y_1 \cdot Y_4 \right) \left(Y_2 \cdot Y_3 \right) \right.$ $\left. + \widetilde{g_L} g_R a^2 m_e^2 \left(Y_1 \cdot Y_2 \right) \right]$
$\nu_{\mu/\tau} + e^- \to \nu_{\mu/\tau} + e^-$	1	$128 \left[\widetilde{g_L}^2 \left(Y_1 \cdot Y_2 \right) \left(Y_3 \cdot Y_4 \right) + g_R^2 \left(Y_1 \cdot Y_4 \right) \left(Y_3 \cdot Y_2 \right) \right.$ $\left. - \widetilde{g_L} g_R a^2 m_e^2 \left(Y_1 \cdot Y_3 \right) \right]$
$\nu_{\mu/\tau} + e^+ \to \nu_{\mu/\tau} + e^+$	1	$128 \left[\widetilde{g_L}^2 \left(Y_1 \cdot Y_4 \right) \left(Y_3 \cdot Y_2 \right) + g_R^2 \left(Y_1 \cdot Y_2 \right) \left(Y_3 \cdot Y_4 \right) \right.$ $\left. - \widetilde{g_L} g_R a^2 m_e^2 \left(Y_1 \cdot Y_3 \right) \right]$

Table 3.2: Squared matrix elements for weak processes involving active neutrinos and electrons/positrons. S is the symmetry factor and $\alpha, \beta \in \{e, \mu, \tau\}$, where $\alpha \neq \beta$. Here: $g_R = \sin^2 \theta_W$, $g_L = 1/2 + \sin^2 \theta_W$ and $\widetilde{g_L} = -1/2 + \sin^2 \theta_W$, with θ_W the Weinberg angle.

3.B.1.2 Three-particle and four-particle meson decays

Process ($1 \to 2 + 3$)	S	$|\mathcal{M}|^2$
$\pi^0 \to \gamma + \gamma$	1	$\alpha_{\text{em}}^2 m_\pi^4 \left[2\pi^2 f_\pi^2\right]^{-1}$
$\pi^+ \to \mu^+ + \nu_\mu$	1	$2G_{\text{F}}^2 \left|V_{ud}\right|^2 f_\pi^2 m_\mu^4 \left[\frac{m_\pi^2}{m_\mu^2} - 1\right]$

Table 3.3: Squared matrix elements for pion decays.

Process ($1 \to 2 + 3 + 4$)	$|\mathcal{M}|^2$	Process ($1 \to 2 + 3$)	$|\mathcal{M}|^2$ [MeV^2]
$K^+ \to \pi^0 + e^+ + \nu_e$	$1.42906 \cdot 10^{-13}$	$K^+ \to \pi^+ + \pi^0$	$3.28177 \cdot 10^{-10}$
$K^+ \to \pi^+ + \pi^- + \pi^+$	$1.85537 \cdot 10^{-12}$	$K^+ \to \mu^+ + \nu_\mu$	$8.78918 \cdot 10^{-10}$
$K_L^0 \to \pi^\pm + e^\mp + \nu_e$	$2.80345 \cdot 10^{-13}$	$K_S^0 \to \pi^+ + \pi^-$	$1.53713 \cdot 10^{-7}$
$K_L^0 \to \pi^\pm + \mu^\mp + \nu_\mu$	$3.03627 \cdot 10^{-13}$	$K_S^0 \to \pi^0 + \pi^0$	$6.71800 \cdot 10^{-8}$
$K_L^0 \to \pi^0 + \pi^0 + \pi^0$	$1.05573 \cdot 10^{-12}$	$\eta \to \gamma + \gamma$	$1.42174 \cdot 10^1$
$K_L^0 \to \pi^+ + \pi^- + \pi^0$	$8.26989 \cdot 10^{-13}$	$\rho^0 \to \pi^+ + \pi^-$	$1.86839 \cdot 10^7$
$\eta \to \pi^0 + \pi^0 + \pi^0$	$8.70984 \cdot 10^{-2}$	$\rho^+ \to \pi^+ + \pi^0$	$1.86390 \cdot 10^7$
$\eta \to \pi^+ + \pi^- + \pi^0$	$6.90629 \cdot 10^{-2}$	$\omega \to \pi^0 + \gamma$	$8.55086 \cdot 10^4$
$\eta \to \pi^+ + \pi^- + \gamma$	$4.66530 \cdot 10^{-3}$	$\eta' \to \rho^0 + \gamma$	$8.04463 \cdot 10^3$
$\omega \to \pi^+ + \pi^- + \pi^0$	$1.14569 \cdot 10^3$	$\phi \to K^+ + K^-$	$1.28798 \cdot 10^6$
$\eta' \to \pi^+ + \pi^- + \eta$	$4.38880 \cdot 10^1$	$\phi \to K_L^0 + K_S^0$	$1.03471 \cdot 10^6$
$\eta' \to \pi^0 + \pi^0 + \eta$	$2.00986 \cdot 10^1$	$\phi \to \rho^0 + \pi^0$	$2.86706 \cdot 10^5$

Table 3.4: Squared matrix elements for meson decays, where the constant matrix element approximation is used. For Majorana particles that can also decay through the charge-conjugated channel, the factor of 2 in the decay width is already taken into account here.

3.B.2 Matrix elements for Heavy Neutral Leptons above Λ_{QCD}

3.B.2.1 Four-particle processes with leptons only

Process $(1 + 2 \to 3 + 4)$	S	$SG_{\mathrm{F}}^{-2}a^{-4}\,\lvert\mathcal{M}\rvert^2$
$N + \nu_\beta \to \nu_\alpha + \nu_\beta$	1	$32\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4)$
$N + \overline{\nu_\beta} \to \nu_\alpha + \overline{\nu_\beta}$	1	$32\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$
$N + \nu_\alpha \to \nu_\alpha + \nu_\alpha$	$\frac{1}{2}$	$64\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4)$
$N + \overline{\nu_\alpha} \to \nu_\alpha + \overline{\nu_\alpha}$	1	$128\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$
$N + \overline{\nu_\alpha} \to \nu_\beta + \overline{\nu_\beta}$	1	$32\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_4)\,(Y_3 \cdot Y_2)$
$N + \overline{\nu_e} \to e^+ + e^-$	1	$128\,\lvert\theta_e\rvert^2\Big[g_L^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $+ g_L g_R a^2 m_e^2\,(Y_1 \cdot Y_2)\Big]$
$N + e^- \to \nu_e + e^-$	1	$128\,\lvert\theta_e\rvert^2\Big[g_L^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_3 \cdot Y_2)$ $- g_L g_R a^2 m_e^2\,(Y_1 \cdot Y_3)\Big]$
$N + e^+ \to \nu_e + e^+$	1	$128\,\lvert\theta_e\rvert^2\Big[g_L^2\,(Y_1 \cdot Y_4)\,(Y_3 \cdot Y_2) + g_R^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4)$ $- g_L g_R a^2 m_e^2\,(Y_1 \cdot Y_3)\Big]$
$N + \overline{\nu_{\mu/\tau}} \to e^+ + e^-$	1	$128\,\lvert\theta_{\mu/\tau}\rvert^2\Big[\widetilde{g_L}^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $+ \widetilde{g_L} g_R a^2 m_e^2\,(Y_1 \cdot Y_2)\Big]$
$N + e^- \to \nu_{\mu/\tau} + e^-$	1	$128\,\lvert\theta_{\mu/\tau}\rvert^2\Big[\widetilde{g_L}^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_3 \cdot Y_2)$ $- \widetilde{g_L} g_R a^2 m_e^2\,(Y_1 \cdot Y_3)\Big]$
$N + e^+ \to \nu_{\mu/\tau} + e^+$	1	$128\,\lvert\theta_{\mu/\tau}\rvert^2\Big[\widetilde{g_L}^2\,(Y_1 \cdot Y_4)\,(Y_3 \cdot Y_2) + g_R^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4)$ $- \widetilde{g_L} g_R a^2 m_e^2\,(Y_1 \cdot Y_3)\Big]$

Table 3.5: Squared matrix elements for weak processes involving HNLs and leptons. S is the symmetry factor and $\alpha, \beta \in \{e, \mu, \tau\}$, where $\alpha \neq \beta$. Here: $g_R = \sin^2\theta_{\mathrm{W}}$, $g_L = 1/2 + \sin^2\theta_{\mathrm{W}}$ and $\widetilde{g_L} = -1/2 + \sin^2\theta_{\mathrm{W}}$, with θ_{W} the Weinberg angle.

Process ($1 + 2 \rightarrow 3 + 4$)	S	$SG_{\mathrm{F}}^{-2}a^{-4}\,\lvert\mathcal{M}\rvert^2$
$N + \overline{\nu_\mu} \rightarrow e^- + \mu^+$	1	$128\,\lvert\theta_e\rvert^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$
$N + \overline{\nu_e} \rightarrow e^+ + \mu^-$	1	$128\,\lvert\theta_\mu\rvert^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4)$
$N + e^- \rightarrow \nu_e + \mu^-$	1	$128\,\lvert\theta_\mu\rvert^2\,(Y_1 \cdot Y_2)\,(Y_3 \cdot Y_4)$
$N + e^+ \rightarrow \nu_\mu + \mu^+$	1	$128\,\lvert\theta_e\rvert^2\,(Y_1 \cdot Y_4)\,(Y_3 \cdot Y_2)$
$N + \overline{\nu_\mu} \rightarrow \mu^+ + \mu^-$	1	$128\,\lvert\theta_\mu\rvert^2\Big[g_L^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $\quad + g_L g_R a^2 m_\mu^2\,(Y_1 \cdot Y_2)\Big]$
$N + \overline{\nu_{e/\tau}} \rightarrow \mu^+ + \mu^-$	1	$128\,\lvert\theta_{e/\tau}\rvert^2\Big[\widetilde{g_L}^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) \qquad\qquad +$ $g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $\quad + \widetilde{g_L} g_R a^2 m_\mu^2\,(Y_1 \cdot Y_2)\Big]$

Process ($1 \rightarrow 2 + 3 + 4$)	S	$SG_{\mathrm{F}}^{-2}a^{-4}\,\lvert\mathcal{M}\rvert^2$
$N \rightarrow \nu_\alpha + \nu_\beta + \overline{\nu_\beta}$	1	$32\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$
$N \rightarrow \nu_\alpha + \nu_\alpha + \overline{\nu_\alpha}$	$\frac{1}{2}$	$64\,\lvert\theta_\alpha\rvert^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$
$N \rightarrow \nu_e + e^+ + e^-$	1	$128\,\lvert\theta_e\rvert^2\Big[g_L^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $\quad + g_L g_R a^2 m_e^2\,(Y_1 \cdot Y_2)\Big]$
$N \rightarrow \nu_{\mu/\tau} + e^+ + e^-$	1	$128\,\lvert\theta_{\mu/\tau}\rvert^2\Big[\widetilde{g_L}^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) \qquad\qquad +$ $g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $\quad + \widetilde{g_L} g_R a^2 m_e^2\,(Y_1 \cdot Y_2)\Big]$
$N \rightarrow \nu_\mu + e^- + \mu^+$	1	$128\,\lvert\theta_e\rvert^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$
$N \rightarrow \nu_e + e^+ + \mu^-$	1	$128\,\lvert\theta_\mu\rvert^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4)$
$N \rightarrow \nu_\mu + \mu^+ + \mu^-$	1	$128\,\lvert\theta_\mu\rvert^2\Big[g_L^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) + g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $\quad + g_L g_R a^2 m_\mu^2\,(Y_1 \cdot Y_2)\Big]$
$N \rightarrow \nu_{e/\tau} + \mu^+ + \mu^-$	1	$128\,\lvert\theta_{e/\tau}\rvert^2\Big[\widetilde{g_L}^2\,(Y_1 \cdot Y_3)\,(Y_2 \cdot Y_4) \qquad\qquad +$ $g_R^2\,(Y_1 \cdot Y_4)\,(Y_2 \cdot Y_3)$ $\quad + \widetilde{g_L} g_R a^2 m_\mu^2\,(Y_1 \cdot Y_2)\Big]$

Table 3.6: Squared matrix elements for weak processes involving HNLs and leptons. Note: low temperatures are assumed here. At high temperatures, reactions such as $N + \mu^- \rightarrow e^- + \nu_\mu$ are possible. The corresponding matrix elements can be trivially deduced from the ones given above.

3.B.2.2 Four-particle processes with leptons and quarks

Process $(1 + 2 \to 3 + 4)$	S	$SG_{\mathrm{F}}^{-2} a^{-4} \left| \mathcal{M} \right|^2$
$N + \ell_\alpha^+ \to U + \overline{D}$	1	$128 \left| \theta_\alpha \right|^2 \left| V_{ud} \right|^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3)$
$N + D \to \ell_\alpha^- + U$	1	$128 \left| \theta_\alpha \right|^2 \left| V_{ud} \right|^2 (Y_1 \cdot Y_2)(Y_3 \cdot Y_4)$
$N + \overline{U} \to \ell_\alpha^- + \overline{D}$	1	$128 \left| \theta_\alpha \right|^2 \left| V_{ud} \right|^2 (Y_1 \cdot Y_4)(Y_3 \cdot Y_2)$
$N + \overline{\nu_\alpha} \to \overline{U} + U$	1	$\frac{32}{9} \left| \theta_\alpha \right|^2 \Big[16 g_R^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3)$ $+$ $(3 - 4g_R)^2 (Y_1 \cdot Y_3)(Y_2 \cdot Y_4)$ $+ 4 g_R \theta_W (4 g_R - 3) a^2 m_U^2 (Y_1 \cdot Y_2) \Big]$
$N + U \to \nu_\alpha + U$	1	$\frac{32}{9} \left| \theta_\alpha \right|^2 \Big[16 g_R^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3)$ $+$ $(3 - 4g_R)^2 (Y_1 \cdot Y_2)(Y_3 \cdot Y_4)$ $- 4 g_R (4 g_R - 3) a^2 m_U^2 (Y_1 \cdot Y_3) \Big]$
$N + \overline{U} \to \nu_\alpha + \overline{U}$	1	$\frac{32}{9} \left| \theta_\alpha \right|^2 \Big[16 g_R^2 (Y_1 \cdot Y_2)(Y_3 \cdot Y_4)$ $+$ $(3 - 4g_R)^2 (Y_1 \cdot Y_4)(Y_3 \cdot Y_2)$ $- 4 g_R (4 g_R - 3) a^2 m_U^2 (Y_1 \cdot Y_3) \Big]$
$N + \overline{\nu_\alpha} \to \overline{D} + D$	1	$\frac{32}{9} \left| \theta_\alpha \right|^2 \Big[4 g_R^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3)$ $+$ $(3 - 2g_R)^2 (Y_1 \cdot Y_3)(Y_2 \cdot Y_4)$ $+ 2 g_R (2 g_R - 3) a^2 m_D^2 (Y_1 \cdot Y_2) \Big]$
$N + D \to \nu_\alpha + D$	1	$\frac{32}{9} \left| \theta_\alpha \right|^2 \Big[4 g_R^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3)$ $+$ $(3 - 2g_R)^2 (Y_1 \cdot Y_2)(Y_3 \cdot Y_4)$ $- 2 g_R (2 g_R - 3) a^2 m_D^2 (Y_1 \cdot Y_3) \Big]$
$N + \overline{D} \to \nu_\alpha + \overline{D}$	1	$\frac{32}{9} \left| \theta_\alpha \right|^2 \Big[4 g_R^2 (Y_1 \cdot Y_2)(Y_3 \cdot Y_4)$ $+$ $(3 - 2g_R)^2 (Y_1 \cdot Y_4)(Y_3 \cdot Y_2)$ $- 2 g_R (2 g_R - 3) a^2 m_D^2 (Y_1 \cdot Y_3) \Big]$

Table 3.7: Squared matrix elements for weak scattering processes involving HNLs, leptons and quarks. Here: U are up-type quarks, D down-type quarks and $g_R = \sin^2 \theta_{\mathrm{W}}$.

Process ($1 \to 2 + 3 + 4$)	S	$SG_{\mathrm{F}}^{-2} a^{-4} \left|\mathcal{M}\right|^2$
$N \to \ell_\alpha^- + U + \overline{D}$	1	$128 \left|\theta_\alpha\right|^2 \left|V_{ud}\right|^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3)$
$N \to \nu_\alpha + \overline{U} + U$	1	$\frac{32}{9} \left|\theta_\alpha\right|^2 \Big[16 g_R^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3) \qquad\qquad +$ $(3 - 4 g_R)^2 (Y_1 \cdot Y_3)(Y_2 \cdot Y_4)$ $+ 4 g_R (4 g_R - 3) a^2 m_U^2 (Y_1 \cdot Y_2) \Big]$
$N \to \nu_\alpha + \overline{D} + D$	1	$\frac{32}{9} \left|\theta_\alpha\right|^2 \Big[4 g_R^2 (Y_1 \cdot Y_4)(Y_2 \cdot Y_3) \qquad\qquad +$ $(3 - 2 g_R)^2 (Y_1 \cdot Y_3)(Y_2 \cdot Y_4)$ $+ 2 g_R (2 g_R - 3) a^2 m_D^2 (Y_1 \cdot Y_2) \Big]$

Table 3.8: Squared matrix elements for weak decay processes involving HNLs, leptons and quarks. Here: U are up-type quarks, D down-type quarks and $g_R = \sin^2 \theta_{\mathrm{W}}$.

3.B.3 Matrix elements for Heavy Neutral Leptons below Λ_{QCD}

In addition to interactions with leptons, HNLs will also decay into mesons.

3.B.3.1 Three-particle processes with single mesons

Process ($1 \to 2 + 3$)	S	$SG_{\mathrm{F}}^{-2} M_N^{-4} \left|\mathcal{M}\right|^2$
$N \to \nu_\alpha + \pi^0$	1	$\left|\theta_\alpha\right|^2 f_\pi^2 \left[1 - \frac{m_\pi^2}{M_N^2} \right]$
$N \to \ell_\alpha^\mp + \pi^\pm$	1	$2 \left|\theta_\alpha\right|^2 \left|V_{ud}\right|^2 f_\pi^2 \left[\left(1 - \frac{m_{\ell_\alpha}^2}{M_N^2} \right)^2 - \frac{m_\pi^2}{M_N^2} \left(1 + \frac{m_{\ell_\alpha}^2}{M_N^2} \right) \right]$
$N \to \nu_\alpha + \eta$	1	$\left|\theta_\alpha\right|^2 f_\eta^2 \left[1 - \frac{m_\eta^2}{M_N^2} \right]$
$N \to \nu_\alpha + \rho^0$	1	$\left|\theta_\alpha\right|^2 \left(1 - 2 \sin^2 \theta_{\mathrm{W}} \right)^2 f_\rho^2 \left[1 + 2 \frac{m_\rho^2}{M_N^2} \right] \left[1 - \frac{m_\rho^2}{M_N^2} \right]$
$N \to \ell_\alpha^\mp + \rho^\pm$	1	$2 \left|\theta_\alpha\right|^2 \left|V_{ud}\right|^2 f_\rho^2 \left[\left(1 - \frac{m_{\ell_\alpha}^2}{M_N^2} \right)^2 + \frac{m_\rho^2}{M_N^2} \left(1 + \frac{m_{\ell_\alpha}^2}{M_N^2} \right) - 2 \frac{m_\rho^4}{M_N^4} \right]$
$N \to \nu_\alpha + \omega$	1	$\left|\theta_\alpha\right|^2 \left(\frac{4}{3} \sin^2 \theta_{\mathrm{W}} \right)^2 f_\omega^2 \left[1 + 2 \frac{m_\omega^2}{M_N^2} \right] \left[1 - \frac{m_\omega^2}{M_N^2} \right]$
$N \to \nu_\alpha + \eta'$	1	$\left|\theta_\alpha\right|^2 f_{\eta'}^2 \left[1 - \frac{m_{\eta'}^2}{M_N^2} \right]$
$N \to \nu_\alpha + \phi$	1	$\left|\theta_\alpha\right|^2 \left(\frac{4}{3} \sin^2 \theta_{\mathrm{W}} - 1 \right)^2 f_\phi^2 \left[1 + 2 \frac{m_\phi^2}{M_N^2} \right] \left[1 - \frac{m_\phi^2}{M_N^2} \right]$

Table 3.9: Squared matrix elements for HNL decays into mesons.

3.C   Collision integrals

Consider the Boltzmann equation in comoving coordinates:

$$\frac{\mathrm{d}f_1}{\mathrm{d}t} = \frac{\mathrm{d}f_1}{\mathrm{d}\ln a}\frac{\mathrm{d}\ln a}{\mathrm{d}t} = \frac{\mathrm{d}f_1}{\mathrm{d}\ln a}H = \sum_{\mathrm{reactions}} I_{\mathrm{coll}} \,, \tag{3.C.1}$$

with

$$I_{\mathrm{coll}} = \frac{a^{7-2Q}}{2g\widetilde{E_1}} \sum_{\mathrm{in,out}} \int \left(\prod_{i=2}^{Q} \frac{\mathrm{d}^3 y_i}{(2\pi)^3 2\widetilde{E_i}} \right) S|\mathcal{M}|^2 F[f](2\pi)^4 \delta^4(Y_{\mathrm{in}} - Y_{\mathrm{out}}) \tag{3.C.2}$$

The delta function can be rewritten as

$$\delta^4(Y_{\mathrm{in}} - Y_{\mathrm{out}}) = \delta^4(s_1 Y_1 + s_2 Y_2 + ... + s_Q Y_Q) \,, \tag{3.C.3}$$

with $s_i = \{-1, 1\}$ if particle i is on the {left, right}-hand side of the reaction. The $Y_i = aP_i$ here are the comoving four-momenta.

3.C.1   Three-particle collision integral

$$I_{\mathrm{coll}} = \frac{a}{2\widetilde{E_1}} \int \frac{\mathrm{d}^3 y_2 \mathrm{d}^3 y_3}{(2g\pi)^6 2\widetilde{E_2} 2\widetilde{E_3}} S|\mathcal{M}|^2 F[f](2\pi)^4 \delta^4(s_1 Y_1 + s_2 Y_2 + s_3 Y_3) \tag{3.C.4}$$

3.C.1.1   Case $y_1 \neq 0$

Since a homogeneous and isotropic universe is assumed, only absolute values of momenta are relevant. Moreover, the matrix element in three particle interactions is independent of the four-momenta. The collision integral becomes:

$$I_{\mathrm{coll}} = \frac{S|\mathcal{M}|^2 a}{8(2\pi)^2 g\widetilde{E_1}} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 \mathrm{d}\Omega_2 \mathrm{d}\Omega_3 y_2^2 y_3^2}{\widetilde{E_2}\widetilde{E_3}} F[f]\delta^4(s_1 Y_1 + s_2 Y_2 + s_3 Y_3) \tag{3.C.5}$$

Using the identity

$$\delta^3(s_1 \mathbf{y_1} + s_2 \mathbf{y_2} + s_3 \mathbf{y_3}) = \frac{1}{(2\pi)^3} \int \mathrm{d}\lambda \mathrm{d}\Omega_\lambda \lambda^2 e^{i(s_1 \mathbf{y_1} + s_2 \mathbf{y_2} + s_3 \mathbf{y_3})\cdot\boldsymbol{\lambda}} \tag{3.C.6}$$

gives

$$
\begin{aligned}
I_{\text{coll}} =& \frac{S|\mathcal{M}|^2 a}{8(2\pi)^5 g\widetilde{E_1}} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 y_2^2 y_3^2}{\widetilde{E_2}\widetilde{E_3}} F[f] \delta(s_1\widetilde{E_1} + s_2\widetilde{E_2} + s_3\widetilde{E_3}) \cdot \\
& \cdot \int \mathrm{d}\lambda \lambda^2 \int \mathrm{d}\Omega_\lambda e^{is_1 y_1 \lambda \cos\theta_\lambda} \int \mathrm{d}\Omega_2 e^{is_1 y_2 \lambda \cos\theta_2} \int \mathrm{d}\Omega_3 e^{is_1 y_3 \lambda \cos\theta_3} \\
=& \frac{S|\mathcal{M}|^2 a}{8(2\pi)^5 g\widetilde{E_1}} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 y_2^2 y_3^2}{\widetilde{E_2}\widetilde{E_3}} F[f] \delta(s_1\widetilde{E_1} + s_2\widetilde{E_2} + s_3\widetilde{E_3}) \cdot \\
& \cdot \int \mathrm{d}\lambda \lambda^2 \left(4\pi \frac{\sin(y_1\lambda)}{y_1\lambda} \right)\left(4\pi \frac{\sin(y_2\lambda)}{y_2\lambda} \right)\left(4\pi \frac{\sin(y_3\lambda)}{y_3\lambda} \right) \\
=& \frac{S|\mathcal{M}|^2 a}{(2\pi)^2 g\widetilde{E_1}y_1} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 y_2 y_3}{\widetilde{E_2}\widetilde{E_3}} F[f] \delta(s_1\widetilde{E_1} + s_2\widetilde{E_2} + s_3\widetilde{E_3}) \cdot \\
& \cdot \int \frac{\mathrm{d}\lambda}{\lambda} \sin(y_1\lambda)\sin(y_2\lambda)\sin(y_3\lambda)
\end{aligned}
\tag{3.C.7}
$$

Rewrite the delta function of energies as

$$
\begin{aligned}
\int \frac{\mathrm{d}y_3 y_3}{\widetilde{E_3}} \delta(s_1\widetilde{E_1} + s_2\widetilde{E_2} + s_3\widetilde{E_3}) &= \int \mathrm{d}y_3 \frac{y_3}{\widetilde{E_3}} \frac{\delta(y_3 - y_3^*)}{\frac{y_3^*}{\widetilde{E_3^*}}} \theta\left(\left((s_1\widetilde{E_1} + s_2\widetilde{E_2})\right)^2 - a^2 m_3^2 \right) \\
&= \int \mathrm{d}y_3 \frac{y_3}{\widetilde{E_3}} \frac{\widetilde{E_3^*}}{y_3^*} \delta(y_3 - y_3^*)\,\theta\left(\left(\widetilde{E_3^*}\right)^2 - a^2 m_3^2 \right),
\end{aligned}
\tag{3.C.8}
$$

where $\left(\widetilde{E_3^*}\right)^2 = (y_3^*)^2 + \frac{x^2}{M^2}m_3^2 = \left(s_1\widetilde{E_1} + s_2\widetilde{E_2}\right)^2$ and $y_3^* = \sqrt{(s_1\widetilde{E_1} + s_2\widetilde{E_2})^2 - a^2 m_3^2}$.

Plugging Eq. (3.C.8) in Eq. (3.C.7) above:

$$
I_{\text{coll}} = \frac{S|\mathcal{M}|^2 a}{(2\pi)^2 g\widetilde{E_1}y_1} \int \frac{\mathrm{d}y_2 y_2}{\widetilde{E_2}} F[f] \int \frac{\mathrm{d}\lambda}{\lambda} \sin(y_1\lambda)\sin(y_2\lambda)\sin(y_3^*\lambda)\theta\left(\left(\widetilde{E_3^*}\right)^2 - a^2 m_3^2 \right)
\tag{3.C.9}
$$

Now, the integral over λ is equal to

$$
\mathcal{X} = \frac{\pi}{8} \left(-\mathrm{Sgn}[y_1 - y_2 - y_3^*] + \mathrm{Sgn}[y_1 + y_2 - y_3^*] + \mathrm{Sgn}[y_1 - y_2 + y_3^*] - 1 \right),
\tag{3.C.10}
$$

with Sgn the signum function and where $y_1 \geq y_2 \geq y_3$ is assumed.
The final form is then

$$
I_{\text{coll}} = \frac{S|\mathcal{M}|^2 a}{(2\pi)^2 g\widetilde{E_1}y_1} \int \frac{\mathrm{d}y_2 y_2}{\widetilde{E_2}} \mathcal{X}\theta\left(\left(s_1\widetilde{E_1} + s_2\widetilde{E_2}\right)^2 - a^2 m_3^2 \right) (F[f])\bigg|_{y_3 = y_3^*}
\tag{3.C.11}
$$

78

3.C.1.2 Case $y_1 = 0$

$$
\begin{aligned}
I_{\text{coll}} =& \frac{S|\mathcal{M}|^2 a}{8(2\pi)^2 gam_1} \int \frac{\mathrm{d}^3 y_2 \mathrm{d}^3 y_3}{\widetilde{E_2}\widetilde{E_3}} F[f]\delta\left(s_1 am_1 + s_2 \widetilde{E_2} + s_3 \widetilde{E_3}\right)\delta^3\left(s_2 \mathbf{y_2} + s_3 \mathbf{y_3}\right) \\
=& \frac{S|\mathcal{M}|^2}{8(2\pi)^2 gm_1} \int \mathrm{d}^3 y_2 F[f]\delta\left(s_1 am_1 + s_2\sqrt{y_2^2 + (am_2)^2} + s_3\sqrt{y_2^2 + (am_3)^2}\right) \cdot \\
& \cdot \left(\sqrt{y_2^2 + (am_2)^2}\sqrt{y_2^2 + (am_3)^2}\right)^{-1} \\
=& \frac{S|\mathcal{M}|^2}{8\pi gm_1} \int \mathrm{d}y_2 y_2^2 F[f]\delta\left(y_2 - y_2^*\right)\left|\frac{s_2 y_2^*}{\sqrt{(y_2^*)^2 + a^2 m_2^2}} + \frac{s_3 y_2^*}{\sqrt{(y_2^*)^2 + a^2 m_3^2}}\right|^{-1} \cdot \\
& \cdot \left(\sqrt{y_2^2 + (am_2)^2}\sqrt{y_2^2 + (am_3)^2}\right)^{-1}\theta\left(\left(s_1 am_1 + s_2 \widetilde{E_2}\right)^2 - a^2 m_3^2\right) \\
=& \frac{S|\mathcal{M}|^2}{8\pi gm_1} y_2^* \left|s_2\sqrt{(y_2^*)^2 + (am_3)^2} + s_3\sqrt{(y_2^*)^2 + (am_2)^2}\right|^{-1}\theta\left(\left(\widetilde{E_3^*}\right)^2 - a^2 m_3^2\right) \cdot \\
& \cdot (F[f])\Bigg|_{y_1=0,\ y_2=y_2^*,\ y_3=-y_2^*} \\
=& \frac{S|\mathcal{M}|^2}{8\pi gm_1} y_2^* \left|s_1 s_2 s_3 am_1\right|^{-1}\theta\left(\left(\widetilde{E_3^*}\right)^2 - a^2 m_3^2\right)(F[f])\Bigg|_{y_1=0,\ y_2=y_2^*,\ y_3=-y_2^*} \\
=& \frac{S|\mathcal{M}|^2}{8\pi gm_1^2}\frac{y_2^*}{a}\theta\left(\left(\widetilde{E_3^*}\right)^2 - a^2 m_3^2\right)(F[f])\Bigg|_{y_1=0,\ y_2=y_2^*,\ y_3=-y_2^*} \quad, \quad (3.\text{C}.12)
\end{aligned}
$$

with $\left(\widetilde{E_3^*}\right)^2 = \left(s_1 am_1 + s_2 \widetilde{E_2}\right)^2$ and $y_2^* = a\sqrt{\frac{\left(m_1^2 - m_2^2 - m_3^2\right)^2 - 4m_2^2 m_3^2}{4m_1^2}}$.

3.C.2 Four-particle collision integral

$$
I_{\text{coll}} = \frac{1}{2g\widetilde{E_1}}\frac{1}{a} \int \frac{\mathrm{d}^3 y_2 \mathrm{d}^3 y_3 \mathrm{d}y_4^3}{(2\pi)^9 8\widetilde{E_2}\widetilde{E_3}\widetilde{E_4}} S|\mathcal{M}|^2 F[f](2\pi)^4 \delta^4(s_1 Y_1 + s_2 Y_2 + s_3 Y_3 + s_4 Y_4)
$$

$$(3.\text{C}.13)$$

As can be seen in Appendix 3.B, $|\mathcal{M}|^2$ can be written as

$$
|\mathcal{M}|^2 = \frac{1}{a^4}\sum_{i\neq j\neq k\neq l}\left[K_1(Y_i \cdot Y_j)(Y_k \cdot Y_l) + K_2 a^2 m_i m_j (Y_k \cdot Y_l)\right] \quad (3.\text{C}.14)
$$

A similar procedure as with the three-particle case is followed here.

3.C.2.1   Case $y_1 \neq 0$

$$
\begin{aligned}
I_{\text{coll}} =& \frac{S}{16(2\pi)^5 g \widetilde{E_1} a} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 \mathrm{d}y_4 y_2^2 y_3^2 y_4^2}{\widetilde{E_2}\widetilde{E_3}\widetilde{E_4}} F[f]\delta\left(s_1\widetilde{E_1} + s_2\widetilde{E_2} + s_3\widetilde{E_3} + s_4\widetilde{E_4}\right) \cdot \\
&\cdot \int \mathrm{d}\Omega_2 \mathrm{d}\Omega_3 \mathrm{d}\Omega_4 \left|\mathcal{M}\right|^2 |\delta^3\left(s_1\mathbf{y_1} + s_2\mathbf{y_2} + s_3\mathbf{y_3} + s_4\mathbf{y_4}\right) \\
=& \frac{S}{64\pi^3 g \widetilde{E_1} y_1 a^5} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 \mathrm{d}y_4 y_2 y_3 y_4}{\widetilde{E_2}\widetilde{E_3}\widetilde{E_4}} F[f]\delta\left(s_1\widetilde{E_1} + s_2\widetilde{E_2} + s_3\widetilde{E_3} s_4\widetilde{E_4}\right) \cdot \\
&\cdot \mathcal{D}(Y_1, Y_2, Y_3, Y_4) \,, \tag{3.C.15}
\end{aligned}
$$

with

$$
\begin{aligned}
\mathcal{D}(Y_1, Y_2, Y_3, Y_4) =& \frac{y_1 y_2 y_3 y_4}{64\pi^5} \int \mathrm{d}\Omega_2 \mathrm{d}\Omega_3 \mathrm{d}\Omega_4 \left|\mathcal{M}\right|^2 |\delta^3\left(s_1\mathbf{y_1} + s_2\mathbf{y_2} + s_3\mathbf{y_3} + s_4\mathbf{y_4}\right) \\
=& \frac{y_1 y_2 y_3 y_4}{64\pi^5} \int \mathrm{d}\lambda\lambda^2 \int \mathrm{d}\Omega_\lambda e^{is_1\mathbf{y_1}\cdot\boldsymbol{\lambda}} \int \mathrm{d}\Omega_2 e^{is_2\mathbf{y_2}\cdot\boldsymbol{\lambda}} \int \mathrm{d}\Omega_3 e^{is_3\mathbf{y_3}\cdot\boldsymbol{\lambda}} \cdot \\
&\cdot \int \mathrm{d}\Omega_4 e^{is_4\mathbf{y_4}\cdot\boldsymbol{\lambda}} \sum_{i\neq j\neq k\neq l} \left[K_1(Y_i\cdot Y_j)(Y_k\cdot Y_l) + K_2 a^2 m_i m_j (Y_k\cdot Y_l)\right] \\
=& \frac{y_1 y_2 y_3 y_4}{64\pi^5} \sum_{i\neq j\neq k\neq l} \int \mathrm{d}\lambda\lambda^2 \int \mathrm{d}\Omega_\lambda e^{is_i y_i \lambda\cos\theta_i} \int \mathrm{d}\Omega_j e^{is_j y_j \lambda\cos\theta_j} \cdot \\
&\cdot \int \mathrm{d}\Omega_k e^{is_k y_k \lambda\cos\theta_k} \int \mathrm{d}\Omega_l e^{is_l y_l \lambda\cos\theta_l} \left[K_1(Y_i\cdot Y_j)(Y_k\cdot Y_l) + \right. \\
&\left. + K_2 a^2 m_i m_j (Y_k\cdot Y_l)\right] \tag{3.C.16}
\end{aligned}
$$

Working out the inner products

$$
\begin{aligned}
Y_i \cdot Y_j &= \widetilde{E_i}\widetilde{E_j} - \mathbf{y_i}\cdot\mathbf{y_j} = \widetilde{E_i}\widetilde{E_j} - y_i y_j \cos\theta_{ij} \\
&= \widetilde{E_i}\widetilde{E_j} - y_i y_j \left(\cos\theta_i \cos\theta_j + \cos(\phi_i - \phi_j)\sin\theta_i \sin\theta_j\right) \,, \tag{3.C.17}
\end{aligned}
$$

where θ_{ij} is the angle between vectors $\mathbf{y_i}$ and $\mathbf{y_j}$, and using that

$$
\int_0^\pi \int_0^{2\pi} \mathrm{d}\theta_i \mathrm{d}\phi_i e^{is_i y_i \lambda\cos\theta_i} \sin^2\theta_i \cos(\phi_i - \phi_j) = 0 \tag{3.C.18}
$$

gives

$$\mathcal{D}(Y_1, Y_2, Y_3, Y_4) = \frac{y_1 y_2 y_3 y_4}{64\pi^5} \sum_{i \neq j \neq k \neq l} \int \mathrm{d}\lambda \lambda^2 \int \mathrm{d}\theta_i \mathrm{d}\phi_i \sin\theta_i e^{is_i y_i \lambda \cos\theta_i} \cdot \tag{3.C.19}$$

$$\cdot \int \mathrm{d}\theta_j \mathrm{d}\phi_j \sin\theta_j e^{is_j y_j \lambda \cos\theta_j} \int \mathrm{d}\theta_k \mathrm{d}\phi_k \sin\theta_k e^{is_k y_k \lambda \cos\theta_k} \cdot$$

$$\cdot \int \mathrm{d}\theta_l \mathrm{d}\phi_l \sin\theta_l e^{is_l y_l \lambda \cos\theta_l} \left[K_1 \left(\widetilde{E}_i \widetilde{E}_j - y_i y_j \cos\theta_i \cos\theta_j \right) \cdot \right.$$

$$\left. \cdot \left(\widetilde{E}_k \widetilde{E}_l - y_k y_l \cos\theta_k \cos\theta_l \right) + K_2 a^2 m_i m_j \left(\widetilde{E}_k \widetilde{E}_l - y_k y_l \cos\theta_k \cos\theta_l \right) \right]$$

The integrals over the angles are given by

$$\int_0^\pi \int_0^{2\pi} \mathrm{d}\theta \mathrm{d}\phi \sin\theta e^{isy\lambda\cos\theta} = 4\pi \frac{\sin(y\lambda)}{y\lambda} \tag{3.C.20}$$

$$\int_0^\pi \int_0^{2\pi} \mathrm{d}\theta \mathrm{d}\phi \sin\theta \cos\theta e^{isy\lambda\cos\theta} = \frac{4\pi}{isy\lambda} \left[\cos(y\lambda) - \frac{\sin(y\lambda)}{y\lambda} \right] \tag{3.C.21}$$

$$\tag{3.C.22}$$

and working out all the brackets gives

$$\mathcal{D}(Y_1, Y_2, Y_3, Y_4) = \sum_{i \neq j \neq k \neq l} \left[K_1 \left\{ \widetilde{E}_1 \widetilde{E}_2 \widetilde{E}_3 \widetilde{E}_4 D_1 \left(y_1, y_2, y_3, y_4 \right) + \widetilde{E}_i \widetilde{E}_j D_2 \left(y_i, y_j, y_k, y_l \right) + \right. \right.$$

$$+ \widetilde{E}_k \widetilde{E}_l D_2 \left(y_k, y_l, y_i, y_j \right) + D_3 \left(y_1, y_2, y_3, y_4 \right) \Big\} +$$

$$+ K_2 a^2 m_i m_j \left\{ \widetilde{E}_k \widetilde{E}_l D_1 \left(y_1, y_2, y_3, y_4 \right) + D_2 \left(y_i, y_j, y_k, y_l \right) \right\} \Bigg],$$

$$\tag{3.C.23}$$

with

$$D_1 (y_i, y_j, y_k, y_l) = \frac{4}{\pi} \int \frac{\mathrm{d}\lambda}{\lambda^2} \sin(y_i\lambda) \sin(y_j\lambda) \sin(y_k\lambda) \sin(y_l\lambda) \tag{3.C.24}$$

$$D_2 (y_i, y_j, y_k, y_l) = s_k s_l \frac{4 y_k y_l}{\pi} \int \frac{\mathrm{d}\lambda}{\lambda^2} \sin(y_i\lambda) \sin(y_j\lambda) \left[\cos(y_k\lambda) - \frac{\sin(y_k\lambda)}{y_k\lambda} \right] \cdot$$

$$\cdot \left[\cos(y_l\lambda) - \frac{\sin(y_l\lambda)}{y_l\lambda} \right] \tag{3.C.25}$$

$$D_3 (y_i, y_j, y_k, y_l) = s_i s_j s_k s_l \frac{4 y_i y_j y_k y_l}{\pi} \int \frac{\mathrm{d}\lambda}{\lambda^2} \left[\cos(y_i\lambda) - \frac{\sin(y_i\lambda)}{y_i\lambda} \right] \left[\cos(y_j\lambda) - \frac{\sin(y_j\lambda)}{y_j\lambda} \right] \cdot$$

$$\cdot \left[\cos(y_k\lambda) - \frac{\sin(y_k\lambda)}{y_k\lambda} \right] \left[\cos(y_l\lambda) - \frac{\sin(y_l\lambda)}{y_l\lambda} \right] \tag{3.C.26}$$

All these three functions are symmetric under the exchange $y_i \leftrightarrow y_j$ and $y_k \leftrightarrow y_l$, which then allows to take $y_i > y_j$ and $y_k > y_l$. Integrating out λ gives the functions in terms of

polynomials for all possible cases (factors $s_k s_l$ and $s_i s_j s_k s_l$ omitted):

- $y_i > y_j + y_k + y_l$ or $y_k > y_i + y_j + y_l$:

$$D_1 = D_2 = D_3 = 0$$

- $y_i + y_j > y_k + y_l$ and $y_i + y_l < y_j + y_k$:

$$D_1 = y_l$$
$$D_2 = \frac{1}{3} y_l^3$$
$$D_3 = \frac{1}{30} y_l^3 \left[5 \left(y_i^2 + y_j^2 + y_k^2 \right) - y_l^2 \right]$$

- $y_i + y_j > y_k + y_l$ and $y_i + y_l > y_j + y_k$:

$$D_1 = \frac{1}{2} (y_j + y_k + y_l - y_i)$$
$$D_2 = \frac{1}{12} \left[(y_i - y_j) \left\{ (y_i - y_j)^2 - 3 \left(y_k^2 + y_l^2 \right) \right\} + 2 \left(y_k^3 + y_l^3 \right) \right]$$
$$D_3 = \frac{1}{60} \left[y_i^5 - y_j^5 - y_k^5 - y_l^5 + 5 \left(-y_i^3 y_j^2 + y_i^2 y_j^3 - y_i^3 y_k^2 + y_i^2 y_k^3 \right. \right.$$
$$\left. \left. - y_i^3 y_l^2 + y_i^2 y_l^3 + y_j^3 y_k^2 + y_j^2 y_k^3 + y_j^3 y_l^2 + y_j^2 y_l^3 + y_k^3 y_l^2 + y_k^2 y_l^3 \right) \right]$$

- $y_i + y_j < y_k + y_l$ and $y_i + y_l > y_j + y_k$:

$$D_1 = y_j$$
$$D_2 = \frac{1}{6} y_j \left[3 \left(y_k^2 + y_l^2 - y_i^2 \right) - y_j^2 \right]$$
$$D_3 = \frac{1}{30} y_j^3 \left[5 \left(y_i^2 + y_k^2 + y_l^2 \right) - y_j^2 \right]$$

- $y_i + y_j < y_k + y_l$ and $y_i + y_l < y_j + y_k$:

$$D_1 = \frac{1}{2} (y_i + y_j + y_l - y_k)$$
$$D_2 = -\frac{1}{12} \left[(y_i + y_j) \left\{ (y_i + y_j)^2 - 3 \left(y_k^2 + y_l^2 \right) \right\} + 2 \left(y_k^3 - y_l^3 \right) \right]$$
$$D_3 = \frac{1}{60} \left[y_k^5 - y_i^5 - y_j^5 - y_l^5 + 5 \left(-y_k^3 y_l^2 + y_k^2 y_l^3 - y_k^3 y_i^2 + y_k^2 y_i^3 \right. \right.$$
$$\left. \left. - y_k^3 y_j^2 + y_k^2 y_j^3 + y_l^3 y_i^2 + y_l^2 y_i^3 + y_l^3 y_j^2 + y_l^2 y_j^3 + y_i^3 y_j^2 + y_i^2 y_j^3 \right) \right]$$

Going back to the collision integral, the same trick as before can be applied to the delta

function of energies, which then gives:

$$
I_{\text{coll}} = \frac{S}{64\pi^3 g \widetilde{E_1} y_1 a^5} \int \mathrm{d}y_2 \mathrm{d}y_3 \frac{y_2 y_3}{\widetilde{E_2} \widetilde{E_3}} \mathcal{D}(Y_1, Y_2, Y_3, Y_4) \cdot
$$
$$
\cdot \, \theta \left(\left(\left(s_1 \widetilde{E_1} + s_2 \widetilde{E_2} + s_3 \widetilde{E_3} \right)^2 - a^2 m_4^2 \right) (F[f]) \Big|_{y_4 = y_4^*} \right) , \tag{3.C.27}
$$

with $y_4^* = \sqrt{\left(s_1 \widetilde{E_1} + s_2 \widetilde{E_2} + s_3 \widetilde{E_3} \right)^2 - a^2 m_4^2}$.

3.C.2.2 Case $y_1 = 0$

$$
I_{\text{coll}} = \frac{S}{64\pi^3 gam_1} \frac{1}{a^5} \int \frac{\mathrm{d}y_2 \mathrm{d}y_3 \mathrm{d}y_4 y_2 y_3 y_4}{\widetilde{E_2} \widetilde{E_3} \widetilde{E_4}} F[f] \delta \left(s_1 a m_1 + s_2 \widetilde{E_2} + s_3 \widetilde{E_3} + s_4 \widetilde{E_4} \right) \cdot
$$
$$
\cdot \, \mathcal{B}(Y_1, Y_2, Y_3, Y_4) , \tag{3.C.28}
$$

with

$$
\mathcal{B}(Y_1, Y_2, Y_3, Y_4) = \frac{y_2 y_3 y_4}{64\pi^5} \int \mathrm{d}\Omega_2 \mathrm{d}\Omega_3 \mathrm{d}\Omega_4 \, |\mathcal{M}|^2 \, |\delta^3 \left(s_2 \mathbf{y_2} + s_3 \mathbf{y_3} + s_4 \mathbf{y_4} \right)
$$
$$
= \frac{y_2 y_3 y_4}{64\pi^5} \int \mathrm{d}\lambda \lambda^2 \mathrm{d}\Omega_\lambda \int \mathrm{d}\theta_2 \mathrm{d}\phi_2 \sin\theta_2 e^{is_2 y_2 \lambda \cos\theta_2} \cdot
$$
$$
\cdot \int \mathrm{d}\theta_3 \mathrm{d}\phi_3 \sin\theta_3 e^{is_3 y_3 \lambda \cos\theta_3} \int \mathrm{d}\theta_4 \mathrm{d}\phi_4 \sin\theta_4 e^{is_4 y_4 \lambda \cos\theta_4} \cdot
$$
$$
\cdot \sum_{i \neq j \neq k \neq l} \left[K_1 \left(\widetilde{E_i} \widetilde{E_j} - y_i y_j \cos\theta_i \cos\theta_j \right) \cdot \left(\widetilde{E_k} \widetilde{E_l} - y_k y_l \cos\theta_k \cos\theta_l \right) + \right.
$$
$$
\left. + K_2 \, a^2 m_i m_j \left(\widetilde{E_k} \widetilde{E_l} - y_k y_l \cos\theta_k \cos\theta_l \right) \right] \tag{3.C.29}
$$

Consider the case that $i = 1$ in one of the terms of $|\mathcal{M}|^2$. Then the $\mathcal{B}-$function can be written as:

$$
\mathcal{B}_{i=1}(Y_1, Y_2, Y_3, Y_4) = \frac{y_2 y_3 y_4}{64\pi^5} 4\pi \int \mathrm{d}\lambda \lambda^2 \sum_{j \neq k \neq l} \int \mathrm{d}\theta_j \mathrm{d}\phi_j \sin\theta_j e^{is_j y_j \lambda \cos\theta_j} \cdot
$$
$$
\cdot \int \mathrm{d}\theta_k \mathrm{d}\phi_k \sin\theta_k e^{is_k y_k \lambda \cos\theta_k} \int \mathrm{d}\theta_l \mathrm{d}\phi_l \sin\theta_l e^{is_l y_l \lambda \cos\theta_l} \cdot
$$
$$
\cdot \left[K_1 a m_1 \widetilde{E_j} \cdot \left(\widetilde{E_k} \widetilde{E_l} - y_k y_l \cos\theta_k \cos\theta_l \right) + \right.
$$
$$
\left. + K_2 \, a^2 m_1 m_j \left(\widetilde{E_k} \widetilde{E_l} - y_k y_l \cos\theta_k \cos\theta_l \right) \right] \tag{3.C.30}
$$

$$\mathcal{B}_{i=1}(Y_1, Y_2, Y_3, Y_4) = K_1 a m_1 \sum_{j \neq k \neq l} \left[\widetilde{E_j} \widetilde{E_k} \widetilde{E_l} B_1 \left(y_j, y_k, y_l \right) + \widetilde{E_j} B_2 \left(y_j, y_k, y_l \right) \right] +$$
$$+ K_2 a m_1 \sum_{j \neq k \neq l} a m_j \left[\widetilde{E_k} \widetilde{E_l} B_1 \left(y_j, y_k, y_l \right) + B_2 \left(y_j, y_k, y_l \right) \right],$$

(3.C.31)

with $B_1 \left(y_j, y_k, y_l \right)$ given by Eq. (3.C.10) and

$$B_2 \left(y_j, y_k, y_l \right) = s_k s_l \frac{4 y_k y_l}{\pi} \int \frac{\mathrm{d}\lambda}{\lambda} \sin(y_j \lambda) \left[\cos(y_k \lambda) - \frac{\sin(y_k \lambda)}{y_k \lambda} \right] \left[\cos(y_l \lambda) - \frac{\sin(y_l \lambda)}{y_l \lambda} \right]$$
$$= \begin{cases} \frac{1}{2} \left[y_k^2 + y_l^2 - y_j^2 \right], & y_j + y_k \geq y_l \ \& \ y_j + y_l \geq y_k \ \& \ y_k + y_l \geq y_j \\ 0, & \text{otherwise} \end{cases}$$

(3.C.32)

This procedure can be done for all the other terms in $|\mathcal{M}|^2$. If $j = 1$, the result is the same, but with $i \leftrightarrow j$. Note that if $k = 1$ or $l = 1$, there is no B_2-term in the part with K_2. The collision integral then becomes

$$I_{\text{coll}} = \frac{S}{64 \pi^3 g m_1 a^6} \int \mathrm{d}y_2 \mathrm{d}y_3 \frac{y_2 y_3}{\widetilde{E_2} \widetilde{E_3}} \mathcal{B}(Y_1, Y_2, Y_3, Y_4) \cdot$$
$$\cdot \left. \theta \left(\left(\left(s_1 m_1 + s_2 \widetilde{E_2} + s_3 \widetilde{E_3} \right)^2 - a^2 m_4^2 \right) \left(F[f] \right) \right|_{y_4 = y_4^*},$$

(3.C.33)

with $y_4^* = \sqrt{\left(s_1 a m_1 + s_2 \widetilde{E_2} + s_3 \widetilde{E_3} \right)^2 - a^2 m_4^2}$.

3.D   Low-level code checks

Units correctness   is a very important sanity check of the code. In practice not a single programming language supports this out of the box, but many of them allow to implement more or less complete units algebra with compile time or runtime verification. Typically this restricts the structure of the code or is too verbose and has a big runtime performance overhead.

Fortunately, there is a very economic way to gain a reasonable control of the units. It is based on the idea that units can be treated as some free constants. For example,

$$m = \frac{E}{c^2}$$

(3.D.1)

$$E = 0.511 \, \text{MeV} \qquad c = 3 \cdot 10^8 \, \text{m/s}$$

(3.D.2)

$$m = \frac{0.511 \, \text{MeV}}{9 \cdot 10^{16} \, (\text{m/s})^2} = 5.67 \cdot 10^{-18} \, \frac{\text{MeVs}^2}{\text{m}^2}$$

(3.D.3)

Some units are related to each other:

$$\frac{eV}{J} = 1.602 \cdot 10^{-19} \qquad J = \frac{\text{kg m}^2}{\text{s}^2} \tag{3.D.4}$$

$$m = 5.67 \cdot 10^{-18} \cdot 1.602 \cdot 10^{-13} \frac{\text{J s}^2}{\text{m}^2} \approx 9 \cdot 10^{-31} \, \text{kg} \tag{3.D.5}$$

To employ this idea, one can write a program in the following way:

```python
class UNITS:
    # Base units
    eV = 1.653e-2   # arbitrary constants
    m = 5.27e3
    s = 6.24
    # Derived units
    MeV = 1e6 * eV
    J = 6.242e18 * eV
    kg = J * s**2 / m**2

class CONST:
    c = 3e8 * UNITS.m / UNITS.s

energy = 0.511 * UNITS.MeV
electron_mass = energy / CONST.c**2
print "Electron mass is", electron_mass / UNITS.kg, "kg"
```

Figure 3.19: Sample unit-preserving code in Python

Basically, some arbitrary (possibly random) constants should be assigned to all base units, all derived units are defined in terms of base units and all dimensionful quantities must be created with correct unit multipliers. Then, right before the output, one divides quantity by the desired units constants and obtains the correct answer.

As long as units are used properly, change of the base units constants does not affect the answer (except for corner cases of rounding errors). By running the code twice with different definitions of base units verifies the correctness of the code.

The advantages of this approach are:

- Simplicity and zero overhead: for each occurrence of the unit, there is a simple multiplication operation. There are additional constructs in the code and an obvious recipe for proper usage

- Easy opt-out: all units can be set equal to 1 such that they have no way to influence the computation

- Free unit conversion

On the other hand this approach is runtime-only and does not produce any exceptions during the compilation or runs of the code.

This idea is implemented in the **numericalunits**[4] package that we use in conjunction with simpler code suitable for natural units and similar to the listing 3.19 to enforce units handling in our code.

Verification of the matrix elements. Our computations involve tens of different quantum processes for which we have computed the matrix elements – App. 3.B. It is not uncommon to find mistakes in papers with similar computations. We have checked our matrix elements using Mathematica symbolic algebra scripts that automate the necessary transformations for Fermi theory interactions.

[4]https://pypi.python.org/pypi/numericalunits

Chapter 4

Structure formation

In this Section we briefly describe the observed *Large-Scale Structure of the Universe* (LSS) and its origin. We define the Large-Scale Structure as those inhomogeneities in the matter distribution of the Universe that are roughly larger than the scale of a galaxy. It appears that galaxies form *groups of galaxies*, individual groups form *clusters of galaxies*, while clusters of galaxies can be a part of *superclusters*. These objects are randomly distributed in the Universe and are connected by quasi-1D structures called *filaments* or quasi-2D structures called *walls*. In addition to these overdense regions there are areas with low density of matter called *voids*. All these features together are called *the Cosmic Web*. We will discuss them in Section 4.1.

It appears that these complicated and non-trivial structures can be described using basic equations of matter dynamics (see Section 4.3) starting from very simple initial conditions (see Section 4.3.4). At early times the Universe is almost homogeneous and evolution of the seeds of the current structures can be describe by a simplified linear equations, see Section 4.3. At later stages the evolution of the overdensities cannot be described analytically because of the non-linear nature of basic equations. The widely used way to overcome this problem is to use numerical simulations, so-called N-body simulations, see Section 4.4.1. In Section 4.6 we describe in some details numerical simulations that will be used in our own work, described in Chapter 5. We conclude that the approach reviewed here and applied for the ΛCDM model can describe the observed LSS with a high precision, see Section 4.7.

4.1 Observations of the Large-Scale Structure

Almost one century ago it has been noted by Edwin Hubble that the distribution of the *nebulae* outside our galaxy is fairly uniform on large scales, but starting from the angular scales $\lesssim 10°$ it becomestak clumpy [112–114]. Let us discuss the main blocks of the LSS. The smallest objects of the LSS that we consider here are individual galaxies. They have different sizes from a few to tens of kpc with the total mass from $10^9 M_\odot$ to $10^{13} - 10^{14} M_\odot$
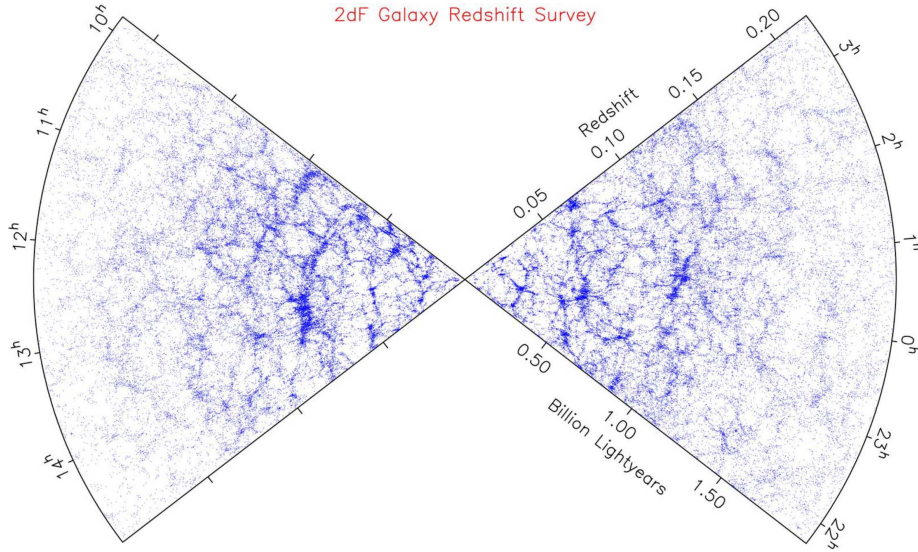
Figure 4.1: The distribution of galaxies found by the 2dF Galaxy Redshift Survey [115].

for the large elliptical galaxies. We expect that there are $\mathcal{O}(10^{11})$ galaxies in the observable Universe [116].

Group of galaxies is a collection of galaxies that consists of $\mathcal{O}(50)$ gravitationally bound members; the collections of galaxies larger than groups in which galaxies are not clustered in some smaller aggregations are called galaxy clusters [117]. The Milky Way galaxy is part of a group of galaxies called the Local Group. A group of galaxies have the typical size of a few Mpc and the typical masses from $10^{13}M_{\odot}$ to $10^{14}M_{\odot}$, while galaxy clusters are from $10^{14}M_{\odot}$ to $10^{15}M_{\odot}$. Typically, galaxy clusters contain $\mathcal{O}(10^3)$ individual galaxies.

The largest known structures in the Universe are superclusters. They are a large group of smaller galaxy clusters or galaxy groups. The large size and low density of superclusters means that they, unlike clusters, expand with the Hubble expansion. The number of superclusters in the observable universe is estimated to be 10 million. The map of the nearby superclusters is shown in Fig. 4.2.

It is known from observations, that small structures such as galaxies, start to form before large structures (see [118] and references therein). The first galaxies are observed at redshifts $z \sim 10$, while the first galaxy clusters start to grow relatively recently, $z \sim 2-3$. This is called the bottom-up structure formation and such a scenario exclude, for example, hot dark matter in which large structures (clusters) are formed before galaxies [33].

Because of the stochastic nature of the LSS the main observables that describe it are different distribution and correlation functions: halo mass function, galaxy-galaxy correlation function, matter power spectrum, etc. The interesting feature that can be found in the galaxy-galaxy correlation function is baryon acoustic oscillations (BAO). The same oscillations that are present in CMB are also imprinted in the density of the baryons and can be observed in the distribution of galaxies at cosmological scales, see Fig. 4.3. The possible
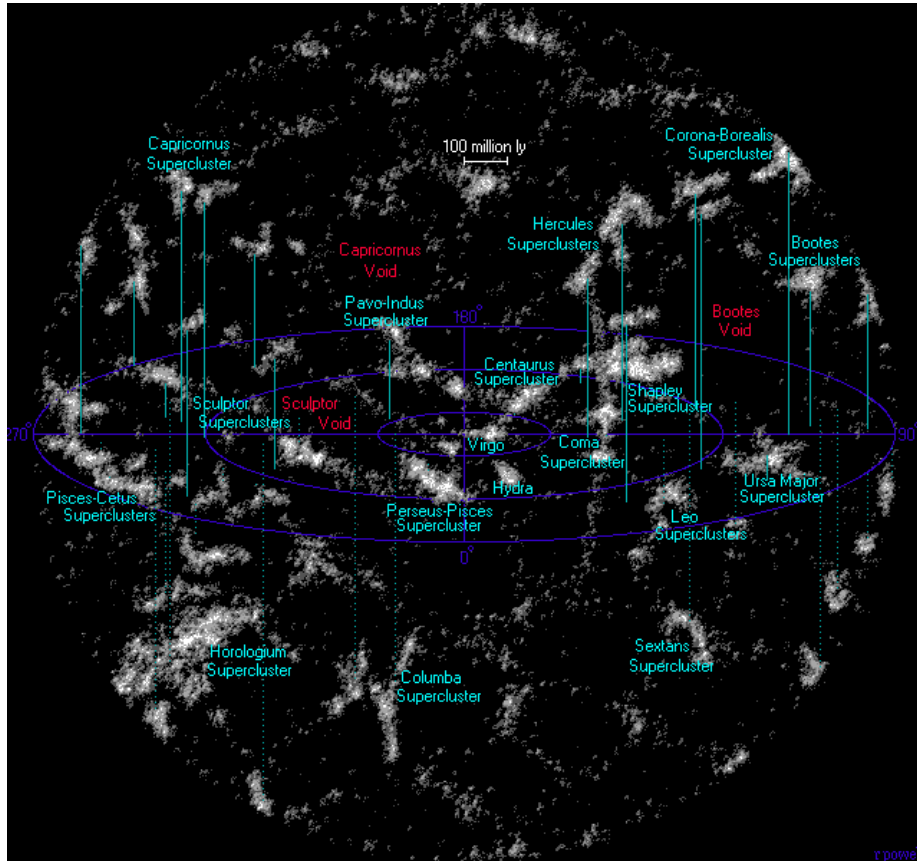
Figure 4.2: A map of superclusters around the Sun.
The figure is from atlasoftheuniverse.com.

application of BAO to cosmology is an independent measurement of the Hubble constant, that can be done using calibration of distances to galaxies using the SN Ia data [120, 121].
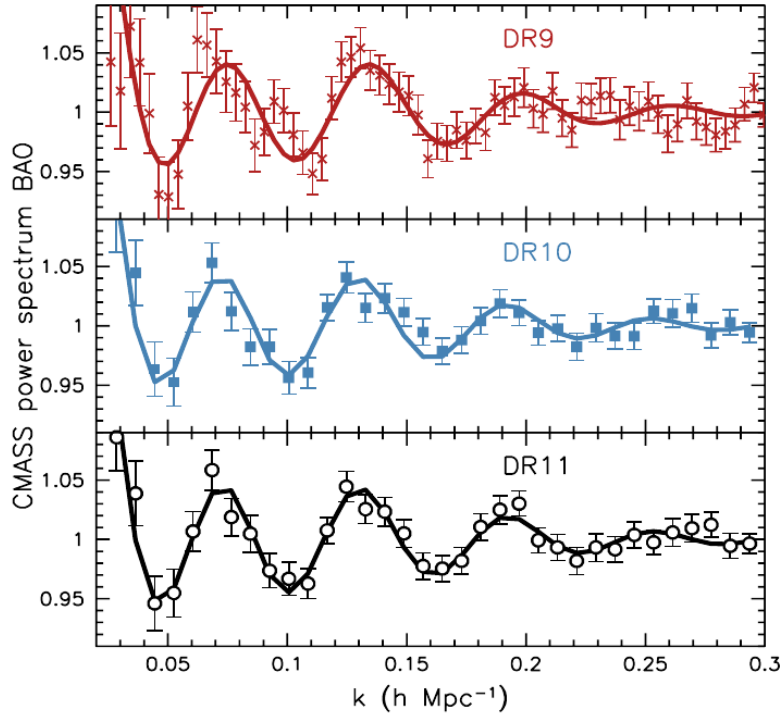
Figure 4.3: The baryon acoustic oscillations in the power spectrum of each of the BOSS data releases, DR9, DR10, and DR11 [119].

4.2 Description of the matter distribution

Matter distribution is defined by its mass density $\rho(\vec{r}, t)$. The inhomogeneities result in appearance of the overdensity $\delta(\vec{x}, t)$ – a relative difference between the local density and the average density of the Universe $\bar{\rho}(t)$:

$$\rho(\vec{x}, t) = \bar{\rho}(t)[1 + \delta(\vec{x}, t)] \quad \Leftrightarrow \quad \delta(\vec{x}, t) = \frac{\rho(\vec{x}, t) - \bar{\rho}(t)}{\bar{\rho}(t)}. \tag{4.2.1}$$

Structure formation was seeded by a quantum process that led to a random distribution of tiny over- and under-densities at very early times. In this case, the particular values of matter density are not instructive – i.e., observing a galaxy at a particular point by itself does not provide useful information. Only the collective properties of the structures, their positions relative to each other, sizes and numbers are important. Hence, we treat $\delta(\vec{x}, t)$ as a random field, the statistical properties of which we want to study. A random field can be completely described by a series of moments – correlations between 1, 2, 3 to infinitely many points defined as:

$$\langle \delta \rangle = \int \delta(\vec{x}) d^3 \vec{x} \tag{4.2.2}$$

$$\langle \delta(\vec{x}_1) \delta(\vec{x}_2) \rangle = \int \delta(\vec{x}_1) \delta(\vec{x}_2) d^3 \vec{x}_1 d^3 \vec{x}_2 \equiv \xi(\vec{x}_1, \vec{x}_2) \tag{4.2.3}$$

etc. Moments of the distribution characterize the generating process leaving out the information about the particular realization of the field. For simple distributions like Gaussian, the first two moments contain all the information while the others are identically zero. Moreover, in general case, the higher-order moments will decrease in amplitude.

When applied to the observed Universe, where the average overdensity is zero by definition $\langle \delta \rangle = 0$ while primordial[1] three-point and higher-order correlations are measured to be consistent with zero. Hence unless we expect some specific non-gaussianity, the two-point correlation function $\xi(\vec{x_1}, \vec{x_2})$ is the main source of information. According to the Cosmological Principle, the Universe we live in is homogeneous and isotropic at large scales. This also means that $\xi(\vec{x_1}, \vec{x_2})$ is invariant under rotations and translations:

$$\xi(\vec{x_1}, \vec{x_2}) \approx \xi(|\vec{x_1} - \vec{x_2}|) \tag{4.2.4}$$

Power spectrum. At small overdensities the gravitational equations can be linearized. This in turn allows for decoupling of scales and simple solutions for these equations in Fourier space. Hence the Fourier image of the correlation function is a particularly useful quantity:

$$\xi(|\vec{x_1} - \vec{x_2}|) = \int \frac{d^3\vec{k_1} d^3\vec{k_2}}{(2\pi)^6} \left\langle \hat{\delta}\left(\vec{k_1}\right) \hat{\delta}\left(-\vec{k_2}\right) \right\rangle e^{-\left(i\vec{k_1}\cdot\vec{x_1} - i\vec{k_2}\cdot\vec{x_2}\right)} \tag{4.2.5}$$

where δ being a real field, means $\delta(-\vec{k}) = \delta^*(\vec{k})$

$$\left\langle \delta\left(\vec{k_1}\right) \delta^*\left(\vec{k_2}\right) \right\rangle = (2\pi)^3 P\left(\vec{k_1}\right) \delta_{\mathrm{D}}\left(\vec{k_1} - \vec{k_2}\right) \tag{4.2.6}$$

$$\xi(|\vec{x_1} - \vec{x_2}|) = \xi(r) = \int \frac{d^3\vec{k}}{(2\pi)^3} P(k) e^{-i\vec{k}\cdot\vec{r}} \tag{4.2.7}$$

4.3 Basic equations governing Large Scale Structure and methods of solution

4.3.1 Equations for a self-gravitating fluid

A simplified description of the density perturbations evolution can be obtained from Newtonian theory if we treat a discrete medium of dark matter particles as a perfect fluid over the expanding background satisfying the Friedmann equations. This approximation is valid as long as we are interested in scales that are both sub-horizon and much larger than the mean free path of the particles.

A perfect fluid is characterized by 3 functions: density distribution $\rho(\vec{r}, t)$, entropy per unit mass $S(\vec{r}, t)$ and velocity field $\vec{v}(\vec{r}, t)$. The density distribution satisfies the continuity

---
[1]Non-gaussianity can be generated during structure formation

equation:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}(\rho \vec{v}) = 0. \tag{4.3.1}$$

The forces acting on a small matter element with a mass M are gravity and pressure p:

$$\vec{F}_{gr} = -M\vec{\nabla}\phi, \tag{4.3.2}$$

$$\vec{F}_{pr} = -\oint p d\vec{\sigma} = -\int_V \vec{\nabla} p dV \approx -\vec{\nabla} p V, \tag{4.3.3}$$

where ϕ is the gravitational potential and $\vec{\sigma}$ is the surface element.

Acceleration of this mass element on the trajectory is

$$a \equiv \frac{d\vec{v}(\vec{r}(t),t)}{dt} = \left(\frac{\partial \vec{v}}{\partial t}\right)_r + \frac{dr^i(t)}{dt}\left(\frac{\partial \vec{v}}{\partial r^i}\right) = \frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v}. \tag{4.3.4}$$

This leads to the equation

$$\frac{\partial \vec{v}}{\partial t} + (\vec{v}\cdot\vec{\nabla})\vec{v} + \frac{\vec{\nabla}p}{\rho} + \vec{\nabla}\phi = 0. \tag{4.3.5}$$

The self-consistent gravitational potential is given by the Poisson equation

$$\Delta\phi = 4\pi G\rho. \tag{4.3.6}$$

If the dissipation is negligible, the entropy is also conserved:

$$\frac{dS(\vec{r}(t),t)}{dt} = \frac{\partial S}{\partial t} + (\vec{v}\cdot\vec{\nabla})S = 0. \tag{4.3.7}$$

The final equation that closes the system is the equation of state that defines the pressure:

$$p = p(\rho, S) \tag{4.3.8}$$

4.3.2 Linearized theory

The system of equations above is non-linear, but it can be considerably simplified in the case when the perturbations are small: $\rho = \rho_0 + \rho_1, \rho_1/\rho_0 \ll 1$. Expanding density, entropy, velocity, pressure and gravitational potential around the averages, we obtain:

$$\frac{\partial \delta\rho}{\partial t} + \rho_0 \vec{\nabla}\delta\vec{v} = 0 \tag{4.3.9}$$

$$\frac{\partial \delta\vec{v}}{\partial t} + \frac{\vec{\nabla}\delta p}{\rho_0} + \vec{\nabla}\delta\phi = 0 \tag{4.3.10}$$

$$\frac{\partial \delta S}{\partial t} = 0 \tag{4.3.11}$$

$$\Delta \delta \phi = 4\pi G \delta \rho \tag{4.3.12}$$

The first-order perturbation of pressure is constrained by the equation of state to be

$$p = p(\rho_0 + \delta\rho, S_0 + \delta S) = p_0 + \delta p \tag{4.3.13}$$

$$\delta p = \left(\frac{\partial p}{\partial \rho}\right)_S \delta\rho + \left(\frac{\partial p}{\partial S}\right)_\rho \delta S \equiv c_s^2 \delta\rho + \sigma \delta S \tag{4.3.14}$$

where c_s is the speed of sound.

The equation for the entropy perturbation state that the entropy is an arbitrary time-independent function. Since the entropy perturbations are not observed, we can put $\delta S = 0$ without loss of generality.

Then the system becomes

$$\frac{\partial \delta \rho}{\partial t} + \rho_0 \vec{\nabla} \vec{\delta v} = 0 \tag{4.3.15}$$

$$\frac{\partial \vec{\delta v}}{\partial t} + \frac{c_s^2}{\rho_0} \vec{\nabla} \delta\rho + \vec{\nabla} \delta\phi = 0 \tag{4.3.16}$$

$$\Delta \delta \phi = 4\pi G \delta \rho \tag{4.3.17}$$

By taking the divergence of the velocity equation and using other equations to express $\vec{\nabla} \vec{\delta v}$ and $\Delta \delta \phi$ in terms of $\delta \rho$, we obtain:

$$\frac{\partial^2 \delta \rho}{\partial t^2} - c_s^2 \Delta \delta\rho - 4\pi G \rho_0 \delta\rho = 0 \tag{4.3.18}$$

Since the coefficients of equations do not depend on the coordinates, a transition to the Fourier space can considerably simplify it:

$$\delta\rho(\vec{r}, t) = \int \delta\rho_k(t) e^{i\vec{k}\cdot\vec{r}} \frac{d^3 k}{(2\pi)^{3/2}} \tag{4.3.19}$$

$$\delta\ddot{\rho}_k + (k^2 c_s^2 - 4\pi G \rho_0)\delta\rho_k = 0 \tag{4.3.20}$$

This equation has two independent solutions

$$\delta\rho_k \propto e^{\pm i \sqrt{k^2 c_s^2 - 4\pi G \rho_0} t} \equiv e^{\pm i\omega(k)t} \tag{4.3.21}$$

The sign under square root governs the behaviour of $\delta\rho_k$ modes. We define the so-called

93

Jeans scale k_J for which $\omega(k_J) = 0$:

$$k_J = \sqrt{\frac{4\pi G \rho_0}{c_s^2}} \qquad (4.3.22)$$

For $k > k_J$ the density perturbation behaves as a sound wave driven by pressure and for $k < k_J$ the perturbation exponentially grows or decays, reflecting the *gravitational instability*:

$$\delta \rho_k \propto e^{\pm |\omega| t} \qquad (4.3.23)$$

When considering large scales $k \to 0$, $|\omega| t \to t/t_{gr} \equiv \sqrt{4\pi G \rho_0} t$. We interpret t_{gr} as the gravitational timescale characteristic to the collapse of the region with density ρ_0.

4.3.3 Perturbations in the expanding Universe

The background density and average velocity fields are subject to Friedmann equations:

$$\dot{\rho}_0 = -3H(t)\rho_0 \qquad (4.3.24)$$

$$\vec{v}_0 = H(t)\vec{r} \qquad (4.3.25)$$

$$\dot{H} + H^2 = -\frac{4\pi G}{3}\rho_0 \qquad (4.3.26)$$

Note that the Hubble flow velocity explicitly depends on the coordinates, which will complicate the form of the equations in the Fourier space. To avoid this, we can switch to the so-called *comoving frame* where coordinates are multiplied by the scale factor $\vec{x} = a(t)\vec{r}$. In this frame the derivatives operators become

$$\vec{\nabla} \to \frac{1}{a}\vec{\nabla} \qquad (4.3.27)$$

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} - \left(\vec{v}_0 \cdot \frac{1}{a}\vec{\nabla}\right) \qquad (4.3.28)$$

The system in the comoving frame takes form

$$\frac{\partial \delta}{\partial t} + \frac{1}{a}\vec{\nabla}\delta\vec{v} = 0, \qquad (4.3.29)$$

$$\frac{\partial \delta\vec{v}}{\partial t} + H\delta\vec{v} + \frac{c_s^2}{a}\vec{\nabla}\delta + \frac{1}{a}\vec{\nabla}\delta\phi = 0, \qquad (4.3.30)$$

$$\Delta\delta\phi = 4\pi G a^2 \rho_0 \delta, \qquad (4.3.31)$$

where $\delta \equiv \left(\frac{\delta\rho}{\rho}\right)$ is called *density contrast*.

Following the logic of the static background, we find the equation for the density

perturbation:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2}\Delta\delta - 4\pi G\rho_0\delta = 0. \tag{4.3.32}$$

Finally, after Fourier transform with the respect to the comoving coordinate $\vec{x}$ the equation becomes

$$\ddot{\delta}_k + 2H\dot{\delta}_k + \left(\frac{c_s^2 k^2}{a^2} - 4\pi G\rho_0\right)\delta_k = 0 \tag{4.3.33}$$

On the scales much larger than the Jeans scale, we can neglect the k-term and find the most general solution

$$\delta = C_1 H \int \frac{dt}{a^2 H^2} + C_2 H \tag{4.3.34}$$

For a flat, matter-dominated universe $a \propto t^{2/3}$ and $H \propto t^{-1}$:

$$\delta = C_1 t^{2/3} + C_2 t^{-1} \tag{4.3.35}$$

Note that in expanding Universe, perturbations grow much slower than in a static Universe – only as a power-law of time instead of exponential.

Multicomponent case. When the Universe is filled with multiple kinds of matter that interact mainly through gravity (e.g. dark matter and baryons), the dynamical equations for each component remain unchanged, while the Hubble rate and the gravitational potential are given by the total energy density.

4.3.4 Initial conditions for inhomogeneities

We have discussed above how the inhomogeneities evolve in the expanding Universe. What is the simplest possible power spectrum?

In the case of $P(k) = 0$ the density field is completely homogeneous. Next to that, what power spectrum would contain the least information?

Let's assume a power-law shape of $P(k) = Ak^n$. First of all, observationally $n \sim 1$. And indeed, for $n = 1$ (known as the Harrison-Zeldovich spectrum) the variation of the gravitational potential is scale-independent. Indeed, if we write the Poisson equation in proper coordinates for a perturbation δ giving rise to a perturbation of the potential $\delta\Phi$, we have

$$\nabla^2\delta\Phi = 4\pi G\bar{\rho}\delta \tag{4.3.36}$$

in Fourier space

$$\delta\hat{\Phi}_k = -4\pi G\bar{\rho}\frac{\hat{\delta}_k}{k^2} \tag{4.3.37}$$

meaning that

$$\Delta^2_{\delta\hat{\Phi}_k} \propto k^3 \left\langle \left|\delta\hat{\Phi}_k\right|^2 \right\rangle \propto k^3 \left\langle |\hat{\delta}(k)|^2 \right\rangle \frac{1}{k^4} \propto k^{n-1} \tag{4.3.38}$$

4.4 From linear theory to N-body simulations

At some point, the growing perturbations achieve level $\delta \sim 1$ breaking the defining assumption of the linear theory. These perturbations have to be described by non-linear gravitational equations for which there are no analytical solutions.

However, it is possible to approximate the solutions by replacing individual DM particles with large super-particles. In this case, we can follow the evolution of each particle using computer simulations. This approach is called N-body simulations.

4.4.1 N-body simulations

Dark matter can be thought of as a collision-less self-gravitating fluid. If we associate a distribution function to it, the equations of motion can be expressed in a form of *collisionless Boltzmann equation*:

$$\frac{\partial f}{\partial t} + \dot{\vec{x}}\frac{\partial f}{\partial \vec{x}} - \vec{\nabla}\phi\frac{\partial f}{\partial \dot{\vec{x}}} = 0 \tag{4.4.1}$$

where $f(\vec{x}, \dot{\vec{x}}, t)$ is defined as number of particles at phase-space point $(\vec{x}, \dot{\vec{x}})$

$$\rho(\vec{x}, t) = \frac{m}{a^3(t)} \int d^3\dot{\vec{x}} f(\vec{x}, \dot{\vec{x}}, t) \tag{4.4.2}$$

We can approximate the solution of the Boltzmann equation by replacing the continuous phase space with a collection of N *pseudoparticles* with mass $M = \rho_0 V/N$. By solving in essence Newtonian equations of motion for the pseudo particles we can recover the resulting distribution of matter up to the resolution scale given by the number of particles.

$$\frac{d\vec{p}_i}{da} = -\frac{\vec{\nabla}\phi}{\dot{a}} \tag{4.4.3}$$

$$\frac{d\vec{x}_i}{da} = \frac{\vec{p}_i}{\dot{a}a^2} \tag{4.4.4}$$

4.4.2 How to put initial conditions?

The starting configuration of DM and gas particles is typically computed at linear stages of the perturbation revolution ($z \sim 100$). The smallest scales to resolve control the precise limit of the starting redshift – since the smallest scales reach the non-linear regime first.

We extract the statistics of the primordial perturbations from the temperature of CMB. The spatial correlations are measured to be Gaussian which means that all information is contained inside the mean and the two-point correlation function (or power spectrum).

Two define the initial conditions for second-order Newton equations, we need to specify the distribution and velocity fields.

Particle velocities depend on the initial distribution function of the matter which is taken to be thermal for baryons but might be different for DM (depending on the model). On larger scales, the gravitational potential causes bulk motion of the matter.

The power spectrum provides the amplitudes for the plane wave decomposition, but not the phases which are considered random. Hence, a single power spectrum corresponds to an infinite number of the density field realizations.

Translation between the density field and the discretized distribution of particles is done using the *Zeldovich approximation* [122].

We represent the proper positions of the particles as a combination of the Hubble flow and some *displacement field*:

$$\vec{x}(t) = a(t)\vec{q} + b(t)\vec{f}(\vec{q}) \tag{4.4.5}$$

The particles are assumed to move along the displacement field $\vec{f}(\vec{q})$ with velocity increasing according to the "growth" function $b(t)$. The Lagrangian coordinates $\vec{q}$ are equal to the initial positions of the particles and have a property that $\rho(\vec{q}) = \text{const}$. The proper density is computed using $\rho(\vec{x})d\vec{x} = \rho(\vec{q})d\vec{q}$:

$$\rho(\vec{x}) = \rho_q \left| \frac{d\vec{q}}{d\vec{x}} \right| = \rho_q \left| \frac{d\vec{x}}{d\vec{q}} \right|^{-1} = \rho_q \left| a(t)\delta_{ij} + b(t)\frac{df_i}{dq_j} \right|^{-1} = \tag{4.4.6}$$

$$= \rho_0 \left| \delta_{ij} + \frac{b(t)}{a(t)}\frac{df_i}{dq_j} \right|^{-1} \approx \rho_0 \left[1 - \frac{b(t)}{a(t)}\vec{\nabla}\vec{f} \right] \tag{4.4.7}$$

$$\delta(\vec{x}) \equiv \frac{\rho(\vec{x}) - \rho_0}{\rho_0} = -\frac{b(t)}{a(t)}\vec{\nabla}\vec{f} \tag{4.4.8}$$

In linear theory the perturbations are non-rotational and displacement field can be written as a gradient of the scalar field:

$$\delta(\vec{x}) = -\frac{b(t)}{a(t)}\Delta\Phi \tag{4.4.9}$$

Notice that this equation becomes the Poisson equation for the density perturbation (4.3.31) if we identify $b(t)\Phi \equiv -\frac{1}{4\pi G a \rho_0}\delta\phi$ at $t = 0$.

Finally, we obtain

$$\vec{x}(t) = a\vec{q} + \vec{\nabla}(b(t)\Phi) = a\vec{q} - \frac{\vec{\nabla}\delta\phi(\vec{q}, 0)}{4\pi G a \rho_0} \tag{4.4.10}$$

This equation provides the displacements of the perturbations at time t using the initial gravitational potential. Velocity field is given by a time derivative $\dot{\vec{x}}(t)$.

So, to define the initial conditions for the simulation, we can define a grid of pseudoparticle positions $\vec{q}_i$ and correct them according to the Zeldovich approximation (4.4.10).

However, it has been shown that Zeldovich approximation might be insufficient to correctly set the initial conditions on the smallest scales. A more precise approach is called the *Second-order Lagrangian Perturbation Theory* and is implemented in a code 2LPT [123].

Glass initial conditions. The choice of a regular grid for initial positions $\vec{q}_i$ can introduce artificial patterns in the simulation that survive until late times (Fig. 4.4).

The way to mitigate this is to choose an explicitly irregular grid. This can be done by applying an N-body code to a regular grid with the direction of the gravity reversed. Particles will attempt to separate as far as possible from each other while still keeping the density more or less unaffected.
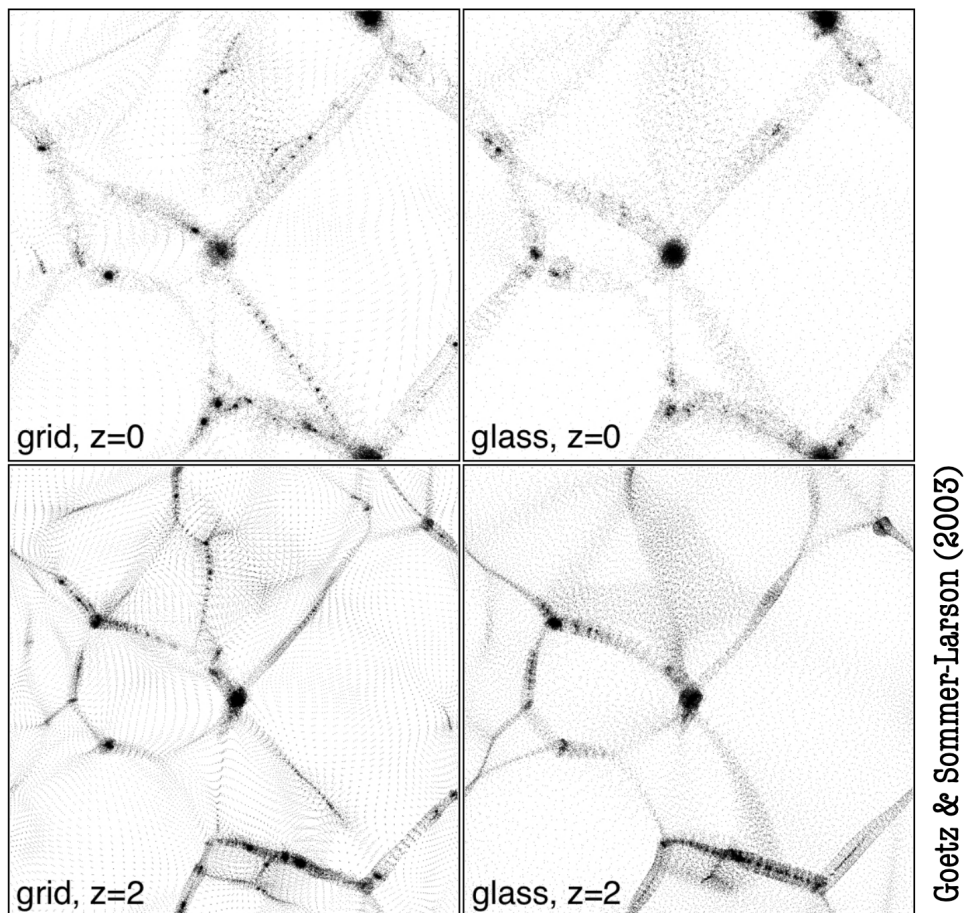
Figure 4.4: Initial conditions built from the evenly spaced *grid* (*left panels*) contain small-scale regularities that survive until late time. *Glass* initial conditions (*right panels*) have the same properties on large scales, but also are disordered on the small scales.

4.5 Numerical methods of N-body simulations

Direct summation approach. The force experienced by each pseudoparticle is the sum of the regular $1/r^2$ Newtonian forces. This poses a difficulty for computation when the particles are close to each other. So to avoid divergences, the gravitational interaction is regularized using *Plummer softening*:

$$\vec{F}_{ij} = GM^2 \frac{\vec{x}_i - \vec{x}_j}{(|\vec{x}_i - \vec{x}_j|^2 + \epsilon^2)^{3/2}} \tag{4.5.1}$$

This corresponds to smearing the particles such that the majority of their mass is contained inside a sphere of radius $\sim 3\epsilon$. This allows the particles to freely pass through each other ($\vec{F}_{ij}(\vec{x}_i = \vec{x}_j) = 0$) without causing numerical problems.

The total force for each particle can be obtained as a *direct sum* over all other particles and the system of equations can be solved by several numerical schemes, including Runge-Kutta, Predictor-Corrector or Leap-Frog (we defer the discussion of them to [45].

Unfortunately, despite the simplicity of this method, it suffers from the $O(N^2)$ complexity which limits its applicability to relatively small particle systems. Even though specialized hardware has been developed to efficiently compute the forces (see GRAPE computer [124], modern simulations do not use this method.

Tree methods. Using the same $1/r^2$ characteristic of the gravitational force, it is possible to make meaningful approximations that reduce the complexity of computation from $O(N^2)$ to $O(N \log N)$.

The main idea lies in separating the nearest neighbors from the particles located far away. Small errors in positions of the latter do not significantly influence the resulting force – hence, we can group them into *clusters* reducing the number of pairs to consider.

One the simplest ways to achieve this is to form a *binary tree* out of particles by grouping together the closest pairs, then pairs of pairs and so on. Each node of the tree is annotated with center-of-mass location, total mass, and its "size". Then, to compute the force one iterates over the tree starting from its root (the largest cluster) considering whether each given node is "far enough" to be considered a point particle. For a node of size L at a distance D this can be expressed as

$$\begin{cases} \text{if } L \ll D & \text{treat node as a point particle} \\ \text{else} & \text{repeat for its descendants} \end{cases} \tag{4.5.2}$$

The precision is controlled by some tolerance parameter equal to the ratio L/D that is "far enough". When tolerance is set to 0, the algorithm becomes the direct summation.

Higher accuracy can be achieved if a low-order multipole expansion of each node is computed in addition to the monopole contribution.

This approach significantly reduces the time of force computation, but at the same time at each simulation step, the tree needs to be rebuilt.

An evolution of this approach is represented by the Barnes-Hut algorithm [125] where the tree is built by successively splitting the simulation volume into octants until each of them contains 1 or 0 particles (Fig. 4.5). This structure allows to efficiently look for nearest neighbors and to only partially rebuild the tree when particles leave their cells.

Spatial Domain   **Quad-Tree Representation**
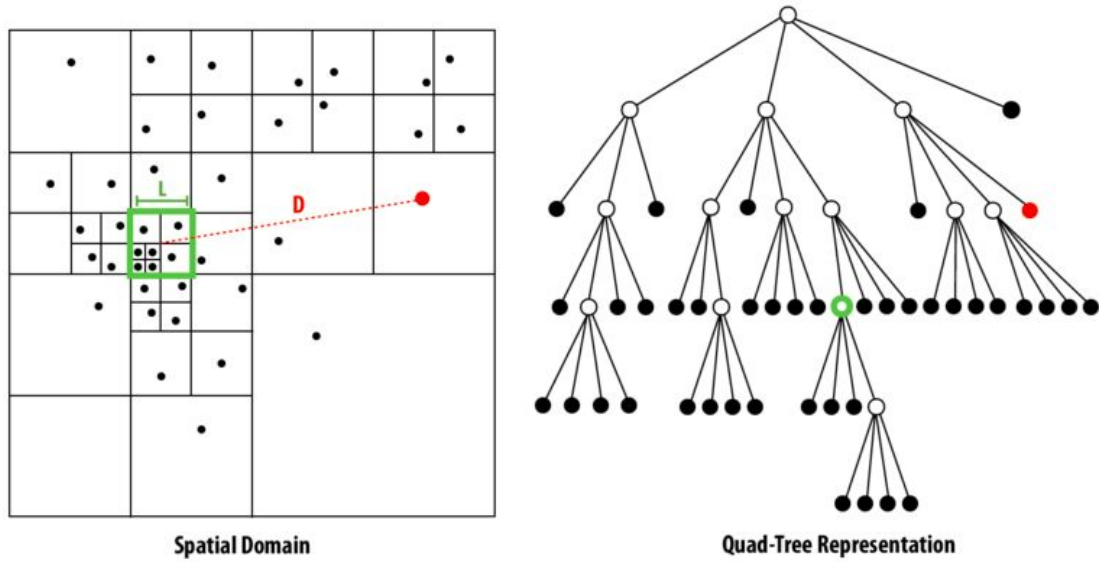
Figure 4.5: In Barnes-Hut algorithm [125], the particles are organized into hierarchical structure (Quad-Tree) that allows for efficient nearest-neighbour search and multipole expansion of the gravitational force.

Particle Mesh methods. An alternative approach to force computation uses the Fast Fourier transform to efficiently solve the Poisson equation for the gravitational potential and compute the resulting force from it.

The first step is to create a regular grid to evaluate the potential. The matter density is interpolated to this grid using some *kernel interpolation function W*:

$$\rho(\vec{g}_i) = \sum_j m_j W(\vec{x}_j - \vec{g}_i) \tag{4.5.3}$$

.

Given a density estimate, the Poisson equation is solved in the Fourier space:

$$\phi_k = -4\pi G a^2 \frac{\rho_k}{k^2} \tag{4.5.4}$$

Then the potential is transformed back to the real space and the force is computed using a discretized derivative operator. Finally, the force is interpolated back to the particles using the same kernel.

This method is very efficient because the Fast Fourier Transform for grid size N_g has the complexity of $O(N_g \log N_g)$. At the same time to achieve a significant speed-up compared to the tree methods, it is desired to have $N_g \ll N$ which significantly limits the force resolution.

P^3M **methods.** A combination of particle mesh and direct summation (or tree) methods can be used to achieve the best of both worlds. The limited force resolution of the particle mesh is compensated by the direct summation on short distances.

4.5.1 Interacting fluid

In the case of an interacting fluid, the Boltzmann equation governing the particle distribution gains the *collision term*

$$\frac{\partial f}{\partial t} + \dot{\vec{x}}\frac{\partial f}{\partial \vec{x}} - \vec{\nabla}\phi\frac{\partial f}{\partial \dot{\vec{x}}} = \left[\frac{df}{dt}\right]_c \tag{4.5.5}$$

The collision term encompasses all the information about the thermodynamical properties and phenomena occurring in the gas: pressure, viscosity, molecular and atomic processes. In the whole generality, one would need to introduce the distribution functions for all states possible in the system like atoms, ions, free electrons, etc. However, thanks to the separation of scales, we can assume that the quantum phenomena are local and the system can be adequately described by a distribution function of *matter* supplemented by position-dependent fields describing the "subgrid" physics (e.g., ionization fraction $\xi(\vec{x})$). One of the notable exceptions to this is the high energy radiation which can have a large mean-free path.

The equations of motion of gas can be expressed as a system containing the continuity, momentum and energy equations:

$$\frac{\partial \rho}{\partial t} + 3H\rho + \frac{1}{a}\vec{\nabla}\rho\vec{v} = -\Gamma_*\rho \tag{4.5.6}$$

$$\frac{\partial \vec{v}}{\partial t} + \frac{1}{a}(\vec{v}\cdot\vec{\nabla})\vec{v} + H\vec{v} = -\frac{1}{a\rho}\vec{\nabla}P - \frac{1}{a}\vec{\nabla}\phi \tag{4.5.7}$$

$$\frac{\partial E}{\partial t} + 2HE + \frac{1}{a}\vec{\nabla}((E+P)\vec{v}) = \frac{Q - \Lambda(\rho, T)}{\rho} \tag{4.5.8}$$

where Γ_* is the gas destruction rate due to star formation, E is the energy of gas per unit comoving volume, Q and Λ account for heating and cooling of the gas. As before, these equations need to be supplemented by the Poisson equation for the gravitational potential and the gas equation of state.

The approaches to solving this system are divided into two groups: Eulerian and Lagrangian methods. The difference between them lies in the way the fields are discretized.

In Eulerian methods, the equations are solved on a static grid and the interactions such as the flow of matter from point to point are computed. While Lagrangian methods are more similar to particle systems that are mobile themselves.

Eulerian methods. This approach is a standard of computational fluid dynamics. These methods are good at conserving integrals of motion and handling hydrodynamical shocks. However, equations are solved on a grid which limits the resolution. In the context of cosmological simulations, this is a problem because of the huge dynamical range: empty voids versus galaxies, stars, etc.

A modern approach to avoid the computational complexity at high resolution is called *Adaptive Mesh Refinement* (AMR) [126].

Lagrangian methods (Smoothed Particle Hydrodynamics). Since the highest resolution is required in regions of the matter collapse, assigning the fluid variables to the particles moving with the fluid is a natural solution. However, this requires the introduction of *artificial viscosity* to handle shocks.

Most methods are based on the *Smoothed Particle Hydrodynamics* by Gingold and Monaghan [127]. The most important idea of the SPH is the interpolation scheme that connects the "physical" fields with their values at a set of particles.

First, let's introduce an estimate of the physical field $A(\vec{x}, t)$ smoothed over scale h:

$$A_S(\vec{x}, t) = \int d\vec{y} A(\vec{y}, t) W(\vec{x} - \vec{y}, h) \tag{4.5.9}$$

$W(\vec{x}, h)$ is the kernel interpolation function satisfying conditions

$$\int W(\vec{x} - \vec{y}, h) d\vec{y} = 1 \tag{4.5.10}$$

$$\lim_{h \to 0} W(\vec{x} - \vec{y}, h) = \delta(\vec{x} - \vec{y}) \tag{4.5.11}$$

The definition of the smoothed field has the advantage that the derivatives of the smoothed quantities are computed through the derivatives of the kernel function

$$\vec{\nabla} A_S(\vec{x}, t) = \int d\vec{y} A(\vec{y}, t) \vec{\nabla}_y W(\vec{x} - \vec{y}, h) \tag{4.5.12}$$

If we discretize the integral by splitting the volume into particles, we can compute the physical quantities using the expression

$$A(\vec{y}) = \sum_j m_j \frac{A_j}{\rho_j} W(\vec{y} - \vec{x}_j, h) \tag{4.5.13}$$

$$\rho(\vec{y}) = \sum_j m_j W(\vec{y} - \vec{x}_j, h) \tag{4.5.14}$$

Example derivation of gas equations can be found e.g. in a SPH review by Monaghan [128].

$$\frac{d\vec{v}_i}{dt} = -\sum_j m_j \left(\frac{P_i}{\rho_i^2} + \frac{P_j}{\rho_j^2} + \Pi_{ij} \right) \vec{\nabla}_i W(\vec{x}_i - \vec{x}_j, h) \vec{\nabla}\phi \tag{4.5.15}$$

$$\frac{d\epsilon_i}{dt} = \sum_j \left(\frac{P_i}{\rho_i^2} + \frac{\Pi_{ij}}{2} \right) (\vec{v}_i - \vec{v}_j) \cdot \vec{\nabla}_i W(\vec{x}_i - \vec{x}_j, h) + \frac{Q_i - \Lambda_i}{\rho_i} \tag{4.5.16}$$

where Π_{ij} represents the mentioned artificial viscosity term.

In this approach, the shocks are spread over several smoothing lengths $\sim 3h$. Because of this, most modern implementations use a spatially-variable smoothing scale $h \propto \rho^{-1/3}$.

4.5.2 Subgrid physics

Equations describing the evolution of baryonic fluid 4.5.6 contain external functions like gas destruction, cooling, and heating rates.

Cooling depends on the thermodynamical quantities as well as chemical composition. It happens through the emission of radiation and redistribution of the kinetic energy between particles during collisions, ionizations, and recombinations. A typical description of the cooling involves a hot cloud of gas that radiates away from the thermal energy. But this cloud is surrounded by medium so ultimately there is no "thermal sink" and energy is just redistributed between the components of the plasma.

Plasma of primordial composition at early times consists from free electrons e^-, chemical elements like H and He and their ionized states H^+, He^+, He^{++}. From primordial nucleosynthesis (Sec. 3) we know that Hydrogen and Helium constitute more than 99% of the nuclei in the Universe, so these two species are enough to satisfactorily describe gas at large scales. Other chemical elements could play a role in special conditions like in galaxies or stars. The cooling rates for this very important case is shown in Fig. 4.6.

Note that the cooling function is proportional to the squared density of the gas (since two-body processes are dominating the plasma). This means that cooling is much more effective in the denser regions. An overdensity also represents a minimum of gravitational potential that attracts particles. In the case of collisionless DM, particles are prevented from falling directly into the center of the overdensity because of the angular momentum conserved for all particles individually. On the other hand, radiating particles like baryons
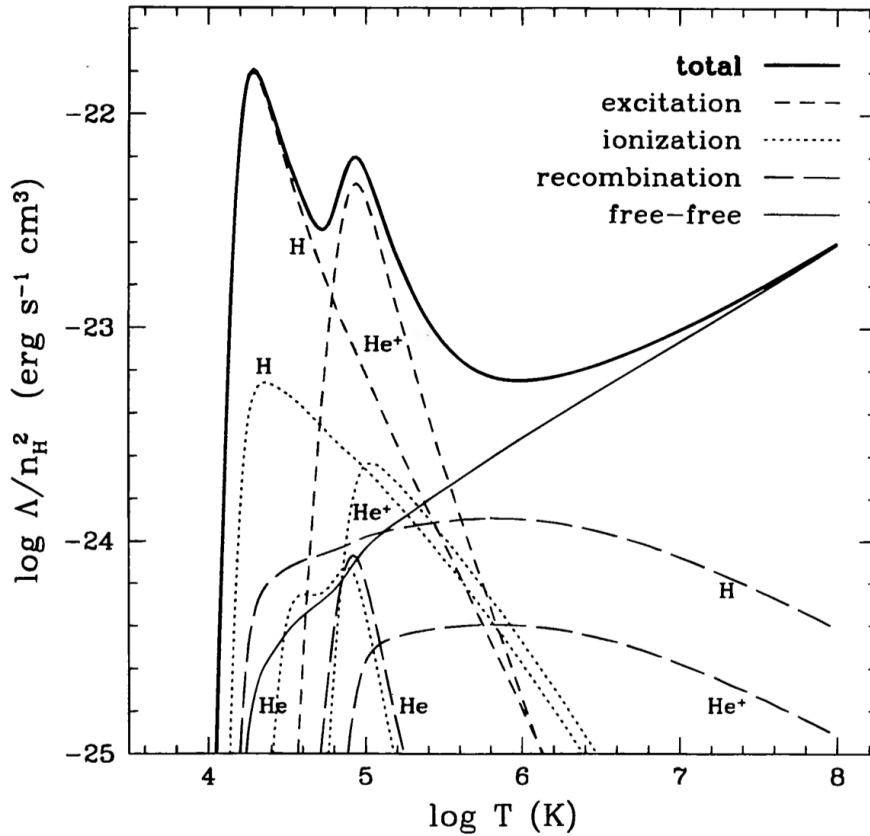
Figure 4.6: Cooling rates as a function of temperature for a primordial composition gas in collisional equilibrium [129]. The heavy solid line shows the total cooling rate. The cooling is dominated by collisional excitation (short-dashed lines) at low temperatures and by free-free emission (thin solid line) at high temperatures. Long-dashed lines and dotted lines show the contributions of recombination and collisional ionization, respectively.

will on average move closer to the center. This, in turn, will increase the density and cooling further.

This mechanism leads to the generation of compact, dense baryonic structures that eventually become stars and galaxies. DM plays an important role in providing the deep potential wells for baryons to fall in.

Star formation. One of the factors playing against the cooling and contraction of baryonic overdensities is the heating due to the increasing pressure. Particles falling in the potential well gain some momentum that compensates for the energy radiated away. At some point, the density and temperature in the center of the cloud might overcome the Coulomb barrier of the nuclei and spark fusion. This would start a chain reaction since fusion is a potent source of energy radiation that will further heat the surrounding plasma.

As a net result, a star is born that on cosmological scales can be treated as a point-like radiation source. The precise details of this process have not been modeled and in any

case, would require tremendous resolution in the context of the cosmological simulation. Typically a single pseudoparticle in simulation is orders of magnitude heavier than stars. Because of this, star formation is treated as a local process happening "inside" a pseudo particle that converts baryons into star particles. Naturally, this requires additional models to specify the conditions and rate of star formation.

For galaxy formation, the feedback processes from the stellar winds and explosions are important and require dedicated modeling. However, they are relevant only to relatively small scales.

Photo-heating and -ionization. Nuclear reactions inside the stars serve as radiation sources capable of heat and ionize the surrounding gas. In general, the treatment of this phenomenon requires solving the radiative transfer equations. However, given a sufficiently homogeneous distribution of sources that is dense in relation to the mean free path of the radiation, we could treat this radiation as a homogeneous field covering all space – the cosmic *Ultraviolet Background* (UVB).

The interaction of the UVB with the gas depends on the cross-section for interaction of a particular atom specie $\sigma_i(\nu)$ and can be encapsulated in photoionization $\Gamma_{\gamma i}$ and photoheating ϵ_i rates:

$$\Gamma_{\gamma i} \equiv \int_{\nu_i}^{\infty} \frac{4\pi J(\nu)}{h\nu} \sigma_i(\nu) d\nu \qquad\qquad [\text{s}^{-1}] \qquad\qquad (4.5.17)$$

$$\epsilon_i \equiv \int_{\nu_i}^{\infty} \frac{4\pi J(\nu)}{h\nu} \sigma_i(\nu) \left(h\nu - \nu_i\right) d\nu \qquad\qquad [\text{ergs s}^{-1}] \qquad\qquad (4.5.18)$$

where $J(\nu)$ is the intensity of the UVB (in units of $\text{ergs}^{-1}\text{cm}^{-2}\text{sr}^{-1}\text{Hz}^{-1}$). These rates describe how frequently atoms are ionized by radiation and what excess energy is introduced into plasma by photoelectrons.

Stars might not be solely responsible for the formation of the UVB. Quasars have been long recognized as important sources of ionizing radiation. Other sources like decaying/annihilating DM or evaporating black holes might also participate.

The intensity of the UVB is another quantity that has to be determined from external models. When supplied, it yields the total heating rate of the gas:

$$Q = \sum n_i \epsilon_i \qquad\qquad (4.5.19)$$

4.6   Simulation codes

Since typical cosmological simulations include both dark matter and gas evolved at the same time, usually the codes layer some kind of an N-body method for DM with either SPH or AMR handling the baryons. Depending on the desired processes, additional numerical

ingredients are added, ranging from subgrid physics for star and black hole formation to radiative transfer schemes for the propagation of radiation.

4.6.1 Code used in this work

In our simulations, we use a modified version of the code P-Gadget3 by Volker Springel et al. [45]. This code uses a TreePM method for collisionless dynamics that complements the accuracy of the tree method on small scales with the efficiency of the particle mesh on large scales. SPH in an explicitly entropy-conserving formulation is used for gas dynamics.

Since we are interested in the physics of the Intergalactic Medium which is either only slightly overdense or underdense, it is not necessary to model precisely the dense regions. This allows us to cut on computing time by using an extremely simple placeholder of the star formation model.

Star formation heuristic. Stars occur only in dense regions and besides their role in the formation of UVB does not participate in structure formation. We use a simple criterion of transforming gas particles into stars: if the temperature of the particle is $< 10^5 K$ while its overdensity is larger than some *critical overdensity* (which we take equal to 10^3).

4.7 Comparison between observations and simulations of Large Scale Structure

As the result of the approach presented in the previous sections it is possible to produce simulations of the LSS at different scales. For example, Fig. 4.7 shows the comparison between the Millennium simulations [130] (red) with the 2dFGRS galaxy survey [115] (blue). We see the remarkable similarity between them, without hint it is not possible to distinguish simulations from observations.

To be more quantitative, let us look at the galaxy-galaxy correlations function produced in the modern IllustrisTNG simulation [132], see Fig. 4.8. Comparing it to the data of the Sloan Digital Sky Survey we see the excellent agreement between them. These, and other comparisons between the simulations and observations convince us that ΛCDM cosmology perfectly fits the data on the LSS.
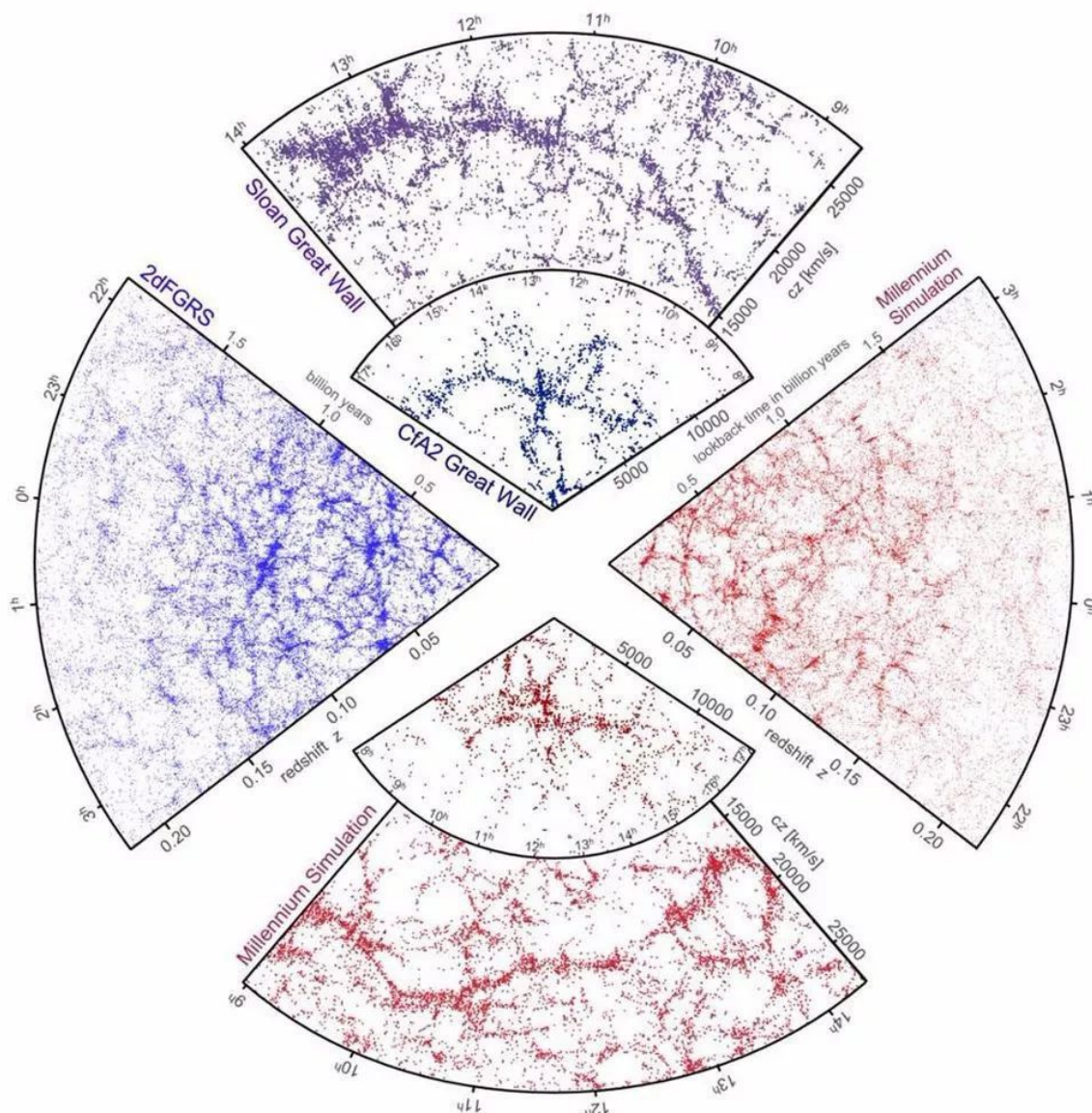
Figure 4.7: Comparison of the galaxy surveys with the simulations. The small slice at the top shows the CfA Great Wall, with the Coma cluster at the centre. Drawn to the same scale is a small section of the SDSS, in which an even larger Sloan Great Wall has been identified. This is one of the largest observed structures in the Universe, containing over 10,000 galaxies and stretching over more than 1.37 billion light years. The cone on the left shows one-half of the 2dFGRS, which determined distances to more than 220,000 galaxies in the southern sky out to a depth of 2 billion light years. The SDSS has a similar depth but a larger solid angle and currently includes over 650,000 observed redshifts in the northern sky. At the bottom and on the right, mock galaxy surveys constructed using semi-analytic techniques to simulate the formation and evolution of galaxies within the evolving dark matter distribution of the Millennium simulation are shown, selected with matching survey geometries and magnitude limits. *Credit: Springel et al.* [131]
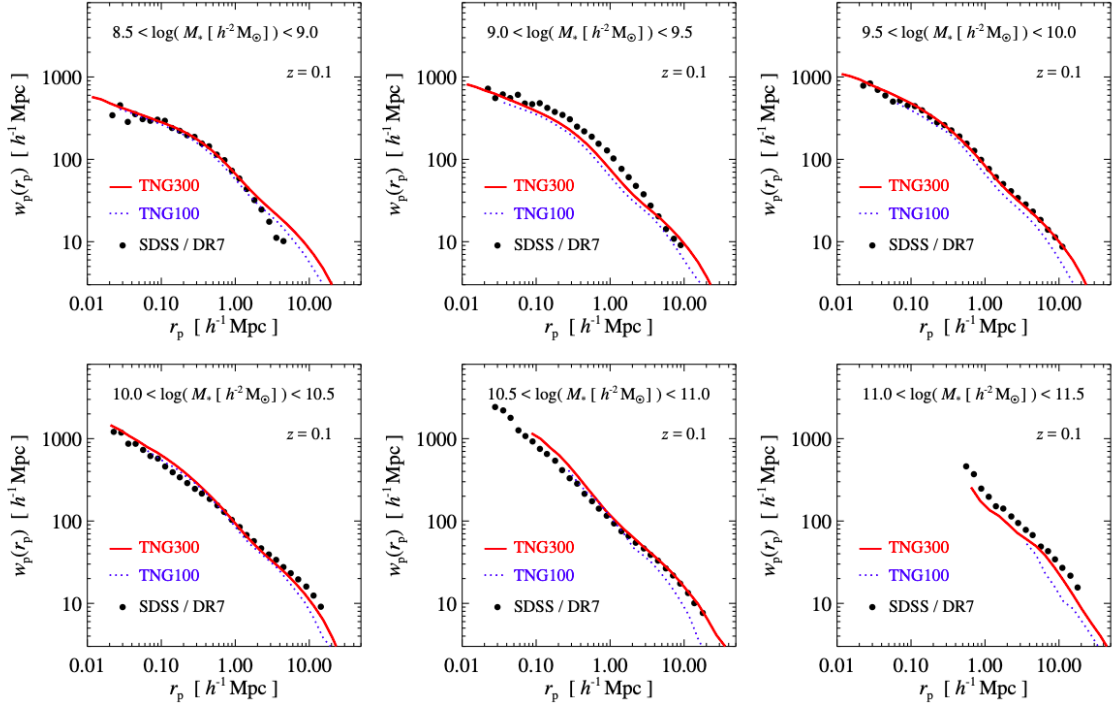
Figure 4.8: Comparison of the projected two-point galaxy correlation functions of the TNG300 simulation (solid) and the TNG100 simulation (dotted) at $z = 0.1$ with the Sloan Digital Sky Survey, in six different stellar mass ranges. Taken from [132].

Chapter 5

Lyman-α forest and
Warm Dark Matter[1]

5.1 Introduction and summary

As we saw in the previous section, ΛCDM cosmology provides an excellent description of the observed Large Scale Structure in the Universe. It also describes very well statistical properties of the CMB, relating the temperature fluctuations detected in the CMB to the density fluctuations in the distribution of galaxies (see *e.g.* [100] for a recent description). Non-baryonic "dark matter" is a crucial ingredient of the Standard Cosmological model, reconciling the low amplitude of the temperature fluctuations in the CMB with the high amplitude of fluctuations detected in the total matter density inferred from the clustering of galaxies (see Chapters 1 and 4).

However, the detailed properties of the DM particle have little impact on the success of the ΛCDM model on large scales. As we have discussed in Chapter 1, depending on the production mechanism of Dark Matter particles in the Early Universe, initial (i.e. primordial) DM velocities may strongly suppress the amplitude of matter fluctuations on scales below a characteristic *free streaming horizon*, λ_{FS} that is more often called 'free-streaming length'(see *e.g.* the discussion by [133]). This quantity is defined in the usual way (see also Section 1.1.1 for details):

$$\lambda_{\mathrm{FS}}(t) \equiv a(t) \int_{t_i}^{t} d\tau \, \frac{v(\tau)}{a(\tau)}. \tag{5.1.1}$$

In CDM model the free streaming length in negligible, but for *Warm Dark Matter* λ_{FS} can be of the order of a co-moving megaparsec. Particle physics candidates for Warm and Cold DM can be very different from the point of view of the particle physics (masses

---

[1]Results of this chapter are presented in papers [1, 2]. The main contributions of Andrii Magalich are a new set of cosmological simulations exploring reionization scenarios in CDM and WDM cosmologies as well as construction of physical Lyman-α forest observables like mock quasar spectra and Flux Power Spectra (in collaboration with Antonella Garzilli).

of particles, generation mechanisms, quantum properties, etc.). But for Cosmology the difference between them could be seen only on the smallest scales – WDM shares all the success of CDM at large scales!

> Therefore, it is a very important fundamental problem for physics in general to identify ways to distinguish between Cold and Warm Dark Matter observationally. In this section, we discuss one promising approach to do this – the *Lyman-α forest* method – and its current results.

The effects of free-streaming on structure formation may be detectable if λ_{FS} is larger than tens of comoving kpc. At smaller scales the effects are completely obscured by baryons and from practical point of view such WDM models are indistinguishable from CDM. To constrain WDM models there are, essentially, 3 main approaches:

1. Determine the number of collapsed structures as a function of their masses and redshifts. A large value of λ_{FS} will also dramatically reduce the abundance of low-mass DM halos (see *e.g.* [134, 135]) and consequently also of the low-mass ('dwarf') galaxies they host. The abundance of Milky Way satellites, for example, therefore provides interesting limits on λ_{FS} [136, 137]. However the impact of relatively poorly understood baryonic physics may ultimately limit the constraining power of both methods. Methods that are largely free from such uncertainties are therefore more promising; these include gravitational lensing by low-mass halos [138], and the creation of gaps in stellar streams by the tidal effects of a passing dark matter subhalo [139].

2. Determine the distribution of matter within individual DM-dominated objects (cores in galactic halos). Particle free-streaming introduces a maximum phase-space density of fermionic DM which could potentially cause Dark Matter halos to have a central density 'core' [140–142]. The smallness of such a core [142], and the potential for baryonic processes associated with star formation and gas cooling to affect the central density profile (see *e.g.* [143–145]), render this route to determining λ_{FS} challenging [146]. We shall not discuss this question in the current review and refer an interested reader to [147].

3. Measurement of the matter power spectrum at small scales. In the work we concentrate on studying statistical properties of DM distribution at small scales. As it was mentioned in Chapter 4, apart from halos of galaxies and clusters, DM forms also filamentary (one dimensional) structures that are organised into the so-called *Cosmic Web* (see Fig. 1.5). In Section 5.2 we briefly describe how the properties of the Cosmic Web differ in the cases of Cold, Warm and Hot Dark Matter and conclude that we would like to observe the effect of WDM on the Matter Power Spectrum (defined in Section 4.2)

The Cosmic Web of Dark Matter is traced by baryons. At large scales (super-Mpc) the distribution of galaxies (Fig. 4.1) follows the Cosmic Web. At smaller scales, the

Cosmic Web is traced by the so-called *Intergalactic Medium* (IGM). The IGM is largely of primordial composition, being constituted by Hydrogen and Helium, with the occasional pollution by metals (all elements beyond Helium). In particular, in Section 5.3 we discuss how we can observe the IGM through its neutral Hydrogen component.

In Section 5.4 we discuss to what extent Hydrogen follows the distribution of Dark Matter. We conclude that thermal effects caused by the formation of first stars and re-ionisation of Hydrogen can prevent it from following the distribution of Dark Matter at small scales and therefore mimic the effect of free-streaming of DM particles, creating therefore major systematic uncertainty for the whole approach. We briefly review what is known about the thermal history of Hydrogen.

We proceed in Section 5.5 by discussing that the distribution of Hydrogen can be observed through its imprints in absorption in the spectra of distant and bright sources. For example, the transitions of an electron between the ground and excited states of neutral Hydrogen are collectively called the *Lyman series* (with the individual transitions from the ground level to the excited ones called Lyman-α, -β, -γ, etc.). Neutral Hydrogen absorbs the light of a distant source (like quasars) and imprints *absorption lines* in the spectrum. The Lyman-α absorption happens in the ultraviolet, at $1215.67\,\text{Å}$. Since the light is traveling a large distance between the quasar and the Earth, it gets redshifted and characteristic quasar spectrum develops a set of absorption features at the locations of the intervening neutral Hydrogen (this happens if, like is the case, the intervening Hydrogen is highly ionized, being the Lyman-α transition sensitive at very small neutral Hydrogen fraction). A collection of these absorption features is called the *Lyman-α forest*.

The Fourier transform of the auto-correlation function of the spectra gives rise to the *Flux Power Spectrum* (FPS) – the main observable of the method (Section 5.6). A generic prediction of the WDM influence in the Lyman-α forest is the cutoff in the FPS at small scales.

The spatial scales that the Lyman-α forest method probes are directly proportional to the spectral resolution of the observing instrument. Initial, medium and low resolution spectra did not exhibit any suppression (the spectrum was rising as CDM would predict; see Section 5.6.2). In this situation the WDM bounds were derived from potential (within statistical error bars) deviation of the observed power spectrum from the one, predicted by the CDM (see e.g. [133, 148, 149]).

The situation has changed recently, when a few high-resolution and high-redshift quasar spectra became available [150, 151]. These data exhibit a very clear cutoff around $k \sim 0.03\,\text{s/km}$ (see Fig. 5.14; for details on units used in spectroscopy, refer to Sec. 5.6). This does not mean, however, that Warm Dark Matter with $\lambda_{\text{DM}} \sim \lambda_{\text{FS}}$ has been discovered! In Section 5.7 we discuss the interpretation of the available data and physical conclusions about the nature of Dark Matter that could be derived from them.

We discuss there that the observed suppression can be mimicked by a number of thermal effects. This can make the data consistent with CDM. In particular, the Doppler broadening introduces its own cutoff, related to the instantaneous temperature of the intergalactic

medium (IGM), while the history of re-ionisation and of the IGM heating manifests itself via another cutoff ("pressure effects"). This, in turn, does not also mean that the "CDM + thermal effects" interpretation of the data is correct. Assuming different (but not excluded neither experimentally nor theoretically) thermal histories of IGM, one could interpret the available data to exclude CDM or to constrain WDM very strongly. We briefly discuss in what way the uncertainty could be removed and how one can use these data before that.

The Lyman-α constraints can be interpreted in terms of constraints on particle physics models. Often 'WDM' refers to a somewhat unphysical scenario where DM is produced in thermal equilibrium and then freezes out while still being relativistic – *(warm) thermal relic* or *thermal WDM* [152]. In this case there is a one-to-one relation between λ_{DM} and the DM particle mass, m_{WDM} ($\lambda_{\mathrm{DM}} \propto m_{\mathrm{WDM}}^{-4/3}$), see *e.g.* [153]. Although it does not correspond to any particular particle physics model, this parametrization is often chosen when deriving WDM constraints and therefore we will refer to it throughout this Chapter.

Many particle physics candidates can play the role of Warm Dark Matter (including axinos [154, 155], gravitinos [156], sterile neutrinos [54], etc.). The focus of this thesis is on a particular DM candidate – *sterile neutrinos, resonantly produced in the presence of a lepton asymmetry* [157, 158]. If such a sterile neutrino (SN in what follows) is sufficiently light (masses of the order $m_{\mathrm{DM}}c^2 \approx \mathrm{keV}$), the 3D linear matter power spectrum exhibits a cutoff below a scale λ_{FS} that is a function of two parameters: the mass of the DM particle, $m_{\mathrm{DM}} \equiv m_{\mathrm{SN}}$, and the primordial lepton asymmetry parameter that governs its resonant production, L_6 (see e.g. [136, 153, 158]); see *e.g.* [54] for a review on keV sterile neutrinos as a DM candidate (see also Section 2.1).

5.2 Cosmic web and Warm Dark Matter

How to distinguish between Cold and Warm Dark Matter? In CDM the clumps of DM of all sizes are possible (up to some very small clumps with masses below one solar mass [159]). Hence there should be a lot of small-mass substructures in the Cosmic Web. If DM is not cold, but has significant free-streaming, it can affect the structure formation and the number of structures that are smaller than the free streaming length is strongly suppressed. So, to distinguish between CDM and WDM it is enough to count the number of small structures. Yet another possibility is that WDM can modify properties of the DM halos themselves.

WDM observation in DM halos.   Several features can be signatures of WDM (see e.g. [160] for a review):

1. **Missing satellites problem** – fewer amount of dwarf galaxies are observed in the Milky Way and M31 than CDM predicts [161–163];

2. **Core-cusp problem** – pure CDM simulations predict behavour of the DM density in the central part of haloes $\rho \propto r^{-\alpha}$ with $\alpha \approx 1$. However, it is claimed that for many objects $\alpha < 1$ describes observational data much better [164, 165];

3. **Too-big-to-fail problem** – large satellites with the maximum circular velocity $V_{\max} > 30\,\mathrm{km/s}$ should contain stars and be observed. However, the naive CDM prediction for the number of such objects in the Milky Way (~ 10) is in a tension with the fact that only 3 such objects were detected [166, 167].

All these problems, however, can be solved within CDM if one takes into account observational uncertainties, baryonic effects and variance in the initial conditions [168–172]. So, to confidently distinguish between CDM and WDM we need to look at the number of small dark halos, that are not so affected by baryonic effects, variance, etc.

The main difficulty is that small haloes may be very difficult to observe. Indeed, for a halo to be visible now it is important that at the early stage of the halo evolution it confines gas and eventually stars are formed. At the same time, if primordial hydrogen is already hot at the time of halo formation, it can fall into a halo only if this halo is massive enough. Primordial hydrogen is heated during reionization by the light of first stars. We do not know exactly to what temperature it is heated. But in simulations, this temperature can be large enough such that we can expect that small haloes cannot confine the gas and today they are made only of DM [172].

There are some clever ideas how to observe these small dark halos using the gravitational lensing. There are two possible methods: search for the perturbations in Einstein rings [173] or look at the flux ratios of multiply-imaged sources [174–176] that can be significantly increased because of the small haloes passing in front of the image. In 2022, the new telescope Euclid will be launched and will provide enough data to observe or constrain the number of small dark halos [177, 178]. This means that very soon we will either significantly constrain WDM or find a very strong evidence for it.

> The study of gravitational lensing is a promising approach that does not depend on the optical visibility of the structures and not as much on the high-resolution hydrodynamical simulations. This would allow either to rule out CDM or to constrain WDM very significantly.

WDM observation in Cosmic Web. Apart from dark matter distribution in the individual haloes, it is possible to infer the free-streaming of DM particles from the structures of the Cosmic Web. At high redshifts the structures of interesting sizes are still not so far from the linear regime ($\delta\rho/\rho$ is not much greater than 1). Also, at larger redshifts we are observing larger volumes within the same solid angle. This is another advantage over the galaxy-scale measurements as the amount of observed structures is much greater.

> Structure of the Cosmic Web at small scales can discriminate between Cold and Warm Dark Matter models.

If DM is cold, its power spectrum to be featureless and defined by a few fundamental parameters at the matter-radiation equality epoch. The Matter Power Spectrum is given by:

$$\begin{cases} P(k < k_{eq}) = Ak \\ P(k > k_{eq}) = B\frac{\ln k^2}{k^3} \end{cases} \tag{5.2.1}$$
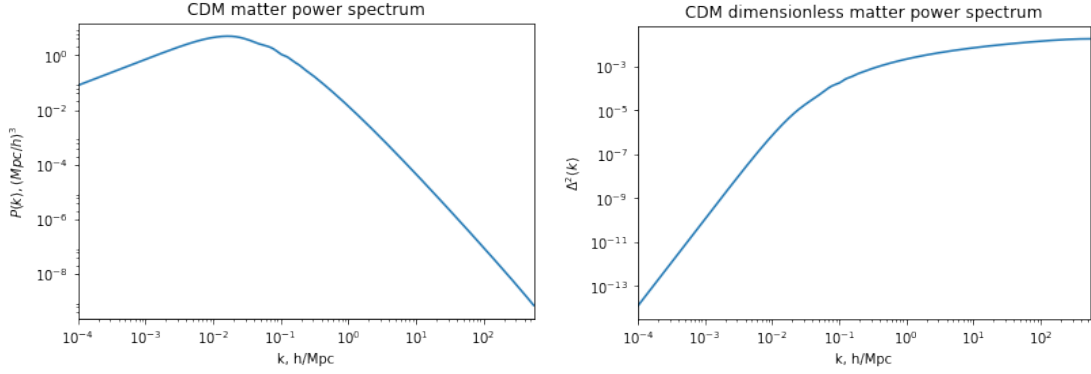
Figure 5.1: CDM Matter Power Spectrum $P(k)$ and dimensionless Matter Power Spectrum $\Delta^2(k)$

For convenience we also introduce dimensionless Matter Power Spectrum $\Delta^2(k)$:

$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2} \tag{5.2.2}$$

$$\begin{cases} \Delta^2(k < k_{eq}) = \frac{Ak^4}{2\pi^2} \\ \Delta^2(k > k_{eq}) = \frac{B\ln k^2}{2\pi^2} \end{cases} \tag{5.2.3}$$

However if DM is warm, below some scales the number of structures will quickly decrease and therefore we expect a *cutoff* in the $P(k)$. The location and the shape of cutoff depends on the particular properties of DM particles (Fig. 5.2).

The effect of WDM on the close-to-linear scales is the suppression of small-scale correlations caused by DM particle free-streaming (see Fig. 1.5). Our goal should be therefore to detect such a cutoff of DM power spectrum.
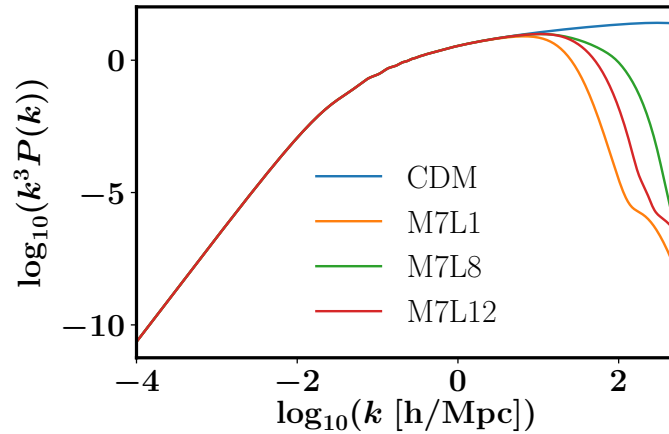
Figure 5.2: Matter Power Spectrum in CDM and various sterile neutrino WDM cosmologies. Sterile neutrinos with mass of $m_{SN} = 7\,\mathrm{keV}$ and lepton asymmetry numbers $L_6 = 1, 8, 12$ are shown.

5.3 Cosmic Web and neutral Hydrogen

Unlike galaxies that make the large-scale Cosmic Web visible, the IGM gas by itself does not shine. Indeed, the average IGM density is sufficiently low for virtually no star formation to take place there.[2] While it is not possible to observe IGM *in emission*, it may be possible to observe it *in absorption*. For this we would need a background 'flashlight'to illuminate the structures. In particular, we can observe the IGM via absorption features in the spectra of very bright and distant objects, typically quasars. In the future, it may be possible to revealing the IGM through the 21cm line in absorption or emission [182].

IGM absorption signal is due to neutral Hydrogen – the strong absorber and the most ubiquitous element in the Universe. In order to learn from these absorption features about the Dark Matter distribution, one should understand at what scales/redshifts does neutral hydrogen follow the Dark Matter.

5.3.1 The timeline

Naively, the evolution of hydrogen and Dark Matter do not have much in common. Indeed

– In radiation-dominated epoch DM density perturbations grow (at a slow logarithmic pace) while protons are tightly coupled with photons and support sound waves, rather than experience gravitational instability.

---

[2]We will discuss later that in some WDM cosmologies it may actually be possible to produce stars in the intergalactic filaments [179, 180]. Alternatively, it may be possible to observe the part of the IGM closer to the galaxies in fluorescent emission [181]. However, for most of the part IGM is not actively emitting.

- After matter-radiation equality DM begins to cluster faster, but baryonic plasma is still interacting rapidly with radiation. Its density perturbations remain small $\delta_b \sim 10^{-5}$ and temperature follows $T \propto a^{-1}$.

- At $z \sim 1000$ baryonic plasma recombines into neutral hydrogen. Although interaction between neutral gas and photons is much weaker than in plasma, the number density of CMB photons is enough to keep the temperature of atoms the same as CMB temperature.

- Around $z \sim 100$ the CMB-baryon scattering is inefficient and the gas becomes independent of CMB. From this point, it is cooling more quickly than photons, as $T \propto a^{-2}$. This allows matter perturbations to catch up with DM. From this point on, gas is following DM.

- At later times, first stars and galaxies form, creating highly non-linear objects and sources of heat and ionizing background.

- When the amount of ionizing radiation becomes significant to affect the ionization balance of the Universe, it triggers the event of Reionization. Relatively quickly gas becomes highly ionized except for in dense collapsed structures. The temperature reaches $\sim 10^4$ K.

5.4 Thermal history of the Intergalactic Medium

The most obvious difference between DM and baryons lies in interactions. The self-interaction cross-section of DM is limited at the level of $\sigma/m_{DM} < 1\,\mathrm{cm^2/g}$ [183] and interaction with the Standard Model is constrained to be many orders of magnitude below this [184, 185]. Self-interaction of neutral gas is of the order of atomic-sized hard-ball cross-section $\sigma/m_p \sim 10^8 \mathrm{cm^2/g}$. Hence despite baryons are gravitationally subdominant and generally follow the distribution of DM, due to self-interactions their distribution can differ in large density regions and on smallest scales.

The extent to which the gas deviates from DM depends on the characteristic pressure scale – the *Jeans length* $\lambda_J \sim \sqrt{\dfrac{kT}{G\rho m}}$. Below this scale the structures are washed-out by pressure. It manifestly depends on the gas temperature which can change rapidly through photoheating. However, redistribution of gas takes *dynamical time* $t_{dyn} \sim 1/\sqrt{G\rho}$ which can be larger. In this case not only the current temperature and pressure define the scale of washed-out structures, but also the whole thermal history.

Ionization state of the gas determines the opacity of the medium – i.e., the amount of absorption. There is a degeneracy between the total gas density and ionization fraction. However, as gas is expected to be in ionization-recombination equilibrium, the information about the thermal state lifts this degeneracy.

5.4.1 Pressure effects in gas

We follow [186] and consider a simplified model of two-component system of DM and baryons. The linear evolution of perturbation in the Dark Matter-baryon fluid is described by

$$\begin{array}{l} \frac{d^2\delta_X}{dt^2} + 2H\frac{d\delta_X}{dt} = 4\pi G\overline{\rho}\left(f_X\delta_X + f_b\delta_b\right) \\ \frac{d^2\delta_b}{dt^2} + 2H\frac{d\delta_b}{dt} = 4\pi G\overline{\rho}\left(f_X\delta_X + f_b\delta_b\right) - \frac{c_S^2}{a^2}k^2\delta_b \end{array}, \tag{5.4.1}$$

where δ_X and δ_b are the Fourier components of density fluctuations in the DM and baryons, f_X and f_b are the respective mass fractions, and c_s is the gas as sound speed.

Pressure term in the baryon perturbation equation suppresses the fluctuations below the characteristic pressure Jeans scale:

$$k_J = \frac{a}{c_S}\sqrt{4\pi G\overline{\rho}} \tag{5.4.2}$$

In the limit of baryon fraction f_b going to zero, this gives a filtering in baryon perturbations:

$$\delta_b(t,k) = \frac{\delta_X(t,k)}{1 + k^2/k_J^2} \tag{5.4.3}$$

5.4.2 Pressure support in expanding Universe

The distribution of matter also depends on the whole thermal history of gas temperature. At every point in time gas pressure is defined by the temperature. If the corresponding Jeans scale is larger than the size of the structure, gas expands and leaves the gravitational well.

However, while the temperature of the gas can change very fast due to radiation moving at the speed of light, the gas itself redistributes at the dynamical timescale $\sim \sqrt{\frac{1}{G\rho}}$. This means that during events like reionization gas temperature and pressure change rapidly, but the Jeans scale does not characterize the distribution of gas if the dynamical timescale is larger than the heating timescale. Vice versa, pressure at early times suppresses small structures and delays their growth until much later.

According to [186], a relevant scale for pressure suppression in the linear regime is given by

$$\frac{1}{k_F^2(a)} = \frac{3}{a}\int_0^a \frac{da'}{k_J^2(a')}\left[1 - \left(\frac{a'}{a}\right)^{1/2}\right] \tag{5.4.4}$$

(above the matter-dominated epoch with $\Omega_m = 1$ is assumed – which is applicable to the high enough redshifts $z \gtrsim 3$)

We consider two illustrative cases for the thermal history:

1. For gas coupled to CMB at $z \gtrsim 100$ with $T \propto 1/a$ the filtering scale equals to the Jeans scale: $k_J \equiv k_F$. In case when gas temperature evolves as $T \propto 1/a^2$ (as happens between reionization and $z = 100$) $k_F < k_J$. And for rapidly heating gas like during reionization $k_F > k_J$.

2. In a case of a schematic thermal history with $T \propto 1/a$, $T \propto 1/a^2$ and reionization regimes, the filtering scale turns out to be about 4 times different from the Jeans scale at the end of reionization: $k_F \sim 4k_J$

> In static conditions, pressure support creates a minimal scale for gas clouds – the *Jeans scale* λ_J that depends on the instantaneous gas temperature. However because of finite time required to redistribute the gas, the minimum cloud size in the expanding Universe with changing temperature is the *filtering scale* λ_F, defined by the whole thermal history.

5.4.3   Epoch of Reionization

Effect of the ionizing radiation in linear approximation can be described by the following system of equations [187]:

$$\frac{dT}{dt} = -2HT + \frac{2T}{3(1+\delta)}\frac{d\delta}{dt} - \frac{T}{\sum \tilde{X}_i}\frac{d\sum \tilde{X}_i}{dt} + \frac{2}{3k_B n_b}\frac{dQ}{dt} \qquad (5.4.5)$$

$$\frac{d\tilde{X}_i}{dt} = -\tilde{X}_i\Gamma_i + \sum \tilde{X}_j\tilde{X}_k R_{jk}\frac{\bar{\rho}_b(1+\delta)}{m_p} \qquad (5.4.6)$$

where equation on temperature accounts for expansion, adiabatic heating, redistribution of energy per particle due to changes in abundances and external heating ($\dot{Q}$). The abundance equations explicitly account for ionizations and recombinations. The quantities $\tilde{X}_i$ are defined by $n_i = \frac{(1+\delta)\bar{\rho}_b}{m_p}\tilde{X}_i$ such that e.g. $X_{HI} = \frac{\tilde{X}_{HI}}{\tilde{X}_{HI}+\tilde{X}_{HII}}$. The photoionization Γ_i and photoheating $\dot{Q}$ rates are given by:

$$\Gamma_i = \int_{\nu_i}^{\infty} 4\pi J(\nu)\sigma_i \frac{d\nu}{h\nu} \qquad (5.4.7)$$

$$\frac{dQ}{dt} = \sum n_i \int_{\nu_i}^{\infty} 4\pi J(\nu)\sigma_i(h\nu - h\nu_i)\frac{d\nu}{h\nu} = n_i\epsilon_i \qquad (5.4.8)$$

Γ_i has units of s^{-1}, ϵ_i is $\mathrm{erg\,s^{-1}}$ and $J(\nu)$ is $\mathrm{erg\,s^{-1}cm^{-2}sr^{-1}Hz^{-1}} = \mathrm{erg\,cm^{-2}sr^{-1}}$.

The quantity $J(\nu)$ is the mean intensity of the radiation that encompasses all information about the sources. Since the cross-sections σ_i are strongly peaked at ν_i, most of the information about the ionization and heating rates is provided by the amplitude and slope of $J(\nu)$ at ionization thresholds ν_i. Radiation intensity is computed based on the properties of the sources, but the intervening material is known to modify the spectrum of radiation as well [188].

After reionization, the gas is highly ionized by collisions with UV-background photons counteracted by recombination of electrons with the nuclei. Demanding the equilibrium of
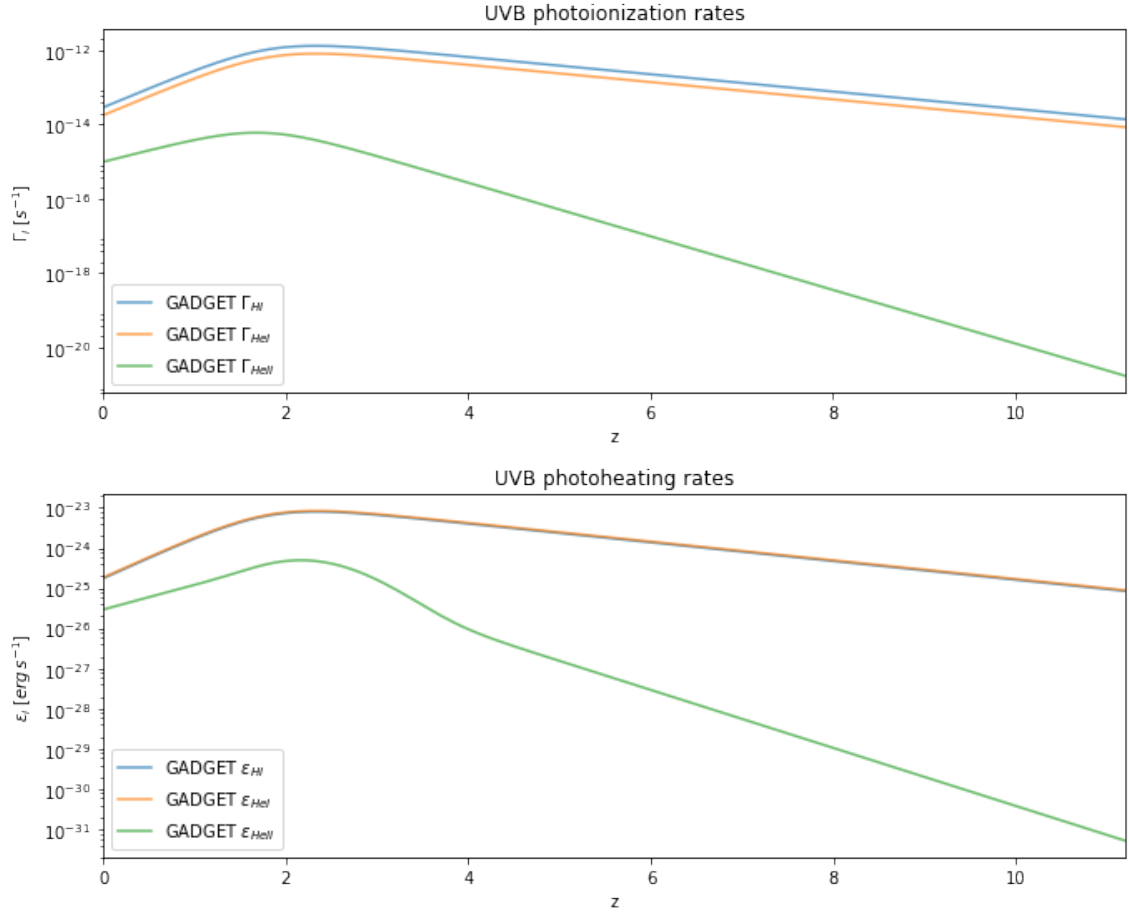
Figure 5.3: Example of an ultraviolet background created by stars typically used in simulations. *Upper panel* shows the photoionization rates for Hydrogen and Helium species, while *lower panel* shows the corresponding photoheating rates.

the system, we can write

$$n_{HI}\Gamma_{\gamma HI} = \alpha_{HII}n_e n_{HII} \tag{5.4.9}$$

In the highly ionized state $n_H \approx n_{HII}$, hence

$$\frac{n_{HI}}{n_H} \approx \frac{n_{HI}}{n_{HII}} = \frac{\alpha_{HII}}{\Gamma_{\gamma HI}}\frac{n_e n_{HII}}{n_{HII}} = \frac{\alpha_{HII}}{\Gamma_{\gamma HI}}n_e \tag{5.4.10}$$

The electron density is given by total density of baryons and non-Hydrogen mass fraction ($\approx 1 - Y$ with $Y \approx 0.24$ being the Helium mass fraction)

$$n_e = n_H + 2n_{He} = n_H\left(1 + 2\frac{Y}{1-Y}\frac{m_H}{m_{He}}\right) \approx n_H\left(1 + \frac{Y/2}{1-Y}\right) = n_H\frac{1-Y/2}{1-Y} \tag{5.4.11}$$

Then,

$$n_{HI} = \frac{\alpha_{HII}}{\Gamma_{\gamma HI}} n_H^2 \frac{1 - Y/2}{1 - Y} \tag{5.4.12}$$

We see that neutral Hydrogen density in ionization equilibrium is proportional to the square of the total Hydrogen density. Also note that $\alpha_{HII} \propto T^{-0.7}$ – so both increasing temperature and higher UVB ionization rate decrease the neutral Hydrogen fraction.

Next, according to [186], as long as the UVB is sufficient to ionize the Universe, the average temperature of the gas does not depend on the precise spectrum of photons and can be expressed as

$$n_{HI} \to 0 \quad \Rightarrow \quad T_0 \to \frac{\epsilon_{HI}}{3k_B \Gamma_{\gamma HI}} \tag{5.4.13}$$

i.e., temperature is given only by the ratio of photoheating and photoionization rates.

The details of galaxies emission are thought to be crucial for the state of the IGM. In fact, the temperature of the gas is thought to result from a balance between photo-heating from the ultraviolet background generated by the galaxies and the adiabatic cooling of the Universe [187, 189]. This balance results in a tight power-law relation between gas temperature and density, the *temperature-density* (or $T - \rho$) *relation*:

$$T = T_0 \left(\frac{\rho}{\bar{\rho}}\right)^{\gamma - 1}, \tag{5.4.14}$$

where $\bar{\rho}$ denotes the mean density. In particular, the details of the hardness of the sources of reionization is crucial for the discussion of the state of the IGM at high redshift.

As gas is impulsively heated during reionisation, the heat input per hydrogen atom is mostly independent of density, driving $\gamma \to 1$. The heating rate then drops as the gas becomes ionised, but more so at low density than at high density. This steepens the temperature-density relation asymptotically to $\gamma - 1 = 1/(1 + 0.7) \sim 0.6$, with the factor 0.7 coming from the temperature dependence of the CASE-A H II recombination coefficient [189, 190]. The characteristic timescale for approaching the asymptotic value is of the order of the Hubble time. If reionisation indeed happens late, $z \sim 7.5$, then we would expect $1 < \gamma < 1.6$.

5.4.3.1 Sources of reionization

What about the mechanism of this *reionization* event? Initially neutral atoms form when the Cosmic Microwave Background decouples from the baryons – i.e., when the interaction rate between plasma and photons becomes less than the Universe expansion rate. This happens at very early times ($z \sim 1100$) when the temperature of radiation is very high ($T \sim 3000\,\mathrm{K}$). Since then, because of the expansion of the Universe, the temperature
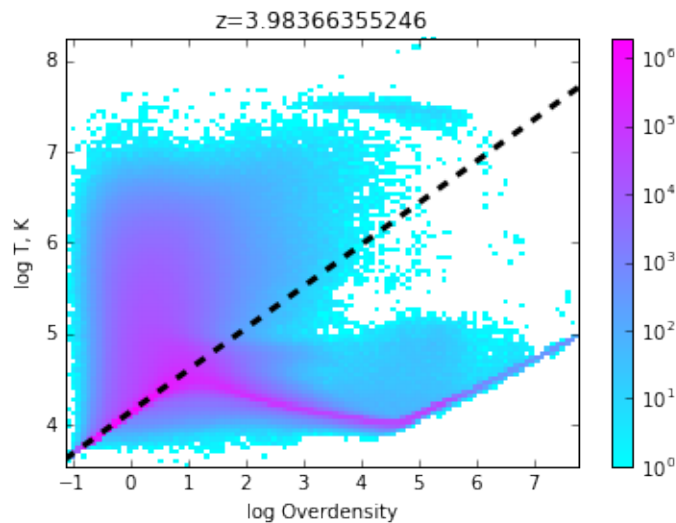
Figure 5.4: Temperature-density relation observed in simulations. Color shows the number of particles. The dashed line shows the linear fit to the range of overdensities $0.1 \ldots 1$

decreased as $\propto 1/a = (1 + z)$, so at $z \ll 100$ the CMB could not possibly be the cause of the reionization.

The reionization of hydrogen and of the first level of helium happen respectively at $E \sim 13.6\,\text{eV}$ and $E \sim 24.6\,\text{eV}$, and they are considered consistent with stellar reionization. The reionization of the second level of Helium happens at $E \sim 54.4\,\text{eV}$, hence it requires much harder sources spectra than the ones of hydrogen and first level of helium. Hence, it is believed to be driven mainly by quasars. However, poor knowledge of the halo mass function (especially at small masses and high redshifts) and quasar population makes it unclear in what proportion quasars, high- and low-mass galaxies contribute at different times.

Among other the possible sources of the ionizing radiation are evaporating black holes, various annihilating or decaying particles (see e.g. [191–193]).
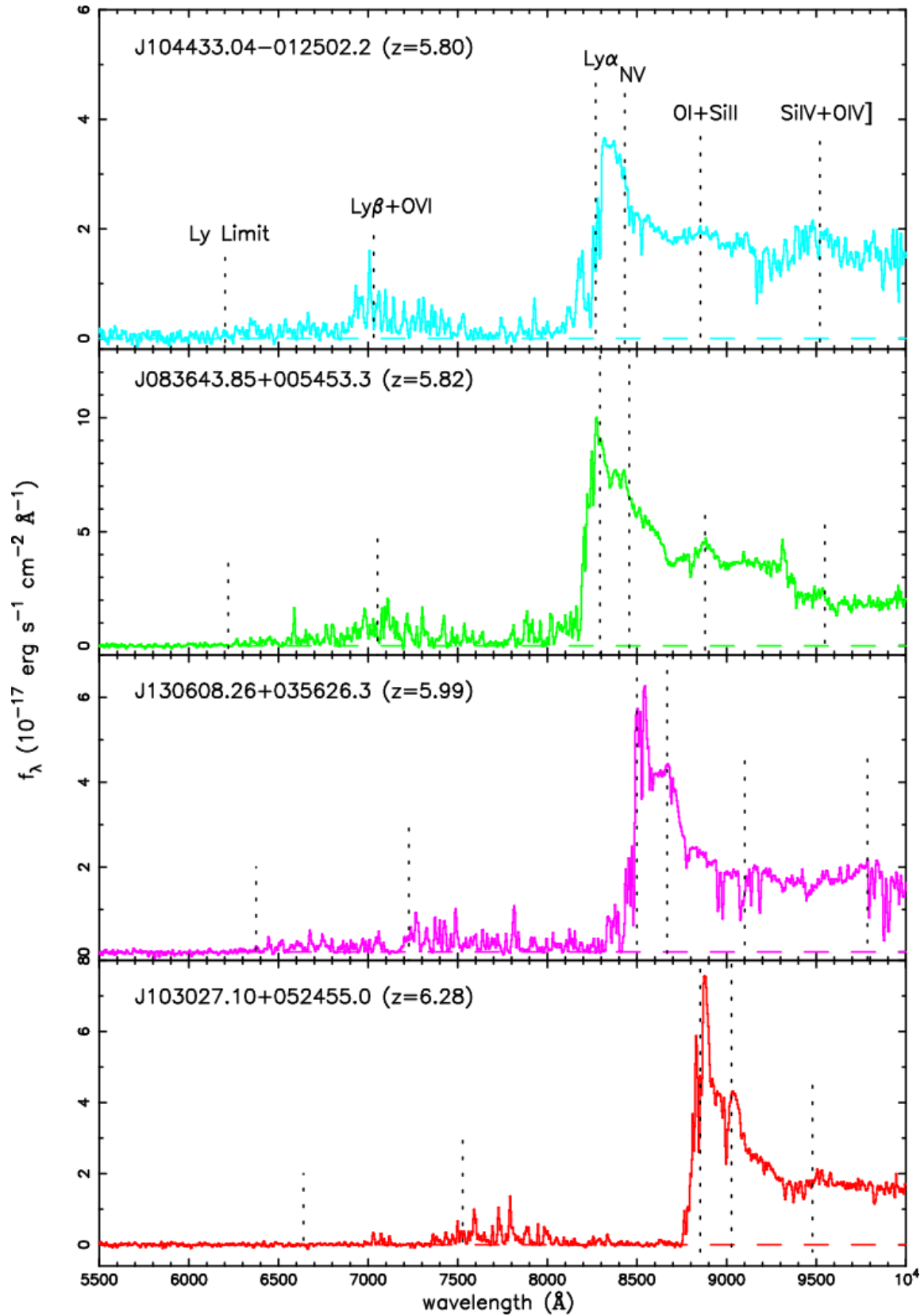
Figure 5.5: Spectra of 4 quasars located around redshift $z \sim 6$. Note how flux to the left of the Lyman-α emission peak disappears after redshift $z = 6$. *Credit: Becker et al.* [194]

5.4.3.2 Timing of reionization

After recombination the Universe becomes cold, neutral and transparent to photons. However, observations indicate that the galaxies and clusters contain mostly hot ionized gas. This discrepancy is associated with heating up of the gas falling into gravitational wells as well as photoheating and photoionization. It is plausible to assume that star formation and winds originating from galaxies are responsible for ionization of the gas.

It has been observed that the flux bluewards of the Lyman-α emission peak in quasar spectra rapidly decreases around redshift $z \sim 6$. This phenomenon can be explained by the fact that Lyman-α transition is a relatively strong one and even a small density of neutral hydrogen can completely absorb all photons. Indeed, given typical cosmological hydrogen density, a neutral fraction of $\gtrsim 10^{-4}$ is capable of creating the region of full absorption – the so-called Gunn-Peterson trough [195] (Fig. 5.5). And vice versa, for transmission, the neutral fraction should not exceed this value. Figure 5.5 shows spectra of four quasars around redshift 6.

Since this happens universally for all distant-enough quasars (albeit with some redshift variation), we can conclude that hydrogen undergoes a process of ionization at $z \gtrsim 6$ that results in appearance of the highly ionized regions and overdense (or slightly more neutral) structures that correspond to absorption features.

Moreover, at $z \lesssim 6$ the IGM is ionized up to a very high degree – fraction of neutral hydrogen is less than 10^{-4}:

$$\tau(z) \sim 6.6 \cdot 10^3 \frac{\Omega_b h}{0.02} (1+z)^{\frac{3}{2}} x_{HI}(1+\delta) \qquad (5.4.15)$$

Given there is enough time to establish the photoionization equilibrium, the gas should also be hot – of the order $10^4 K$. This is corroborated by the Doppler broadening of individual absorption features.

From this we can conclude that the Universe was somehow heated up and ionized around $z \sim 6$. This event is referred to as the *Epoch of Reionization*.

5.4.4 Summary of the reionization

We observe that at the time of formation of CMB the Universe became non-ionized, while at the moment it is ionized up to a very high degree. This means that at some point reionization happened. Among the possible sources of reionization are star-forming galaxies and quasars. The Lyman-α data point that reionization has been completed by $z \sim 6$.

However, the reionization history is weakly constrained by the experimental data. In [196] the constraints on reionization were put using data from CMB and Lyman-α emission at redshifts 7-8. The results are shown in Fig. 5.6. We see that very different histories of ionization fractions are possible, including even non-monotonic ones. As we will see in Sec. 5.7, these different thermal histories may cause deviations of the Hydrogen distribution from the distribution of dark matter at very different scales.
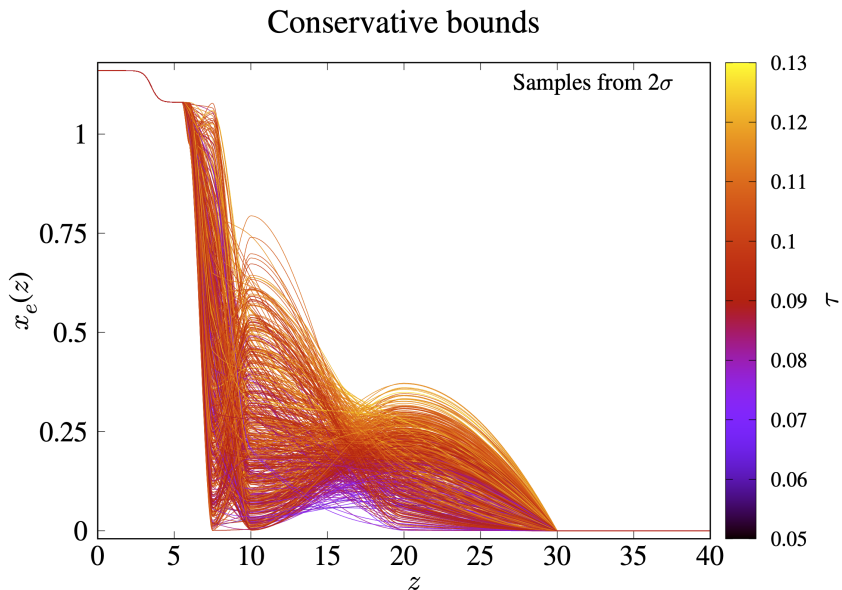
Figure 5.6: A collection of reionization histories consistent with CMB and Lyman-α emission data. Curves present individual histories of free electron fraction evolution $x_e(z)$ coloured by the corresponding optical depth to reionization τ. *Credit: Hazra et al.* [196]

We conclude that thermal effects caused by the formation of first stars and reionization of Hydrogen can prevent it from following the distribution of Dark Matter at small scales and therefore mimic the effect of free-streaming of DM particles. Our current knowledge about reionization and thermal history of intergalactic medium does not allow to predict at what scales these thermal effects will show up. This creates a major systematic uncertainty for the whole approach. Once we see a cutoff in the distribution of Hydrogen, we have to investigate its origin before interpreting the results.

5.5 How to measure Hydrogen power spectrum

Even if at the smallest scales neutral Hydrogen can be affected by thermal effects that take place during (and after) reionization (see Sec. 5.4), it can still be used to trace dark matter distribution in the Cosmic Web for a wide range of scales.

There are two main observational approaches to do this – the *21 cm line* [182] (related to the transition between hyperfinely split levels of the Hydrogen atom ground state) and the *Lyman-α absorption*. To date, the measurement of the 21cm signal is an extremely active area of research, but even the most rough density distributions obtained in this manner are still years in the future [197]. Hence here we discuss the second approach. For this we need distant sources, preferably with a featureless spectrum in the corresponding part of the spectrum. The best available candidates are quasars.

5.5.1 Quasi-Stellar Objects
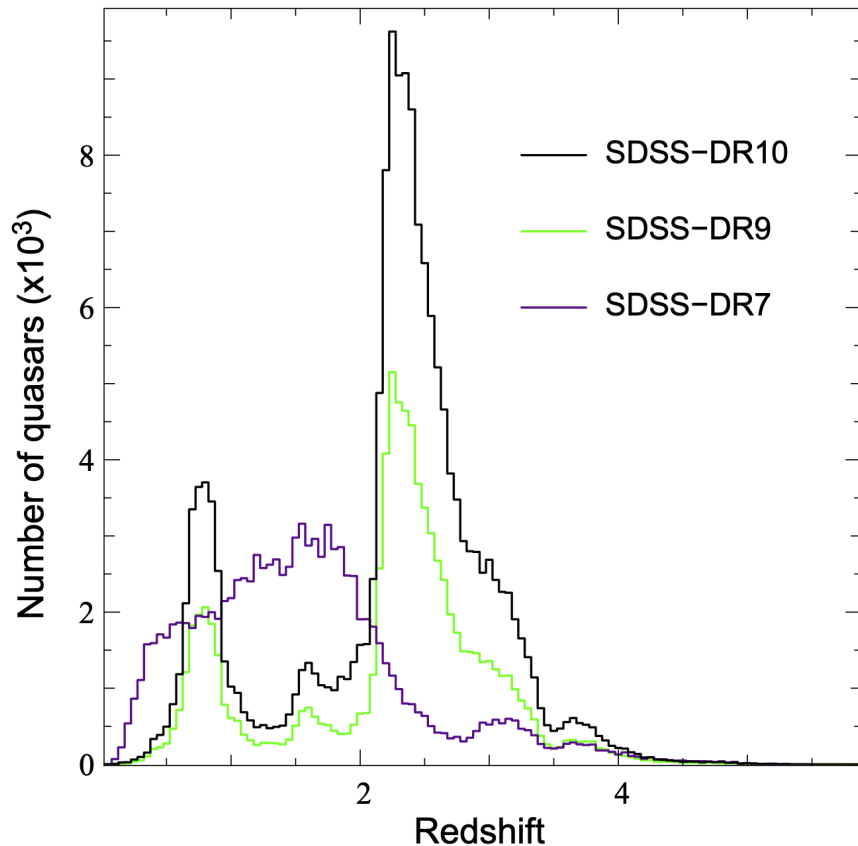
Figure 5.7: Quasar redshift distribution observed by Sloan Digital Sky Survey. *Credit: Paris et al.* [198]

Quasars are extremely bright objects found at huge range of redshifts up to $z \sim 10$, see Fig. 5.7 with quasars distribution by redshift. Most of the quasars are located around $z \sim 2.5$, but their numbers sharply fall at larger z. They emit significant radiation at wide range of wavelengths (as opposed to the stellar black-body spectrum centered around the temperature).

Quasar spectra contain numerous emission peaks corresponding to atomic transitions, see Fig. 5.8. As the light emitted travels from the quasar, it gets redshifted. If the light at the redshifted frequency is close to the one of the baryonic transitions, the photons might be absorbed by the intervening atoms, and later re-emitted in a random direction. This means that *bluewards* of each emission feature, we expect a set of absorption features, due to scattering of the light with the IGM. Moreover, as quasars emit at large range of frequencies, these sets of features can significantly overlap, making the identification difficult.

Luckily, some ranges of wavelengths with little interference were identified. They belong to Lyman series of transitions of the lightest elements – hydrogen and helium. Due to their primordial large abundances, these features are the most intense.
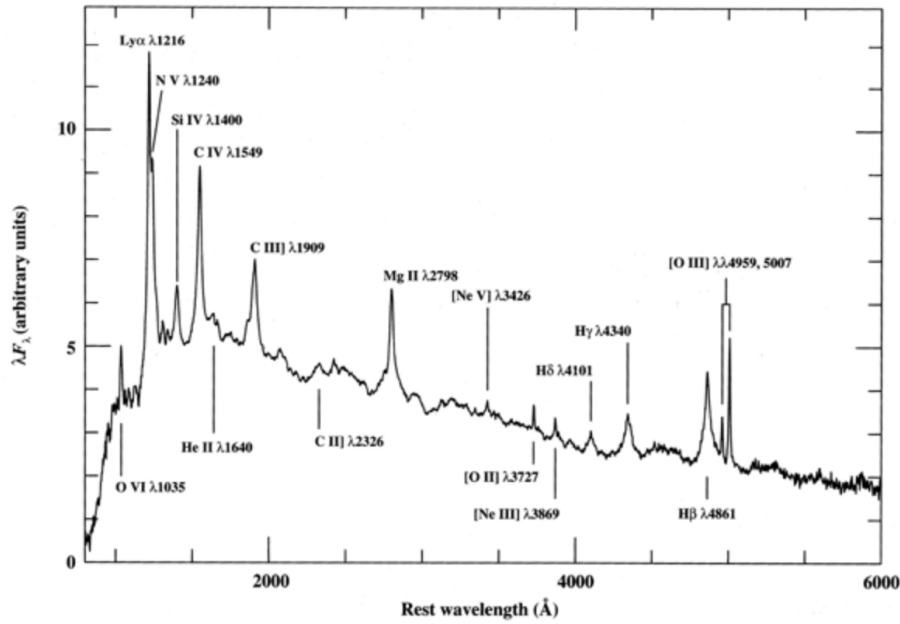
Figure 5.8: A mean QSO spectrum formed by averaging spectra of over 700 QSOs from the Large Bright Quasar Survey. Prominent emission lines are indicated. *Credit: Francis et al.* [199]

Especially interesting is the Lyman-α transition with wavelength $\lambda_\alpha = 1215.67\,\text{Å}$. It corresponds to transition between the ground level and first excited level of the electron in the atomic hydrogen. This is the most prominent peak in quasar spectra, as shown in Figure 5.9. Large numbers of absorption features at all redshifts can be identified between Lyman-α frequency and the next one in the series – Lyman-β.

We can use the section of the quasar spectrum between Lyman-α and Lyman-β emission peaks to measure the distribution of H I in the Universe.

5.5.2 Absorption features

The properties of the absorbers are defined by the matter distribution as well as the thermal state of the gas.

For an isolated cloud of gas, the resulting quasar spectrum feature is defined by:

- Size of the cloud along the line of sight (light passing along the filament will get a wide feature)

- Neutral hydrogen density (ionized clouds are transparent; recombined gas in galaxies is opaque)

- Peculiar motion of the cloud with respect to the Hubble flow (similar to an error in determination of redshift)
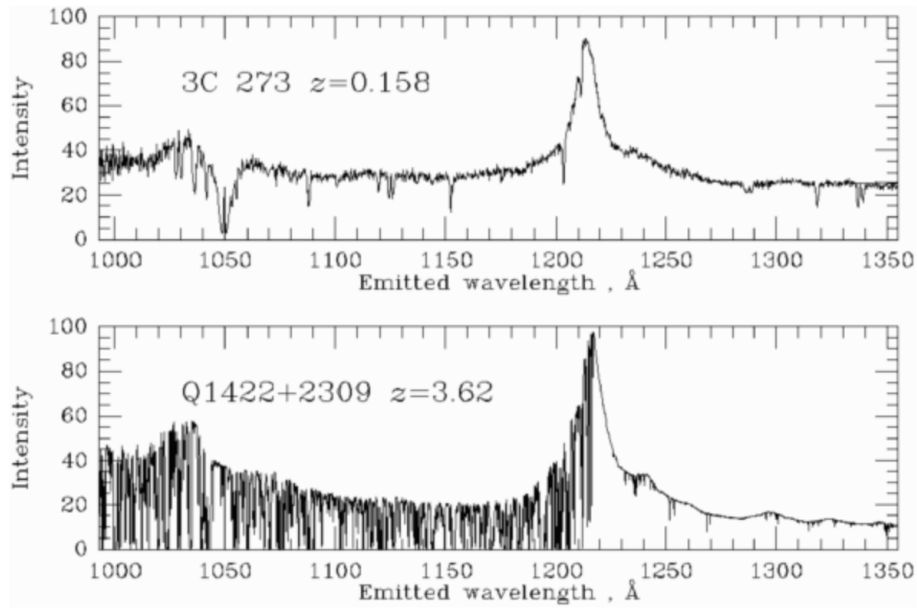
Figure 5.9: Two examples of quasar spectra in the emission wavelength. In the top panel we show the quasar spectra around of a quasar at $z = 0.158$, in the bottom panel the spectra of a quasar at $z = 3.62$. In both panels we observe the Lyman-α emission peak. In the bottom panel, bluewards of the peak, we observe a prominent Lyman-α forest, whereas in the top panel we observe more isolated absorption line. In both panel we observe on the left the Lyman-β peak. *Credit:* [`www.astro.ucla.edu`]

- Velocity gradient inside of the cloud (gas in non-collapsed structures retains a portion of Hubble expansion; gas in virialized halos has a narrower distribution of velocities)

- Gas temperature (motions of individual atoms)

> Each absorption line has two main characteristics: its depth and width. The depth of the absorption line is defined by the column density of the neutral Hydrogen along the line of sight. The line width depends on the Doppler effect (peculiar velocities and temperature of the gas) and physical size of the structure.

5.5.3   Optical depth

Observed quasar spectra are similar to the rays of light shining through the Universe which are focused on the Earth. When we study them, we aim to abstract from the properties of the particular quasar and to measure the fractional transmission $F(\nu) \equiv N_{observed}/N_{emitted}$. Hence we are not required *per se* to model the quasars. We will be modelling the quasar spectra using the *optical depth*

$$\tau = -\ln F. \tag{5.5.1}$$

127

The optical depth of light at some observed wavelength λ through Lyman-α transition by the neutral hydrogen gas with density n_{HI} along some distance R is given by the integral transition cross-section σ multiplied by the column density of the gas:

$$\tau(\lambda) = \int_R^0 \sigma\left(\frac{c}{\lambda}\right) n_{HI} dR, \tag{5.5.2}$$

where Lyman-α transition cross-section is given by:

$$\sigma(\nu) = \frac{\pi e^2}{m_e c} f \phi(\nu - \nu_\alpha), \tag{5.5.3}$$

where ν_α is the Lyman-α transition frequency, ϕ is the normalized line profile and $f = 0.416$ is the oscillator strength. ϕ can be taken either as delta-function, or as a Lorentz profile $\phi(\nu - \nu_\alpha) = \frac{1}{\pi} \frac{\gamma/2}{(\nu - \nu_\alpha)^2 + (\gamma/2)^2}$

Effect of cosmological expansion. To adapt this statement to the expanding Universe, we need to establish the relation between redshift z and proper distance R:

$$dR = a\,dr = c\,dt = c\frac{\dot{a}\,dt}{\dot{a}} = c\frac{da}{\dot{a}} = -c\frac{dz}{(1+z)H} \tag{5.5.4}$$

where $dr = \frac{c\,dt}{a}$ follows from the definition of the light-like interval $(-c^2 dt^2 + a^2 dr^2 = 0)$ Additionally, absorption at some distance happens at earlier moment of time, implying additional blue-shift of the argument of σ:

$$\tau(\lambda) = \int_0^z \sigma\left(\frac{c(1+z)}{\lambda}\right) n_{HI}(z) \frac{c\,dz}{(1+z)H(z)} \tag{5.5.5}$$

meaning that the light was absorbed at some smaller wavelength which was stretched by the expansion of the Universe.

Effect of thermal broadening and peculiar velocities. Let's assume that the medium is characterized by some temperature $T(z)$ and include the Doppler broadening into consideration. For simplicity, we will include this effect into the transition cross-section σ.

The one-dimensional distribution of velocity for a gas with average mass m and temperature T is the following Maxwell distribution:

$$f(v_z) = \sqrt{\frac{m}{2\pi kT}} e^{-\frac{m v_z^2}{2kT}} = \frac{1}{\sqrt{\pi} b_T} e^{-\frac{v_z^2}{b_T^2}} \tag{5.5.6}$$

with $b_T = \sqrt{\frac{2kT}{m}}$ being the thermal dispersion.

Now, let's integrate the cross-section with the velocity distribution substituting the wavelength with the value from the Doppler formula:

$$\sigma_T\left(\frac{c}{\lambda}\right) = \int_{-\infty}^{\infty} \sigma\left(\frac{c}{\lambda}\gamma(1-\beta)\right) \frac{1}{\sqrt{\pi}b_T} e^{-\frac{v_z^2}{b_T^2}} dv_z \tag{5.5.7}$$

We also can include the peculiar velocities along the radial direction by shifting the exponent

$$\sigma_T\left(\frac{c}{\lambda}\right) = \int_{-\infty}^{\infty} \sigma\left(\frac{c}{\lambda}\gamma(1-\beta)\right) \frac{1}{\sqrt{\pi}b_T} e^{-\frac{(v_z-v_0)^2}{b_T^2}} dv_z \tag{5.5.8}$$

Intrinsic broadening of the line together with Doppler effect constitute the so-called Voigt profile of the Lyman-α line.

Substituting modified Lyman-αtransition cross-section (5.5.8) into Eq. (5.5.5) we can obtain the optical depth that include effects of thermal broadening and peculiar velocities:

$$\tau(\lambda) = \int_0^z \frac{cdz}{(1+z)H(z)} n_{HI}(z) \int_{-\infty}^{\infty} dv\, \sigma\left(\frac{c(1+z)}{\lambda}\sqrt{\frac{1-v/c}{1+v/c}}\right) \frac{1}{\sqrt{\pi}b_T(z)} e^{-\frac{(v-v_0)^2}{b_T(z)^2}}$$

$$\tag{5.5.9}$$

5.5.4 Lyman-alpha forest in simulations

The integrand in Eq.5.5.9 is strongly peaked around $v = v_0$ and $\frac{c}{\lambda_\alpha} = \frac{c(1+z)}{\lambda}\sqrt{\frac{1-v/c}{1+v/c}}$ (where λ_α is the rest frame Lyman-α wavelength). It is convenient to use the 'Doppler' velocity instead of the redshift as the integration variable:

$$\lambda = \lambda_\alpha(1+z)(1+v/c) \tag{5.5.10}$$

This parametrization corresponds to the distance along the line of sight according to the Hubble law $\vec{v} = H\vec{r}$ under assumption of no peculiar velocities and Doppler broadening.

If we split the velocity space into bins of the size Δv, the Lyman-α optical depth in the velocity bin v_i will get contributions from species X in velocity bins v_j like so:

$$\tau(i) = \sum_X \sum_j \sigma_X \sqrt{\frac{m_X c^2}{2\pi k_B T(j)}} \rho_X(j) a \Delta v \exp\left(-\frac{m_X(v(i)-v(j))^2}{2k_B T(j)}\right) \tag{5.5.11}$$

Gas quantities associated to the velocity bins are estimated from the SPH particles

using the following definitions [200]:

$$\rho_X(i) = a^3 \sum_j \chi_X(j) W_{ij} \tag{5.5.12}$$

$$(\rho T)_X(i) = a^3 \sum_j \chi_X(j) T(j) W_{ij} \tag{5.5.13}$$

$$(\rho v)_X(i) = a^3 \sum_j \chi_X(j)(a\dot{x}(j) + \dot{a}(x(j) - x(i))) W_{ij} \tag{5.5.14}$$

where W_{ij} is the SPH kernel interpolation for locations $x(i)$ and $x(j)$ and $\chi_X(i)$ is the relative abundance of the specie X at the particle i.

The elements abundances are computed assuming the ionization-recombination equilibrium.

5.5.5 Our simulations of Intergalactic Medium

Hydrodynamical cosmological simulations usually expose the gas in the IGM to a uniform (homogeneous and isotropic) but evolving ionising background that mimics the combined emissivity of radiation from galaxies and quasars [see e.g. 201]. As a result, the mean neutral fraction is very low: $x \equiv n_{\mathrm{HI}}/n_{\mathrm{H}} \ll 1$. Without such an ultraviolet background (UVB), the effective optical depth would be much higher than observed [195].

Assuming that the UVB is uniform may be a good approximation long after reionisation, when fluctuations around the mean photoionisation rate, Γ_{HI}, are small [202, 203]. However, this may no longer be the case closer to reionisation when the UVB may be much more patchy [e.g. 204, 205]. The current best-estimate for the redshift of reionisation is $z_{\mathrm{reion}} = 7.82 \pm 0.71$, with a reionisation history consistent with a relatively rapid transition from mostly neutral to mostly ionised, and suggesting the presence of regions that were reionised as late at $z \sim 6.5$ [206]. These inferences obtained from the CMB are also consistent with hints of extended parts of the IGM being significantly neutral, $x \sim 0.1 - 0.5$, in the spectra of $z \gtrsim 7$ quasars [207, 208]. Such late reionisation, and the patchiness associated with it, make it much harder to perform realistic simulations of the IGM that yield robust constraints on λ_{FS}. In fact, the impact of large fluctuations in Γ_{HI} is not just restricted to inducing fluctuations in x, the neutral fraction, because the UVB also heats gas.

In this work, we have considered a suite of dedicated cosmological hydrodynamical simulations, and one of the simulations from the Eagle simulation suite. Our dedicated simulation suite has been performed using the simulation code used by [214]. This code is a modified version of the publicly available GADGET-2 TREEPM/SPH code described by [45]; the runs performed are summarised in Table 5.1. The values of the cosmological parameters used are in Table 5.2; runs labelled 'Planck' use parameters taken from [213], those labelled 'Viel' use parameters taken from [150] to allow for a direct comparison with the latter work.

Name	$L\,[\mathrm{Mpc}/h]$	N	Dark Matter	UVB	Cosmology
CDM_L128N64	128	64^3			
CDM_L20N512	20	512^3	CDM	no UVB	Viel
CDM_L20N896	20	896^3			
CDM_L20N1024	20	1024^3			
M7L1			$m_{SN} = 7\,\mathrm{keV}, L_6 = 1$		
M7L8	20	1024^3	$m_{SN} = 7\,\mathrm{keV}, L_6 = 8$	no UVB	Viel
M7L12			$m_{SN} = 7\mathrm{keV}, L_6 = 12$		
CDM_Planck_Late			CDM	*LateR*	
CDM_Planck_Early	20	1024^3	CDM	*EarlyR*	Planck
M7L12_Planck_Late			$m_{SN} = 7\,\mathrm{keV}, L_6 = 12$	*LateR*	
EAGLE_REF	$100\,/h$	1504^3	CDM	Eagle	Planck

Table 5.1: Hydrodynamical simulations considered in this work together with corresponding parameters. All simulations were performed specifically for this work, except `EAGLE_REF` [209]. Columns contain from left to right: simulation identifier, co-moving linear extent of the simulated volume (L), number of Dark Matter particles (N, there is an equal number of gas particles), type of Dark Matter (CDM or sterile neutrino WDM with the indicated particle mass, m_{SN} – expressed in natural units – and lepton asymmetry parameter, L_6), ultra-violet background imposed during the simulation (no UVB indicates no UVB was imposed; *LateR* and *EarlyR* refer to the UVBs from the *LateR* and *EarlyR* reionization models in [210], *Eagle* indicate the standard UVB from [211]), choice of cosmological parameters from Table 5.2, and figure where the particular simulation is used. The gravitational softening length for gas and Dark Matter is kept constant in co-moving coordinates at $1/30^{\mathrm{th}}$ of the initial interparticle spacing. All simulations were started from the initial conditions generated by the `2LPTic` [212] with the same 'glass'like particle distribution generated by GADGET-2 [45].

Cosmology	*Planck* [213]	*Viel* [150]
Ω_0	0.308 ± 0.012	0.298
Ω_Λ	0.692 ± 0.012	0.702
$\Omega_b h^2$	0.02226 ± 0.00023	0.022393
h	0.6781 ± 0.0092	0.7
n_s	0.9677 ± 0.0060	0.957
σ_8	0.8149 ± 0.0093	0.822

Table 5.2: Cosmological parameters used in our simulations. *Planck* cosmology is the conservative choice of TT+lowP+lensing from [213] (errors represent 68% confidence intervals), while *Viel* cosmology corresponds to the bestfit model in [150].

Initial conditions for the runs were generated using the `2LPTic` code described by [212], for a starting redshift of $z = 99$ that guarantees all sampled waves are still in the linear regime. The initial linear power spectrum for the CDM cosmology was obtained with the linear Boltzmann solver CAMB [215]. Sterile neutrino Dark Matter is also modelled as

non-interacting massive particles, with the effects of free streaming imprinted in the initial transfer function as computed with the modified CAMB code described by [153], using the primordial phase-space distribution functions for sterile neutrinos computed in [158]. Using instead results from the most recent computations [216, 217] would not change our results. We neglect the effects of peculiar velocities of the WDM particles other than the cutoff they introduce in the transfer function. The linear matter power spectra for the different models used in this work are shown in Fig.5.2.

Although it is possible to carry out approximately self-consistent simulation of the IGM during reionisation (e.g. [218]), such calculations are still relatively computationally demanding. We therefore use the following strategy in this work: we perform some of the simulations *without* imposing a UVB, meaning that effectively $\lambda_p = 0$. We then apply an 'effective' UVB in post-processing, by imposing a given temperature-density relation of the form given by Eq. (5.4.14) and scaling the neutral fraction x to obtain the observed effective optical depth (as described in more detail below). We stress therefore that many of our runs are not realistic, nor are they intended to be. Quite the opposite, we work in an idealised scenario that allows us to vary individually every relevant effect separately. In addition to these runs, we also carry out simulation that do impose a UVB on the evolving IGM – we use these to demonstrate that our limits on λ_{FS} are also valid in this more realistic scenario.

For simulations that include a UVB, we specify the redshift-dependent values of the photoionisation and photoheating rates for hydrogen and helium as input parameters. The version of GADGET that we use solves for the radiative heating and cooling of the photoionised gas, given these input rates. Imposing the rates of [210] results in a $T - \rho$ relation that is consistent with that of the latter authors. We use the same UVB in the SN cosmology as an *example* of the reionisation history with a small filtering scale.

SPH (gas) particles are converted to collisionless 'star' particles when they reach an overdensity $\rho/\bar{\rho} > 1000$ provided their temperature $T < 10^5$ K. This 'quick-Lyman-α' set-up reduces run time by avoiding the formation of dense gas clumps with short dynamical times, that would in reality presumably form stars in a galaxy. We can do so, because the impact of forming galaxies on the IGM is thought to be small, particularly at high redshifts and for the low density gas regions to which our analysis is sensitive [214, 219].

5.5.5.1 Calculation of mock spectra

We compute mock spectra of the simulations using the SPECWIZARD code that is based on the method described by [189]. This involves computing a mock spectrum along a sight line through the simulation box along one of the coordinate axis.

For simulations without a UVB (CDM_L20N1024, M7L12), we first impose a temperature-density relation of the form of Eq. (5.4.14) on all gas particles. At the high redshifts that we are considering, the Lyman-α transmission is non-negligible only for sufficiently small overdensities, $\delta \lesssim 1$. We checked explicitly that the effect of cooling at the highest densities is negligible for our analysis. Therefore, one can safely apply the temperature-density rela-
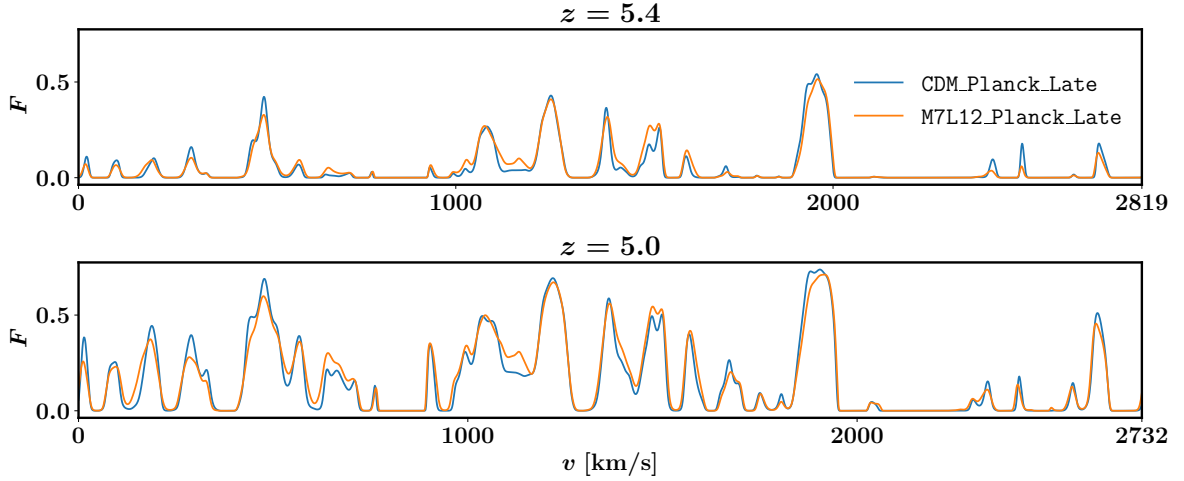
Figure 5.10: Example mock spectra extracted along the same line of sight in `CDM_Planck_Late` (blue line) and `M7L12_Planck_Late` (orange line), simulations at redshifts 5.4 (*top panel*) and 5.0 (*bottom panel*). The temperature T_0 of the gas at the mean density at these redshifts is ~ 7700K for both redshifts. Note that a sightline through the full extent of the box corresponds to a different velocity extent at different redshifts. The evolution of the mean transmission is apparent. The CDM and WDM spectra look quite similar, nevertheless on closer inspection it is clear that the CDM spectrum has some sharper features.
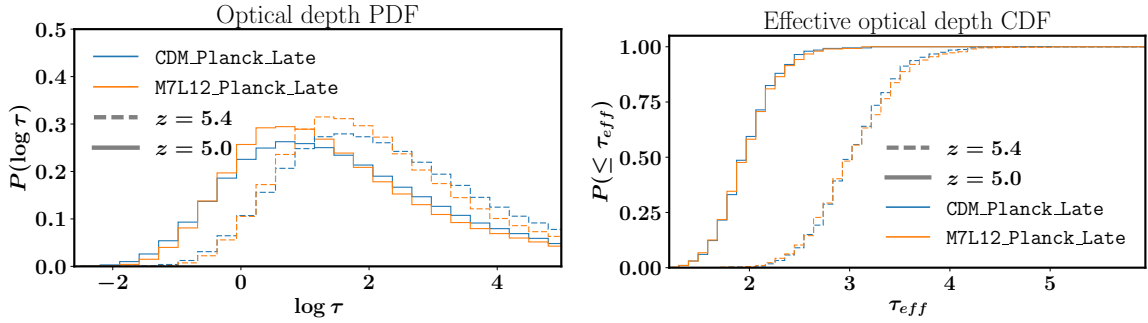
Figure 5.11: *Left panel:* probability distribution function of the optical depth per pixel. *Right panel:* cumulative probability distribution of the effective optical depth, τ_{eff}, measured in chunks of 50 Mpc h^{-1}. The `CDM_Planck_Late` model is plotted in *blue*, the `M7L12_Planck_Late` in *orange*, redshift $z = 5.4$ corresponds to *dashed lines* and $z = 5.0$ to *full lines*.

tion to the whole range of densities considered, without worrying about it being applicable only in the range $\delta \lesssim 10$ [187].

We use the same post-processing also for simulations which do include a UVB. The rationale behind this is the following. As already mentioned, we use [210] ionisation history only as an example of the model with small pressure effects, not as a holistic model. We then vary the T_0 in post-processing and determine the range of admissible temperatures in

CDM and WDM cosmologies. We verify *a posteriori* that the actual temperature predicted by the *LateR* model lies within the range of admissible temperatures.

Given T and ρ of each particle, we compute the neutral fraction x using the interpolation tables from [220], which assume photoionisation equilibrium,

$$\frac{dn_{\mathrm{H\,I}}}{dt} = -\Gamma_{\mathrm{H\,I}} n_{\mathrm{H\,I}} - \Gamma_c\, n_e\, n_{\mathrm{H\,I}} + \alpha(T)\, n_e\, n_{\mathrm{H\,II}} = 0\,. \qquad (5.5.15)$$

Here the terms from left to right are photoionisation by the imposed UVB, collisional ionisation, and recombination (with $\alpha(T)$ the temperature-dependent case-A recombination coefficient); n_e is the electron density; the photoionisation rate is that described by [211].

We then interpolate the temperature, density, and peculiar velocity to the sight line in bins of $\Delta v = 1\ \mathrm{km\ s^{-1}}$ using the Gaussian method described by [221]. We verified that this spectral resolution is high enough to give converged results. We then compute the optical depth as function of wavelength, $\tau(v)$, thus accounting for Doppler broadening and the effects of peculiar velocities.

To allow for a fair comparison to the observed spectra, we convolve the mock spectra with a Gaussian to mimic the effect of the line-spread function, and rebin to the observed pixel size with parameters as described in Section 5.6.2. The Gaussian white noise has a uniform relative Standard Deviation of $\sigma = 0.066$, corresponding to a signal to noise ratio of $S/N = 15$ per pixel at the continuum level, following [150]. Further details on the application of noise to mock spectra and comparison with previous work are given in Appendix 5.C. We calculate a set of such spectra for the snapshot at redshifts $z = 5$, and $z = 5.4$.

After repeating this procedure for $\mathcal{N} = 10^3$ sight lines, we compute the mean transmission, $\langle F \rangle = \langle \exp(-\tau) \rangle$ and scale the optical depth so that the ensemble of mock spectra reproduces the observed value of $\langle F \rangle$ discussed in Section 5.6.2.

We compare spectra along the same sight line for the CDM and the `M7L12_Planck_Late` models in Fig. 5.10 (blue and orange curves, respectively), at redshifts $z = 5.4$ (top panel), and $z = 5.0$ (bottom panel); the temperature and thermal history are the same for both models. The Lyman-α spectra look very similar in these models, although it can be seen that the CDM model has some sharper features. The probability distribution function (PDF) of the optical depth is compared between these two models in Fig. 5.11. It is unlikely that we will be able to distinguish the CDM and WDM using the information in individual spectra.

Ideally, we would like to measure a quasar spectrum and use it to reconstruct the Hydrogen density along the line of sight. However in reality beside the finite instrument resolution, the broadening of the absorption lines caused by thermal effects is typically larger than the distance between the lines.

This means that we cannot use the individual absorption features to study the small-scale statistics of the matter distribution. Instead, we will use statistical methods to characterise the properties of the Lyman-α forest from many absorbers and lines of sight.

5.6 Comparison of simulations and observations

Using quasar spectra as a proxy to Dark Matter distribution, our goal is to reconstruct the Matter Power Spectrum.

Having a sufficiently dense distribution of quasars on the sky, it would in principle be possible to estimate the 3D distribution of gas by interpolating between the sightlines in the transverse direction. Then we could compute the corresponding 2-point correlation function – the Matter Power Spectrum.

Unfortunately, quasars are not numerous enough for this exercise (however, correlation between close quasar pairs provides an important test for the structure sizes since the transverse correlations are independent of the thermal Doppler effect and peculiar velocities).

5.6.1 Flux power spectrum

As outlined above, in the present work we compare the mock FPS computed from simulations to the observed FPS presented by [150]. Traditionally the FPS is computed in 'velocity space'. Integrating the Doppler shift relation between wavelength and velocity, $dv/c = d\lambda/\lambda$, the redshift or wavelength along a line-of sight to a quasar can be written in terms of a 'Hubble' velocity v as

$$v = c \, \ln \left(\frac{\lambda}{\lambda_0(1+z)} \right) = \frac{H(z)}{1+z} y \,, \qquad (5.6.1)$$

where $\lambda_0 = 1215.67 \,\text{Å}$ is the laboratory wavelength of the Lyman-α transition, and z is a constant reference redshift. The zero-point of v is defined by z and is arbitrary. In data, z is often chosen to be the mean redshift of the data or the quasar's emission redshift, in simulations we take it to be the redshift of the snapshot. In this equation, $H(z)$ is the Hubble constant at redshift z, and the right-hand side also defines a co-moving position y along the spectrum.

The input to the FPS (either observed or obtained from simulations) is then flux as function of velocity, *i.e.* $F(v)$, over some velocity interval V (in the data set this interval is chosen so that one avoids the Lyman-β forest, the quasar near zone, and potentially some strong absorbers; in the simulations it is set by the linear extent of the simulated volume).

Given F and its mean, $\langle F \rangle$, we calculate the 'normalised flux'

$$\delta_F \equiv \frac{F - \langle F \rangle}{\langle F \rangle} \, . \tag{5.6.2}$$

The FPS is written in terms of the dimensionless variance $\Delta_F^2(k)$ (strictly speaking a variance in δ_F per dex in k), defined by

$$\Delta_F^2(k) = \frac{1}{\pi} k P_F(k) \tag{5.6.3}$$

$$P_F(k) = V \left\langle |\tilde{\delta}_F(k)|^2 \right\rangle \tag{5.6.4}$$

$$\tilde{\delta}_F(k) = \frac{1}{V} \int_0^V dv \, e^{-ikv} \delta_F(v) \, . \tag{5.6.5}$$

Here, $\langle \cdot \rangle$ denotes the ensemble average, and $k = 2\pi/v$ is the Fourier 'frequency' corresponding to v and has dimensions of (s/km). To find the conversion to a wave-vector in inverse comoving Mpc, k_x, recall that the Hubble law of Eq. (5.6.1) states that $\Delta v = H(z) \Delta y / (1+z)$. Then, since $k_y \, y = k_v \, v$, where $k_v \equiv k$, we find that

$$k_y = k_v \frac{H(z)}{1+z} \, . \tag{5.6.6}$$

Individual quasar spectra provide only 1D information, similar to the Matter Power Spectrum projected along the line of sight:

$$\Delta_{1d}^2(q) = \frac{q}{2\pi} \int d^2 k_\perp \frac{\Delta_{3d}^2(q, k_\perp)}{\left(q^2 + k_\perp^2 \right)^{3/2}} = q \int_q^\infty \frac{dk}{k^2} \Delta_{3d}^2(k) \tag{5.6.7}$$

with Δ_{3d}^2 being the 3D Matter Power Spectrum.

Residual neutral hydrogen gas in the intergalactic medium (IGM) produces a series of absorption lines in the spectrum of a background source such as a quasar, through scattering in the $n = 1 \rightarrow 2$ Lyman-α transition (see *e.g.* the review by [222]). The set of lines for which the column density of the intervening absorber is low, $N_{\mathrm{H\,I}} \leq 10^{16} \mathrm{cm}^{-2}$, is called the Lyman-$\alpha$ forest. The transmission F, i.e. the fraction of light of the background source that is absorbed, is often written in terms of the optical depth τ, as $F = \exp(-\tau)$; we will refer to this quantity that is independent of the quasar spectrum and only depends on the intervening distribution of neutral gas, as the flux[3]. The observed power spectrum of F exhibits a cutoff on scales below $\lambda_F \approx 30 \, \mathrm{km \, s}^{-1}$ at high redshift, and currently

---

[3]Let $\mathcal{F}$ be the observed quasar flux, and $\mathcal{C}$ what would be the observed flux in the absence of absorption, then $F \equiv \mathcal{F}/\mathcal{C}$ is the transmission. This quantity is commonly but somewhat inaccurately referred to as the 'flux', we will do so as well. Since $\mathcal{C}$ is not directly observable, neither is F. Estimating F from $\mathcal{F}$ is called 'continuum fitting'.

provides the most stringent constraints on λ_{FS} [133, 148, 150, 223–225]. The reason that the Lyman-α forest provides such tight constraints on λ_{FS} is that the neutral gas follows the underlying Dark Matter relatively well, because the absorption occurs in regions close to the cosmological mean density, particularly at higher redshifts $z \geq 5$. Nevertheless there are complicating factors, which include:

(*i*) the density is probed along a single sightline; the measured one dimensional (1D) power spectrum is an integral of the 3D underlying matter power spectrum (as discussed in details in Appendix 5.7.1.1);

(*ii*) the flux is related to the density by a non-linear transformation [226];

(*iii*) absorption lines are Doppler broadened;

(*iv*) the gas distribution is smoothed compared to the Dark Matter due to its thermal pressure [227].

As a consequence, $\lambda_{\mathrm{DM}} \neq \lambda_F$, and numerical simulations that try to account for all these effects are used to infer λ_{FS} by calculating mock absorption spectra, and comparing λ_F from the simulations to the observed value. However, the temperature of the gas, and hence the level of Doppler broadening, λ_b, that needs to be applied, is not accurately known (see *e.g.* [228, 229]), especially at higher redshifts, $z \gtrsim 5$, where the density field is more linear which makes it easier to simulate the IGM more accurately. The smoothing due to gas pressure [230] can be described in linear theory [227] and the smoothing scale, λ_p, depends on the thermal *history* of the gas; that history is not well constrained.

5.6.2 Experimental detection

The study of the FPS has been applied for many years to distinct releases of data from SDSS (BOSS). This dataset contains a few thousands quasar spectra: this allows to reconstruct the FPS very well (Fig. 5.12). However, spectral resolution is limited and the smallest scales resolved are $\Delta v \sim 50\mathrm{km/s}$. The accurate measurement of the FPS was used to constrain WDM in [233].

In this work we compare our simulation results to the same flux power spectrum (FPS) computed from a set of $z \gtrsim 4.5$ quasar spectra previously analysed by [150, 234–236], and [237] (Fig. 5.13). These data are based on 25 high-resolution quasar spectra with emission redshifts in the range $4.48 \leq z_{\mathrm{QSO}} \leq 6.42$ obtained with the HIRES spectrograph on KECK, and the Magellan Inamory Kyocera Echelle (MIKE) spectrograph on the Magellan Clay telescope. We do not analyse the original spectra – they are not yet publicly available – but simply compare to the published FPS. We note that for $z = 5.0$ MIKE dataset contains 4 QSOs with the emission redshifts $z > 4.8$ [238, 239], while the HIRES dataset consists of 16 QSOs [238–240]. At this redshift the interval $\Delta z = 0.4$ used for binning in [150] corresponds to $\sim 140\,\mathrm{Mpc}/h$. Taking into account quasar proximity zones these quasar spectra cover $\sim 240\,\mathrm{Mpc}/h$ (MIKE) and $1230\,\mathrm{Mpc}/h$ (HIRES) at $z = 5$ and $\sim 810\,\mathrm{Mpc}/h$
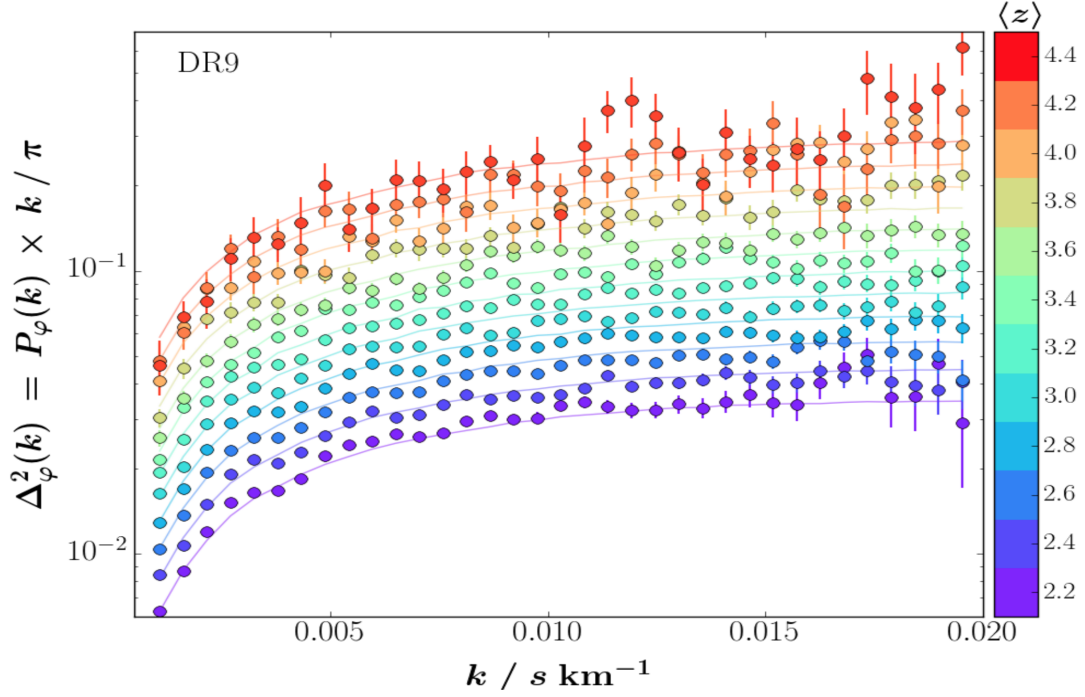
Figure 5.12: FPS measured by SDSS BOSS DR9 [231]. *Credit: Baur et al. [232]*

for HIRES at $z = 5.4$. From this we can already anticipate that the sample variance errors will be quite large for both datasets. We will use this information in Section 5.4 below when estimating errors due to this finite sampling.

The HIRES and MIKE spectra have a spectral resolution of 6.7 and 13.7 $\mathrm{km\,s^{-1}}$ full width at half maximum (FWHM), and pixel size of 2.1 and $5.0\,\mathrm{km\,s^{-1}}$, respectively. The median signal-to-noise ratios at the continuum level are in the range 10–20 per pixel [150]. We generate mock FPS with similar properties, as described below. The finite spectral resolution introduces another cutoff scale in the FPS, $\lambda_s \sim \mathrm{FWHM}$.

The ionisation level of the IGM is quantified by the *effective optical depth*, $\tau_{\mathrm{eff}} \equiv -\ln\langle F \rangle$, where $\langle F \rangle$ is the observed mean transmission, averaged over all line-of-sights. [150] report values of $\tau_{\mathrm{eff}}(z = 5.0) = 1.924$ and $\tau_{\mathrm{eff}}(z = 5.4) = 2.64$, without quoting associated uncertainties which can be quite large, stemming from the systematic errors in continuum fitting and statistical errors due to sample variance. We provide our own estimates of the statistical errors due to sample variance on $\langle F \rangle$ in Appendix 5.D. For details on the properties of the dataset, the associated noise level, and the way the FPS and its covariance matrix were estimated, we refer the reader to [150].

In our analysis we also use a more recent data on FPS measured by upgraded HIRES with only three redshift bins, see Fig. 5.14. For a description of this dataset see [151].
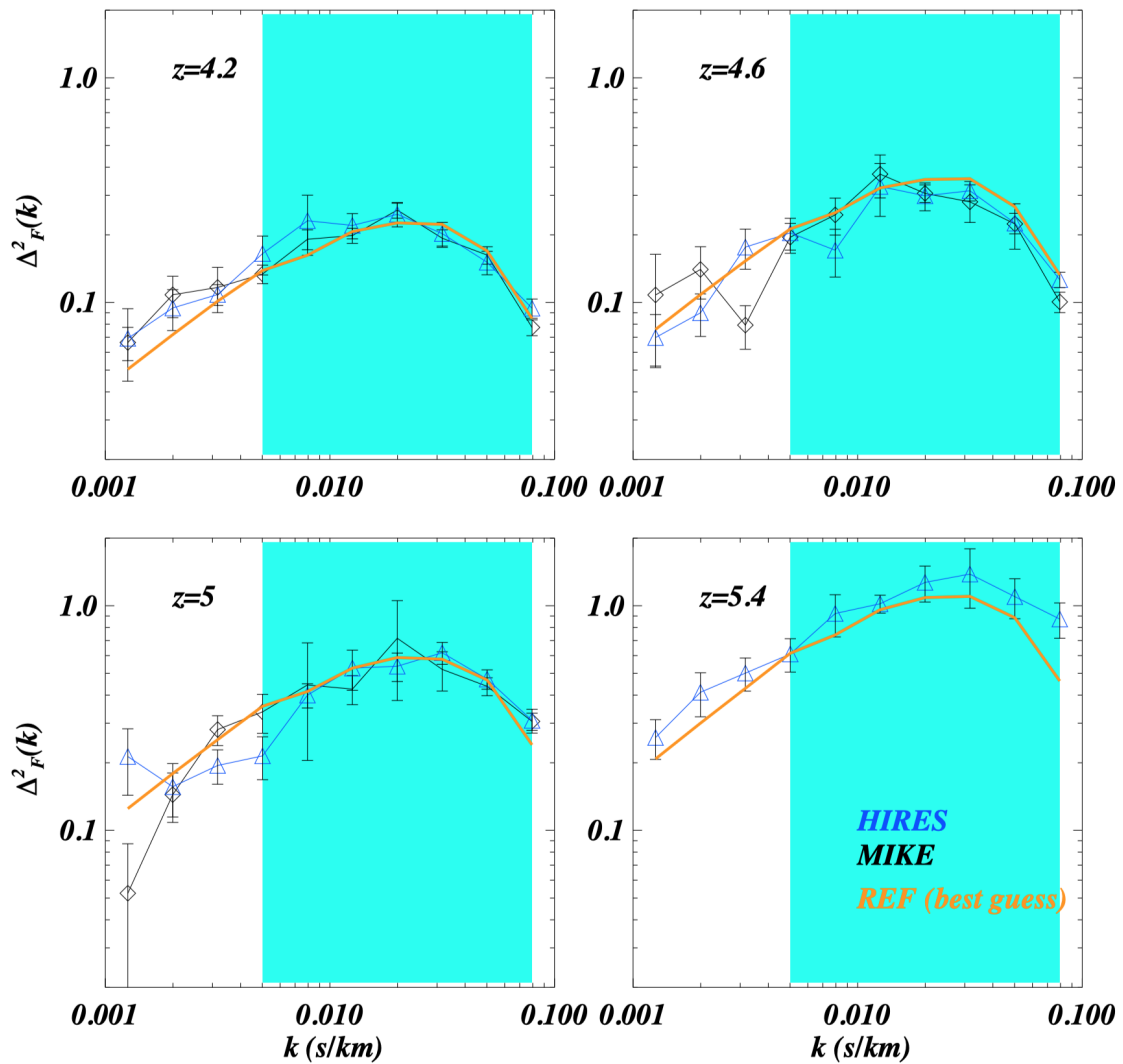
Figure 5.13: FPS measurements by Viel et al. [150]. Triangles and squares represent the actual data from MIKE and HIRES experiments, while the lines are models discussed in the paper. The highlighted regions show the range of wavenumbers used in analysis (large scale points are subject to continuum fitting systematic errors).

Using the high-resolution data we observe a cutoff in the FPS. We will discuss the possible interpretation and bounds on the WDM from it in the next section.
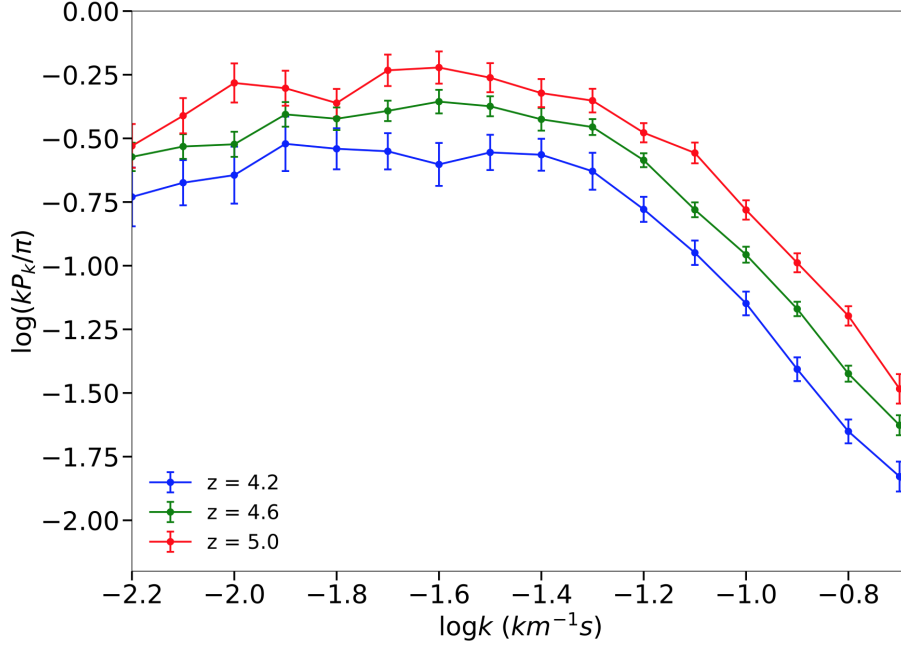
Figure 5.14: The most recent and high-resolution measurements of FPS by Boera et al. [151]

5.7   Interpretation and constraints

Before the availability of high-resolution quasar spectra, the constraints on WDM were put by constraining the deviations of the predicted flux power spectrum from the experimental data (e.g. SDSS BOSS on Fig. 5.12). The FPS in BOSS data grew with the wavenumber being consistent with CDM.

Constraints derived from the low-resolution SDSS BOSS data put a limit for thermal WDM relic at $m \gtrsim 3.5 \,\mathrm{keV}$ [241].

The high-resolution quasar spectra uncover a clear cutoff at the scales $\lambda \lesssim 30 \,\mathrm{km/s}$ (Fig. 5.14). It could be generated by the thermal effects alone – e.g. indicating the temperature of the IGM at the level $T \gtrsim 10000 \,\mathrm{K}$. These values are compatible with the typically considered thermal histories.

Does this mean that the cutoff is primarily due to thermal effects? As we saw before (Section 5.5.2) this kind of feature can be generated by thermal effects (even in CDM) or by the free-streaming of Dark Matter.

5.7.1   At what scale do we expect a cutoff in Cold Dark Matter?

5.7.1.1   Effect of peculiar velocities on Flux Power Spectrum

In CDM cosmologies the real-space MPS, $\Delta^2_{r,3d}(k)$, is a monotonically increasing function of k. However, in *velocity space* over which the FPS observable is built, an additional

effect – the *redshift space distortions* (RSD) – affect the shape [242–245]. RSD may erase small-scale power in the FPS because peculiar velocities of baryons are non-zero.

At linear level MPS in velocity space is related to real space by:

$$\Delta^2_{s,3d}(k) = \Delta^2_{r,3d}(k)(1 + \beta(\vec{k} \cdot \hat{z})^2)^2 \tag{5.7.1}$$

where $\hat{z}$ is the direction of observation and constant β for linear scales is given by expression $\delta_r = -\beta^{-1}\vec{\nabla} \cdot \vec{v}$ [242].

Real-space MPS projected along the line of sight is given by:

$$\Delta^2_{r,1d}(q) = \frac{q}{2\pi} \int d^2k_\perp \frac{\Delta^2_{r,3d}(q, k_\perp)}{\left(q^2 + k_\perp^2\right)^{3/2}} \tag{5.7.2}$$

$$= q \int_q^\infty \frac{dk}{k^2} \Delta^2_{r,3d}(k) \tag{5.7.3}$$

Clearly, in CDM linear $\Delta^2_{r,1d}(k)$ remains a monotonic function of k. Non-linear MPS experiences additional growth at small scales, therefore $\Delta^2_{r,1d}(k)$ does not exhibit a cutoff also at non-linear level.

Beyond the linear regime it is not possible to compute analytically the effect of RSD on the MPS. [246] have attempted to address this case, by considering a fitting formula calibrated to N-body simulations by [247]:

$$\Delta^2_{s,1d}(q) = q \int_q^\infty dk \frac{\Delta^2_{r,NL}(k)}{k^2} \left[1 + \beta \left(\frac{q}{k}\right)^2\right]^2 D\left[q\sigma_{12}(k)\right] \tag{5.7.4}$$

$$D\left[x\right] = \left[1 + \frac{1}{2}x^2 + \eta x^4\right]^{-1} \tag{5.7.5}$$

where $\sigma_{12}(k)$ is a pairwise velocity dispersion of dark matter particles, $\Delta^2_{s,NL}$ is a nonlinear $3d$ MPS and η is a constant. [246] predicted a cutoff on the scales similar to the cutoff observed in the HIRES and MIKE data.

In order to verify the predictions of [246], we have performed simulations where thermal effects were switched off, (Figure 5.15). Obviously, the simulation results for *e.g.* the IGM temperature are unrealistic in this case. The purpose of this exercise was to identify the position of a RSD-induced cutoff, which might have been obscured by thermal broadening. We find that the resolution of simulations by [247] stays significantly below the required resolution of our convergence analysis: number of particles $N = 128^3$ and box-size $L = 100 \, \mathrm{Mpc}/h$ [247] against $N = 1024^3$, $L = 20 \, \mathrm{Mpc}/h$. We conclude that the relevant scales have not been resolved in past simulations. To support this claim, we compare the FPS for various resolutions in model cosmologies designed to remove baryonic effects as much as possible, see Figure 5.15. Since our high-resolution simulations exhibit a cutoff at a position k's that is significantly larger than the reach of the data, we conclude
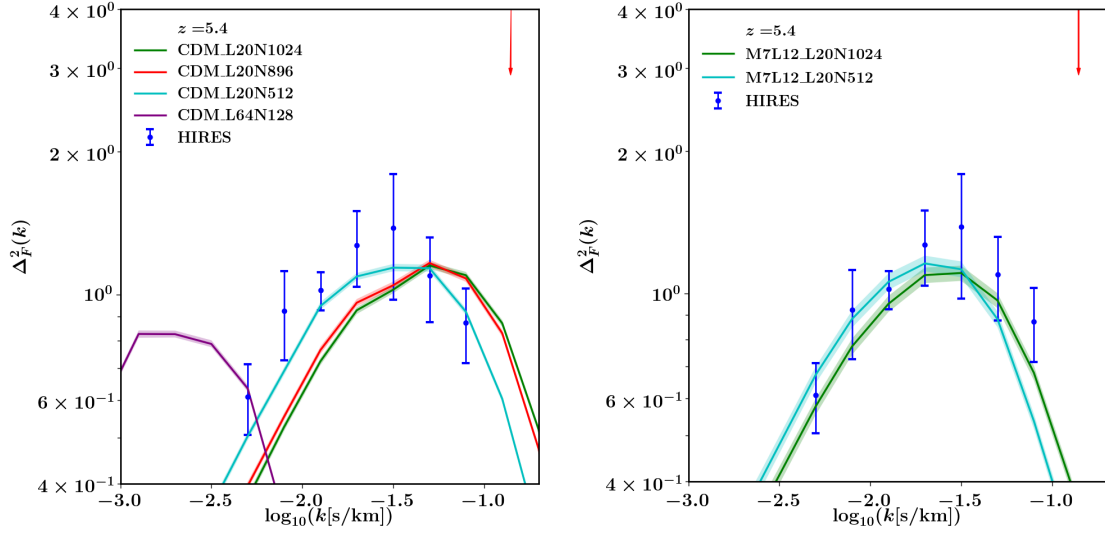
**Figure 5.15**: Effect of numerical resolution on the mock FPS for CDM (*left panel*) and WDM (*right panel*) of simulations performed without an imposed UVB. Both models are for the imposed power-law $T - \rho$ relation of Eq. (5.4.14) with $(T_0, \gamma) = (25\ \mathrm{K}, 1)$, are scaled to the observed value of the effective optical depth, $\tau_{\mathrm{eff}} = 3.0$ for $z = 5.4$, and mimic the spectral resolution and pixel size of the HIRES spectrograph on the KECK telescope (FWHM=6.7 km s^{-1}, pixel size=2.1 km sec^{-1} (see Section 5.6.2) but without adding noise. The data points show the error bars as reported by [150] that do not take into account sample variance (see below). The different colours correspond to different numbers of particles N, as per the legend. The observed FPS from [150] (blue) is plotted to indicate the range of relevant wave numbers. There is a numerical resolution-dependent cutoff in each simulation. Increasing the number of particles, the position of this cutoff shifts to larger k values. In our highest resolution simulations, $N = 1024^3$ DM and gas particles (green line), the resolution-dependent cutoff is outside the range of scales probed by the Lyman-α data, the corresponding Nyquist scale $k_{\mathrm{max,sim}}$ is outside the boundary of the plot. Therefore, we use such resolution in all subsequent simulations. The red arrow shows the scale associated with $k_{\mathrm{max,DM}}$. The figure also demonstrates that the simulations considered by [247] (*purple line*) lacked the necessary resolution to be used in [246].

that the role of RSD in the formation of the cutoff is negligible.

> Although the redshift space distortions produce a cutoff in FPS, the corresponding scale is $\lambda_z \ll 1\,\mathrm{km/s}$. This effect cannot be responsible for the cutoff in the data.

5.7.1.2 Cutoff due to thermal effects

The temperature T of a photoionised IGM depends on the density and on the spectral shape of the ionising radiation [248, 249]. Unlike the more familiar case of galactic H II regions, T is not set by a balance between photoheating and radiative cooling, but by the
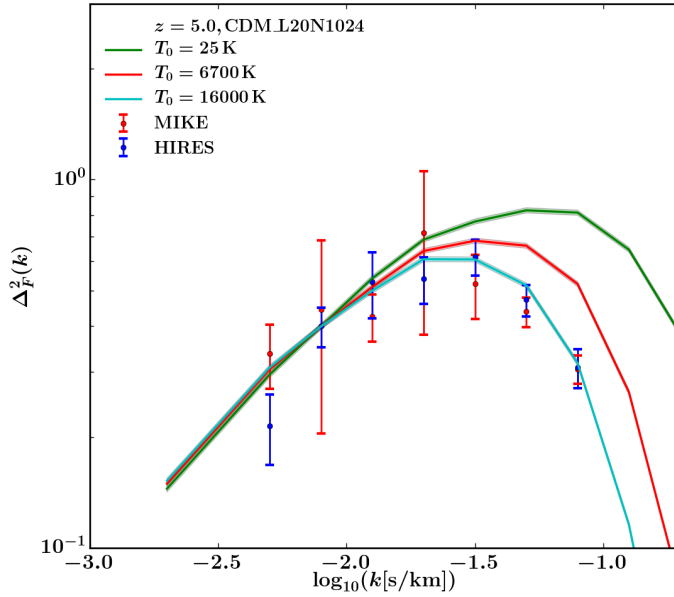
Figure 5.16: The cutoff in the mock flux power spectrum compared to the HIRES and MIKE data. We show a CDM model with three imposed temperature-density relations for $T_0 = 25\,\text{K}, 6700\,\text{K}$ and $16000\,\text{K}$.

mostly impulsive heating during reionisation and the adiabatic expansion of the Universe. Nevertheless, the temperature T_0 in the temperature-density relation of Eq. (5.4.14) is expected to be of the order $T_0 \sim 10^4$ K with $\gamma \approx 1$ close to reionisation. Once heated, pressure will smooth the gas distribution relative to the underlying Dark Matter introducing the filtering scale λ_p discussed previously, below which the amplitude of the density power spectrum is strongly suppressed. The patchiness of reionisation will therefore introduce large-scale fluctuations in the neutral fraction x, but also in the value of λ_p, as well as in that of the Doppler-broadening λ_b.

Doppler broadening cutoff. To describe the cutoff using the thermal Doppler broadening alone (Section 5.5.3) , we need the temperature to be $T \gtrsim 16000\,\text{K}$ (Fig. 5.16). While this temperature is not improbable, we don't have evidence for it at the moment.

On the other hand, how low the IGM temperature could be? When neutral gas is overrun with an ionisation front during reionisation, the difference between the energy of the ionising photon and the binding energy of H I, $\Delta E = h\nu - 13.6$ eV, heats the gas. In the case of H II regions, gas will also cool through line excitation and collisional cooling, resulting in a temperature immediately following reionisation of $T_{0,\text{reion}} \leqslant 1.5 \times 10^4$ K [248, 250]. In the case of reionisation, the low density of the IGM suppresses such in-front cooling, and the numerical calculations of [251] suggest $T_{0,\text{reion}} = 1 - 4 \times 10^4$ K, depending on the spectral slope of the ionising radiation. Following reionisation, the IGM cools adiabatically while being photoheated, preserving some memory of its reionisation history [219, 252]. Therefore the value of T_0 at $z = 5.4$ is set by $T_{0,\text{reion}}$, the redshift

$z_{\rm reion}$ when reionisation happened, and the shape of the ionising radiation that photoheats the gas subsequently. For T_0 to be sufficiently low then requires that $T_{0,\rm reion}$ is low, that $z_{\rm reion} \gg 5.4$, and that the ionising radiation is sufficiently soft.

Taking $z_{\rm reion} = 7.82$ from [100] and $T_{0,\rm reion} = 1.5 \times 10^4$ K yields a guesstimate for the lower limit of $T_0 \sim 0.8 \times 10^4$ K at $z = 5.4$, consistent with the value of $T_0 \sim 10^4$ K suggested by [253] that we used in the previous section. There is now good evidence that He II reionised at $z \sim 3.5$, much later than H I and He I [254–257], as the ionising background hardens due the increased contribution from quasars. This suggests that the ionising background during reionisation was unable to ionise He II significantly and hence was relatively soft. So conditions for low T_0 seem mostly satisfied.

Observationally, the IGM temperature is constrained to be at the level $T_0 \gtrsim 8000$ K at $z \lesssim 4.6$ [238, 255, 258, 259] (see e.g. [190] for a recent discussion). At $z \approx 6.0$ there is a single measurement in the near zone of a quasar that yields $5000 < T_0 < 10000$ K (68% CL, [260]). Fundamentally, all of the techniques used to infer T_0 observationally are based on identifying and computing the statistics of sharp features in Lyman-α forest spectra, and comparing these to simulated spectra. This implies that the T_0 inferred implicitly depends on $\lambda_{\rm DM}$.

Combining the theoretical prejudice and the measurements, we conclude that a value of $T_0 \sim 8000$ K or even colder at redshifts around $z = 5$ is not unreasonable and definitely not ruled out. Using Eq. (5.7.6), such a value of T_0 yields $k_{\rm max,b} = 0.12\,{\rm s/km}$. The maximum temperature still consistent with the cutoff is $T_0 = 16000$ K that yields $k_{\rm min,b} = 0.09\,{\rm s/km}$.

Pressure support cutoff. What do we know about pressure effects (Section 5.4.1) and the filtering length λ_p?

Using the analytical model by [186] for the evolution of the filtering length, we can attempt to bracket the pressure effects as well. In general, for the gas cooling in the expanding Universe $\lambda_F > \lambda_J$ and for rapid heating $\lambda_F < \lambda_J$.

Hence the absolute minimal pressure scale can be computed assuming infinitely fast reionization ($\lambda_p \sim 0.2\lambda_J$) and the maximum is given by the system with no reionization ($T \propto 1/a^2$, $\lambda_p \sim 2\lambda_J$).

Inhomogeneous reionization. The plausible patchiness of reionisation introduces complications. For example the large-scale amplitude of the FPS may be more a measure of the scale and amplitude of temperature fluctuations or of fluctuations in the mean neutral fraction, rather than being solely due to density fluctuations that we simulate. If that were the case, then our simulations should not match the measured FPS on large-scales, since

we have not included these effects (see e.g. [261]). Furthermore, what is the meaning of λ_b or λ_p in such a scenario, where these quantities are likely to vary spatially? Matching the cutoff in the FPS might pick-out in particular those regions where both λ_b and λ_p are unusually small.
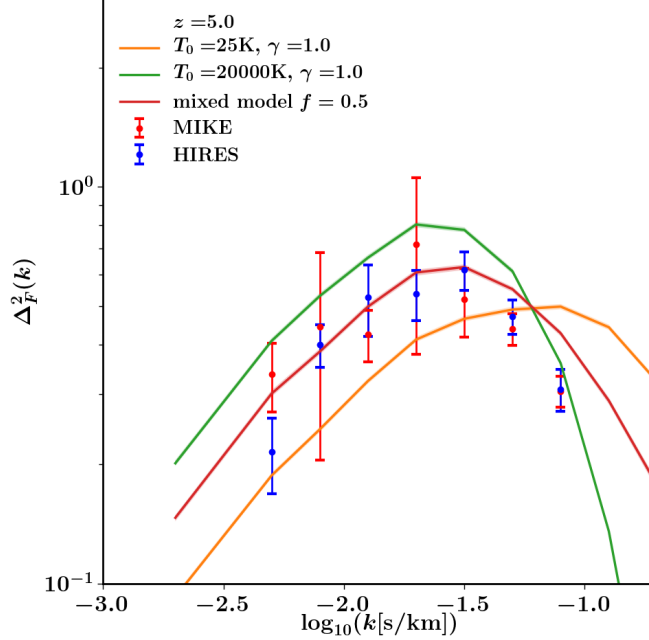
Figure 5.17: The effect of temperature fluctuations on FPS, for the case that bubbles are much larger than our simulated volume at $z = 5.0$. This case corresponds to a mixing fraction $f = 0.5$. For reference, we have drawn the data points of the MIKE and HIRES samples.

To illustrate the effect of fluctuations on the FPS, we contrast the FPS of two sets of mock spectra with an imposed temperature-density relation with different values of T_0: 25 K (i.e. negligible Doppler broadening and $T_0 = 2 \times 10^4$ K in Fig. 5.17, as well as a mock sample that uses half of the spectra from each of the two models. The FPS for the single-temperature models are normalized to have the same mean effective optical depth, $\tau_{\text{eff}} = 2.0$, the mixed-temperature model is computed from the two normalized single-temperature models, and it is not normalized further. We find that in the mixed model the FPS is intermediate between the FPS of the hot and cold models. Hence, if the hot model represents the recently reionized regions in the IGM and the cold model the patches that were reionized previously and then cooled down, the mixed model looks like a model that is colder than the regions in the IGM that were reionized more recently.

Fig. 5.17 illustrates that fluctuations essentially decouple the behaviour of the FPS at large and small scales. If this is the case of the real IGM, then what we determine to be T_0 from fitting the cutoff does not correspond to either the hot or the cold temperature.

We leave a more detailed investigation of patchiness on the FPS and how that impacts on constraints on λ_{FS} to future work.

5.7.1.3 Redshift dependence of the cutoff

Finally, we need to check the redshift dependence of the cutoff scales.

The simplest case is that of Doppler broadening. Consider a sharp feature in $F(v)$, smoothed by Doppler broadening due to gas being at temperature T. The width of the smoothed feature in velocity space will be of order $\Delta v_b = (2k_{\mathrm{B}}T/m_{\mathrm{H}})^{1/2}$ (where k_{B} is Boltzmann's constant and m_{H} the proton mass). In terms of the Fourier transform of $F(v)$, this will correspond to a feature at the proper wavenumber[4]

$$k_{\mathrm{max},b} = \frac{\sqrt{2}}{\Delta v_b} = 0.11 \left(\frac{T}{10^4\,\mathrm{K}} \right)^{-1/2} (\mathrm{km\ s^{-1}})^{-1} \,, \qquad (5.7.6)$$

which is independent of z, provided that T is constant.

How about pressure smoothing? The extent of the smoothing is approximately of order of the Jeans length [262], which in proper units is

$$\lambda_J = \sqrt{\frac{c_s^2 \pi}{\mathrm{G}\rho}} \,. \qquad (5.7.7)$$

Here, ρ is the total mass density (Dark Matter plus gas) of the absorber and c_s the sound speed. The corresponding velocity broadening is then $\Delta v_p = H(z)\lambda_J/(2\pi)$ [228]. At high enough redshift, the Hubble parameter scales like $\propto (1+z)^{3/2}$, and the density dependence of λ_J also scales like $\rho^{-1/2} \propto (1+z)^{3/2}$, making Δv_p also independent of redshift[5]. The corresponding value of k_{max} is

$$k_{\mathrm{max},p} = \frac{\sqrt{2}}{\Delta v_p} = 0.0760 \left(\frac{T}{10^4\,\mathrm{K}} \right)^{-1/2} \left(\mathrm{km\ s^{-1}}\right)^{-1} \,. \qquad (5.7.8)$$

The width of a feature due to Dark Matter free-streaming, λ_{FS}, is imprinted in the linear transfer function, and is therefore constant in co-moving (as opposed to proper) coordinates. The velocity extent of such a feature is therefore $\Delta v_\lambda = H(z)\lambda_{\mathrm{DM}}/(1+z) \propto (1+z)^{1/2}$ at high-enough z, and in the FPS scales like $k_{\mathrm{max},\mathrm{DM}} \propto \Delta v_\lambda^{-1} \propto (1+z)^{-1/2}$ and hence is *not* independent of z. We can write its value as

$$k_{\mathrm{max},\mathrm{DM}} = \frac{1+z}{H(z)}\frac{1}{\lambda_{\mathrm{DM}}} = 0.007 \left(\frac{\lambda_{\mathrm{DM}}}{h^{-1}\mathrm{cMpc}} \right)^{-1} \left(\frac{6}{1+z} \right)^{1/2} \left(\mathrm{km\ s^{-1}}\right)^{-1} \,. \quad (5.7.9)$$

---

[4]This is the case for Gaussian smoothing in the linear regime, with the factor 2 arising from the fact that the power spectrum is the square of the Fourier transform.

[5]We note that this no longer true at low redshift, where Δv_b and Δv_p scale differently with z.

The free-streaming scale λ_{FS} can be estimated as a position of the maximum of the linear matter power spectrum, see Fig. 5.2. For a particular case of 7 keV sterile neutrino that we will investigate in this work, this scale can be found e.g. in [136] as a function of lepton asymmetry. For the model with lepton asymmetry parameter $L_6 = 12$ [see 133, for the definition of L_6] the resulting scale is $\lambda_{\mathrm{DM}} \sim 0.07\,\mathrm{Mpc}/h$ which corresponds to $k_{\mathrm{DM,max}} \approx 0.1\,\mathrm{sec/km}$ at $z = 5$.

Finally, the finite resolution of the spectrograph imprints a feature that is constant in velocity space since the spectral resolution has a given value of $\mathcal{R} \equiv \Delta\lambda/\lambda = c/\Delta v_s$. The feature occurs at the redshift independent wavenumber

$$k_{\mathrm{max},s} = \frac{\sqrt{2}}{\Delta v_s} = 0.21 \left(\frac{6.6\,\mathrm{km\,s^{-1}}}{\Delta v_s}\right)^{-1} \left(\mathrm{km\,s^{-1}}\right)^{-1} . \qquad (5.7.10)$$

When simulating the above effects using a hydrodynamical simulation, yet another scale enters: the Nyquist frequency, set by the mean interparticle spacing. For a simulation with N^3 particles in a cubic volume with linear extent L, the corresponding scale is $\lambda_{\mathrm{sim}} = L/N^{1/3}$, and is constant in co-moving units. The corresponding k_{max} is of order

$$k_{\mathrm{max,sim}} = \frac{(1+z)}{H(z)} \frac{N^{1/3}}{L} \approx 0.27\,(\mathrm{km\,s^{-1}})^{-1} , \qquad (5.7.11)$$

where the numerical value is for $z = 5$, $L = 20\,h^{-1}\mathrm{Mpc}$ and $N = 512^3$, suggesting that the numerical resolution needs to be at least this good in order not to compromise the location of any cutoff in mock spectra.

The conclusion of this is that the effects of free-streaming, compared to those of thermal broadening, pressure smoothing, finite spectral or numerical resolution, scale differently with z. The redshift dependence is sufficiently weak so to make little difference between $z = 5.4$ and $z = 5$, but the difference does become important comparing the FPS at $z = 3$ versus $z = 5$. The numerical values also suggest that free-streaming, Doppler and pressure broadening set-in at very similar values of k, and that the finite spectral resolution of KECK is unlikely to compromise the measurements.

5.7.1.4 Summary about cutoff in Cold Dark Matter

We find that indeed there are thermal histories that have a cutoff at scales smaller than those of the observed cutoff ($\lambda \sim 30\,\mathrm{km/s}$). These are the so-called late reionization (LateR) models described by [210] shown in Fig. 5.18. Mock FPS computed from CDM simulations with LateR reionization history and temperature of the order $T_0 \sim 5000\,\mathrm{K}$ exhibits a cutoff at the scale $\lambda \sim 18\,\mathrm{km/s} < \lambda_F \sim 30\,\mathrm{km/s}$ (Figures 5.16 and 5.19).
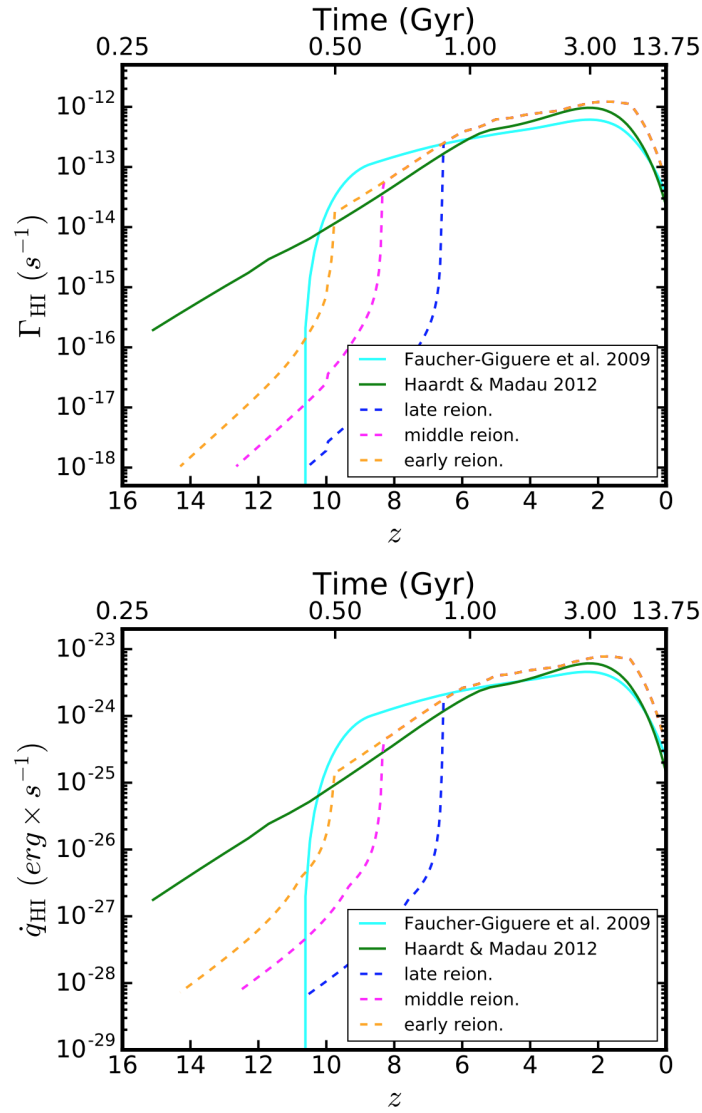
Figure 5.18: Photoionization rates Γ_{HI} and photoheating rates $\dot{q}_{HI}$ of some of the UVB models consistent with observations by [210] (*dashed lines*). Solid lines show analytical models by [263, 264] for comparison.

> There exist realistic thermal histories in which the cutoff due to thermal effects is located at smaller scales. If a true thermal history is such, the cutoff needs to be explained by a different mechanism.

5.7.2 Cutoff in Warm Dark Matter

Even in the absence of thermal effects, WDM free-streaming introduces a cutoff in the FPS which resembles the observed cutoff for sufficiently 'cold' WDM models. Sterile neutrino
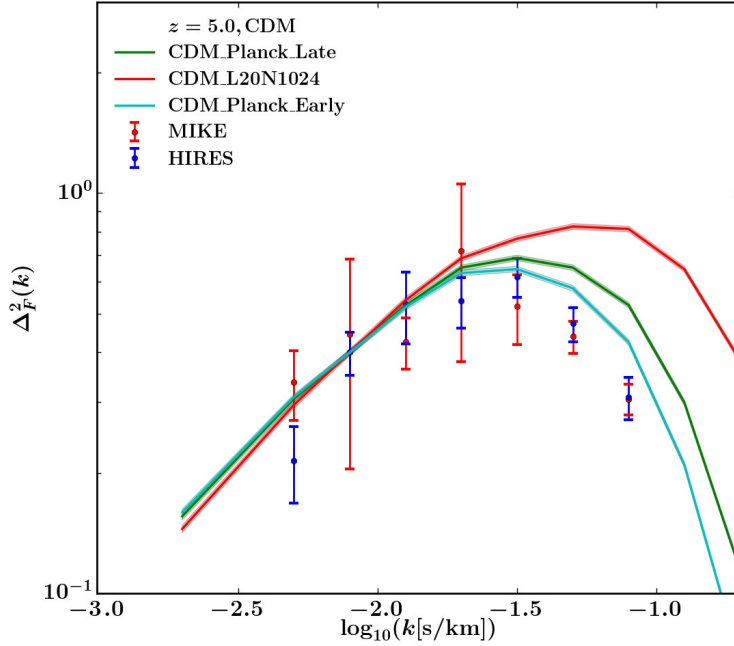
Figure 5.19: Mock FPS computed in simulations with late (*green line*), early (*cyan line*) and no reionization (*red line*). To demonstrate the effects of pressure, the temperature of the gas is set to $T_0 = 25\,\mathrm{K}$.

models with mass $7\,\mathrm{keV}$ and lepton asymmetry parameter $L_6 = 8$ or 12 (Fig. 5.20), appear consistent with the HIRES data.

But can WDM in combination with thermal effects produce the observed cutoff? As we saw in Section 5.7.1.3, the cutoffs due to different effects evolve differently. In addition to that, we expect both the temperature of the IGM and the filtering length to change with time. Since the data is available for 4 different redshifts, it is no longer practical to fit the data manually.

We have varied the parameters of our models to obtain the best fit to the cutoff in the FPS by performing a χ^2 analysis. To this end we use the evolution of the photo-ionisation and photo-heating rate of the *LateR* reionization model of [210], impose the temperature-density relation with $\gamma = 1$ in post-processing, and scale the simulated mean transmission to a range of values characterised by $\tau_{\mathrm{eff}} \equiv -\log\langle F \rangle$. As described in Section 5.5.5.1, we convolve the mock spectra with a Gaussian to mimic instrumental broadening, rebin to the pixel size of the spectrograph, and add Gaussian noise with standard deviation independent of wavelength and flux, corresponding to a signal to noise of 15 at the continuum level. We compute a grid of mock FPS, varying T_0 and τ_{eff} for CDM and WDM models. We compare the mock FPS to the observed FPS at redshifts $z = 4.2, 4.6, 5.0$ and 5.4. When doing the comparison we take into account that the scattering between different realisations is large due to the small size of QSO samples (see Section 5.6.2 for details). We take into account the sample variance by computing the χ^2 of a model using the covariance matrix computed from EAGLE_REF (as the boxsize of our reference simulation is not large
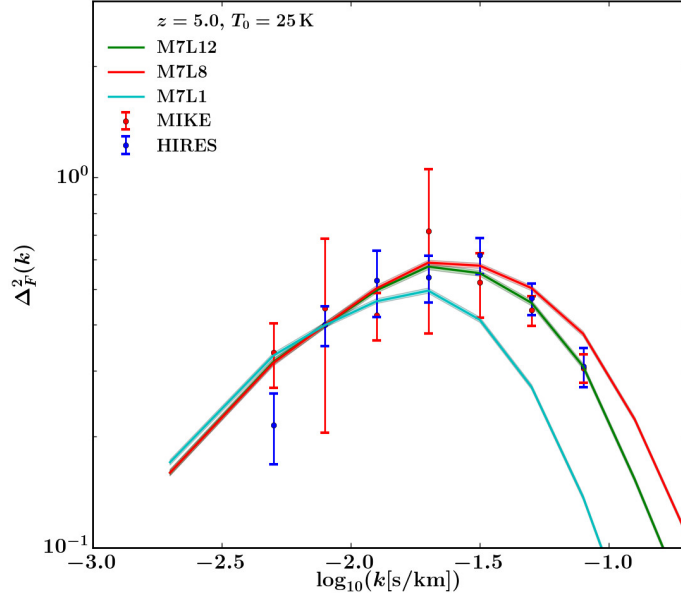
Figure 5.20: The cutoff in the mock flux power spectrum for various models, compared to the HIRES (*blue dots with error bars*) and MIKE data (*red dots with error bars*) at redshifts $z = 5.0$. For illustration purposes, we have scaled the amplitudes of the mock FPS in all cases such that it agrees with the HIRES value for the second point from the left, as a result different FPS in the same panel have different $\tau_{\rm eff}$. Here WDM simulations WDM_L1, WDM_L8 and WDM_L12 (*cyan, red and green curves, respectively*), with negligible Doppler broadening, $T_0 = 25$ K. DM free-streaming alone produces a cutoff in the FPS that resembles the observed cutoff for $L_6 = 8$ and 12.

enough to compute the covariance matrix). The rationale behind choosing EAGLE_REF was its large boxsize. the total length of the lines-of-sight in simulation was chosen equal to the total length of the observed QSO sample for each redshift range. Although EAGLE simulations do not have sufficient resolution at the smallest scales, we expect that the covariance is reproduced correctly.

The resulting contours for 68% and 95% confidence levels for HIRES data are shown in Fig. 5.21. In Table 5.3 we have compiled the values of the χ^2 for the best-fitting models.

As can be seen already from Fig. 5.20, the WDM model M7L12 has the FPS suppression due to the free-streaming that is consistent with the data. Therefore when varying T_0 in post-processing, WDM prefers temperatures with the scale $\lambda_b \ll \lambda_{\rm DM}$, see Fig. 5.21. At the same time, our simulation M7L12_Planck_Late predicts a temperature $T_0^{\rm sim} \simeq 6500$ K at all redshifts (this is also in agreement with findings of [210]). From Fig. 5.21 we see that the HIRES data is consistent with $T_0^{\rm sim}$ within its 95% confidence interval. Thus our procedure of post-processing is self-consistent – the temperature predicted by the simulations is consistent with the data. We show in Fig. 5.22 WDM model with this T_0 K as an example of a model with realistic thermal history, compatible with the data.

Table 5.3: Values of χ^2 for the best-fitting models shown in Figure 5.21. The number of dof is 5.

model	z	χ^2
CDM_Planck_Late	4.2	9.91
	4.6	4.61
	5.0	2.20
	5.4	3.25
M7L12_Planck_Late	4.2	6.04
	4.6	3.99
	5.0	3.44
	5.4	2.85

> We cannot distinguish CDM and WDM without precise knowledge of the true thermal history. Determination of the temperature history during the Epoch of Reionization will categorically exclude one of the models.
>
> Until there are observations allowing to identify the thermal history, the constraints being put on the Dark Matter candidates are model-dependent.

Marginalising over thermal histories.

We have demonstrated [1] that the cutoff in the FPS can be explained by any of the three effects (Doppler, pressure, DM free-streaming). This means that in order to obtain *robust* constraints on the properties of WDM particles one would need to marginalize over all astrophysical effects that can produce similar suppression in the FPS. The marginalisation over thermal histories (*i.e.* pressure effects) cannot be computed by brute-force: each thermal history requires its own full scale hydrodynamical simulation. On the other hand, marginalisation over Doppler broadening is possible in post-processing and is not computationally expensive.

> To put robust constraints on the WDM, we need to *marginalize* over thermal histories. However, dependence of pressure effects on the whole thermal evolution forbids the direct exploration of the parameter space. The in order to constrain the WDM, we aim to find the *boundary* WDM model that would be excluded under any thermal history.

Therefore, we suggest the following procedure for obtaining *robust Lyman-α forest bounds [2]:*

(1) We have modelled the IGM using cosmological numerical simulations, using GADGET-2 [45]. The details about the numerical simulations can be found in [1]. In order to minimise the pressure effects, we have run our simulations with model of ultraviolet
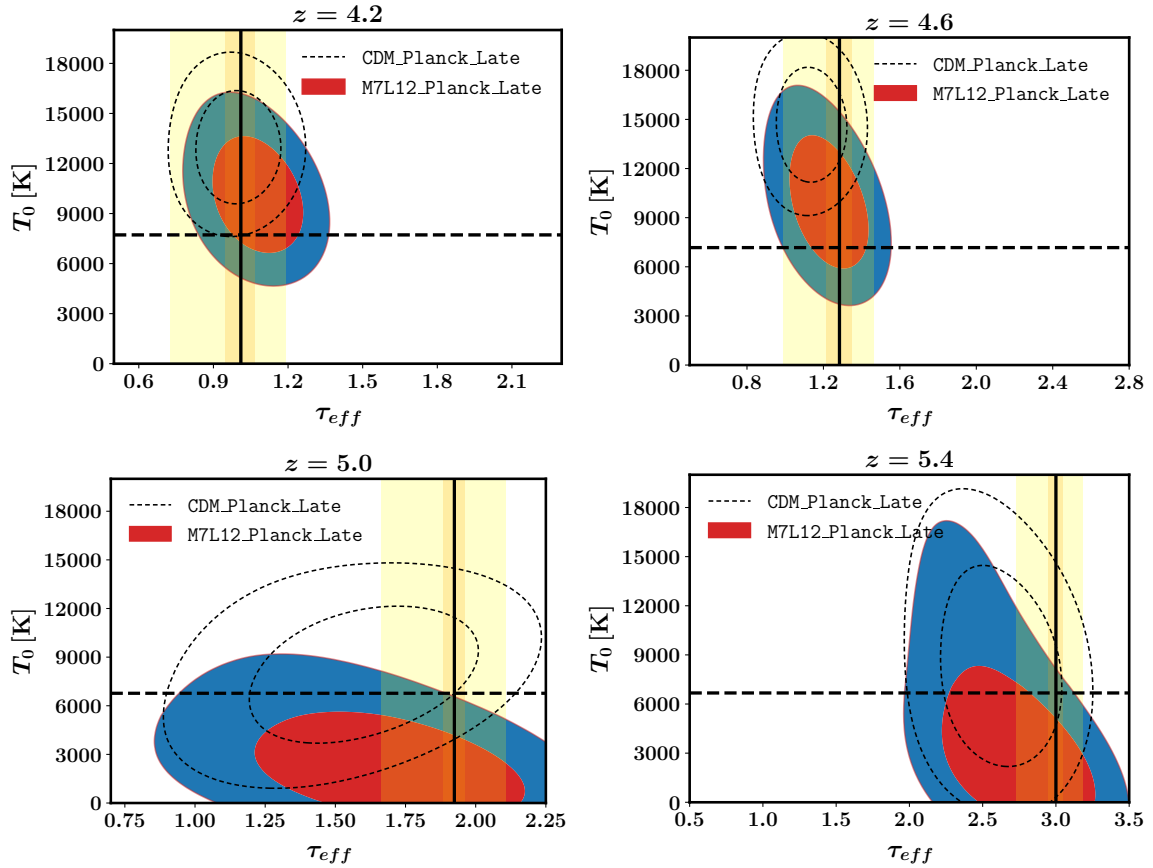
Figure 5.21: Confidence levels of mock FPS compared to the observed FPS of HIRES. We vary the temperature at the mean density, T_0, keeping $\gamma = 1$, and the value of the effective optical depth τ_{eff}. Solid lines and colour shaded areas correspond to 68% and $95\%$ uncertainty intervals for the $m_{\text{SN}} = 7\,\text{keV}$ and $L_6 = 12$ WDM model, dashed lines are the same for the CDM model. Both models used the late reionisation model *LateR* from [210]. The contours take into account both HIRES error bars as reported by [150] and additional errors due to finite number of quasars in the dataset. The black solid vertical line is the directly estimated τ_{eff} as reported in [150]. The horizontal line shows the value of T_0 as obtained in simulations with *LateR* UVB and without post-processing. It is in full agreement with the results of [210]. The systematic uncertainty on τ_{eff} coming from the sample variance is estimated to be $\sim 10\%$, and we have indicated the resulting uncertainty on τ_{eff} with the orange shade. The uncertainty on $\langle F \rangle$ due to continuum fitting is reported to be at the level $\sim 20\%$, and we have indicated the resulting uncertainty on τ_{eff} with the yellow shade.

background (UVB) LATECOLD from [210]. This model starts at redshift $z = 6.7$, later than other thermal histories, considered in [210] but reproduces the measured temperature at late redshift. This choice of the thermal history is compatible with the constraint on reionization time [210]. By construction it gives the minimal filaments size.[6]

---

[6]We thank J. Onorbe for sharing with us the data of the LATECOLD thermal history that were not published
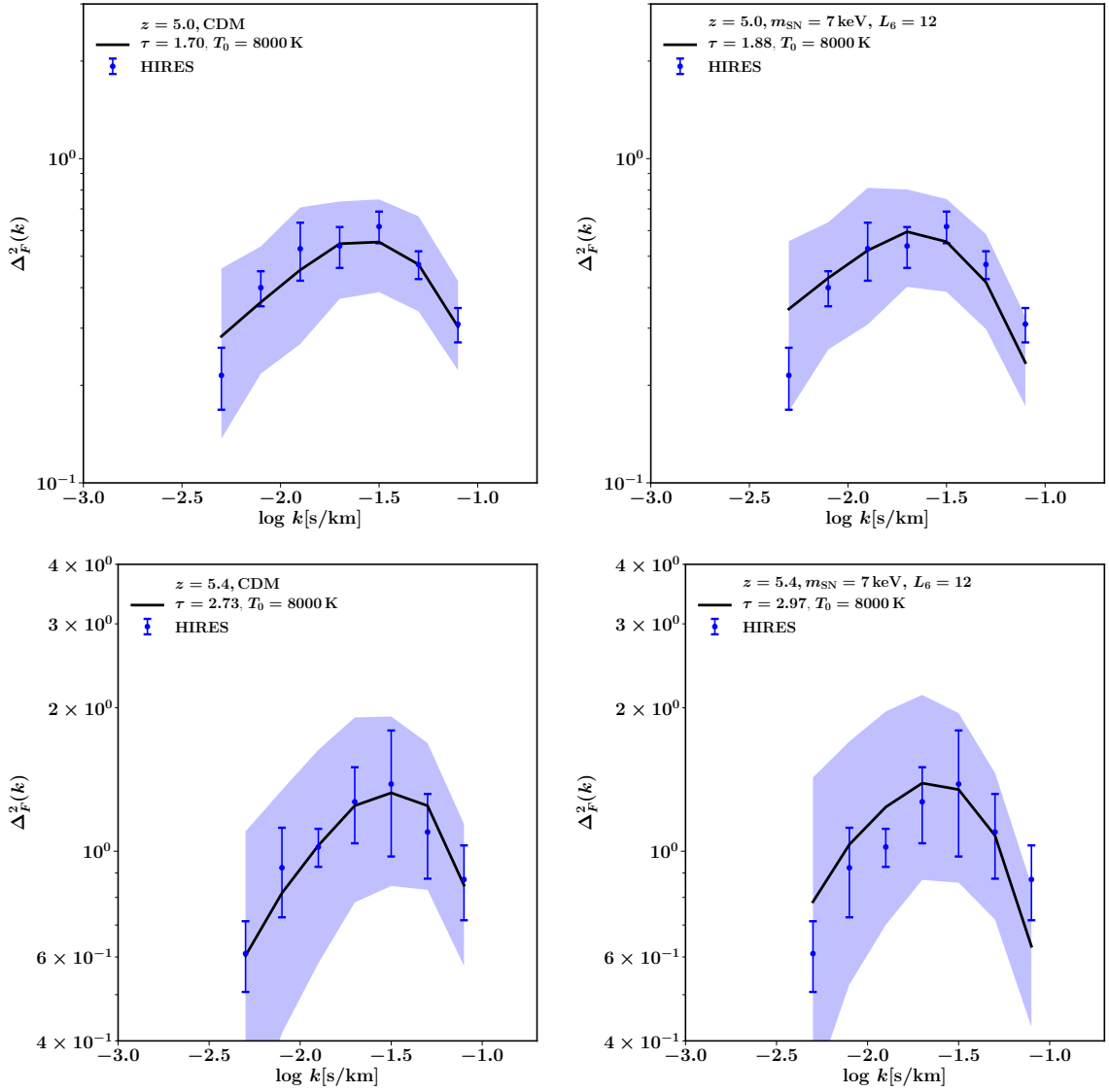
Figure 5.22: Examples of CDM and WDM models with realistic thermal histories, consistent with the high-resolution Lyman-α data. For both models we choose $T_0 = 8000$ K as predicted by our simulations with *LateR* UVB from [210]. The observed FPS inferred from HIRES is plotted as blue symbols with error bars as reported by [150]. One should keep in mind that the data points are correlated and therefore do not fluctuate independently. Shaded regions around the model show the variance due to different realisations of mock FPS (with the total length of the lines-of-sight in simulations equal to the length of observed spectra in the dataset for each redshift interval). The mock spectra have best-fit effective optical depth $\tau_{\rm eff} \equiv -\log\langle F \rangle$ for the fixed uniform temperature T imposed in post-processing. Top panels are for redshift $z = 5$ and bottom panels – for redshift $z = 5.4$ for CDM (*left panels*) and M7L12 SN model (*right panels*). The simulations are CDM_Planck_Late (left panels) and M7L12_Planck_Late (right panels).

---

together with the other thermal histories.

(2) We explicitly marginalise over the IGM temperature (*i.e.* Doppler broadening) adding their effects to the FPS in post-processing.

(3) We also marginalise over the *effective optical depth* $\tau_{\mathrm{eff}} = -\ln\langle F \rangle$. τ_{eff} encompasses the information about the average absorption level (*i.e.* overall level of ionization of the IGM). The work [151] provided measurements of τ_{eff} in each of the redshift bins together with the errorbars. We can vary τ_{eff} in the post-processing of the spectra, by rescaling the optical depth in the spectra by a suitable factor.

(4) Our resulting model describing evolution of FPS contains 7 parameters: WDM mass m_{WDM}, IGM temperature at cosmic mean density T_0 and τ_{eff} (the latter two quantities are evaluated at redshifts $z = 4.2, 4.6, 5.0$). We consider logarithmic priors on the temperature and on the mass of Warm Dark Matter. We perform cosmological hydrodynamical numerical simulations, using GADGET-2 [45], for two distinct cosmological models: Cold Dark Matter (CDM) and WDM constituted of pure thermal relics with masses of $m_{\mathrm{WDM}} = 2.0\,\mathrm{keV}$. The details about the numerical simulations can be found in [1]. Starting from our simulations, we compute in post-processing the FPS for the parameter arranged on a tri-dimensional regular grid. Our final theoretical model is obtained by interpolating linearly the FPS across the grid. We perform a joint analysis on all the redshift intervals.
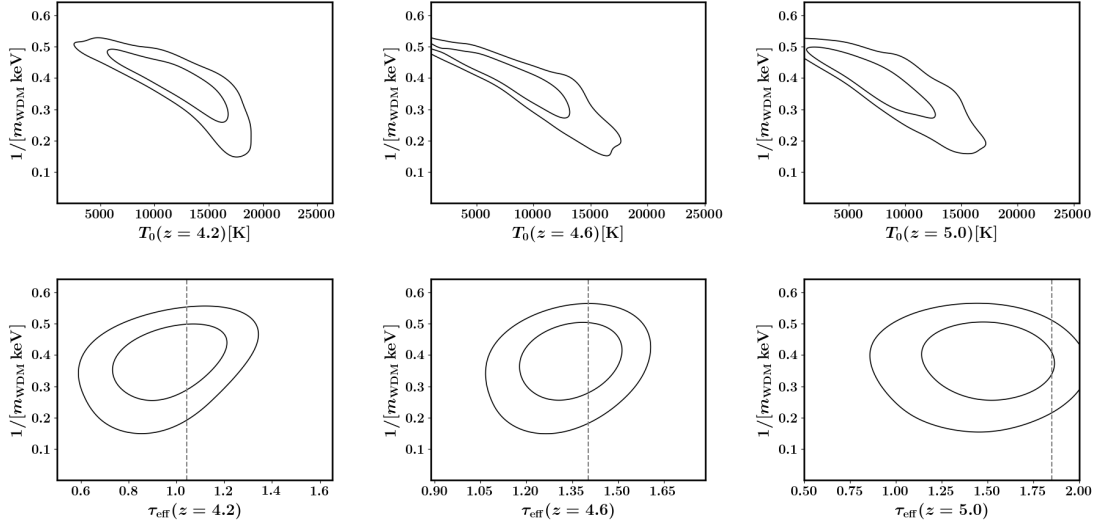
Figure 5.23: Confidence regions between the WDM mass, m_{WDM}, the IGM mean temperature, T_0 and the effective optical depth τ_{eff} at redshifts $z = 4.2, 4.6, 5.0$. Our analysis shows that if LATECOLD were the true history of reionisation, then the CDM would be ruled out. However, it is not possible to use this analysis to determine the IGM temperature in CDM for reionisation histories outside LATECOLD. By the same our analysis does not allow to determine the *lower* bounds on T_0 for a given WDM mass.

We run MCMC analysis and our 2σ contours are shown in Fig. 5.24 (full triangle) and Fig. 5.23 for the most relevant parameters. Two important comments are in order here:

1) Our contours do not reach CDM values ($1/m_{\mathrm{WDM}} = 0$). This does not mean that the CDM cosmology is excluded by the data. This simply indicates that in the LATECOLD reionizaiton scenario, the temperature alone is not sufficient to explain the suppression of the FPS.

2) Unlike the usual case, our temperature contours are only meaningful as the upper temperature limits. Indeed, they demonstrate that when minimizing pressure effects (as in LATECOLD) and switching off free streaming one is not able to explain the shape of the observed cutoff even with the temperatures in excess of $20000\,\mathrm{K}$. Best fit or higher temperature limits are exaggerated as they compensate pressure history that may be too low.

Results. We find that $m_{\mathrm{WDM}} \geq 2.0\,\mathrm{keV}$ at 2-σ level. Only the lower bounds for the temperature are shown. As mentioned above, the upper bound has no physical meaning, being an artifact of our procedure that fixed pressure effects at their lowest value. We show the 1 and 2-σ contours of the cosmic mean temperature T_0, the optical depth τ_{eff}, and the mass of the WDM m_{WDM} in Figure 5.24.
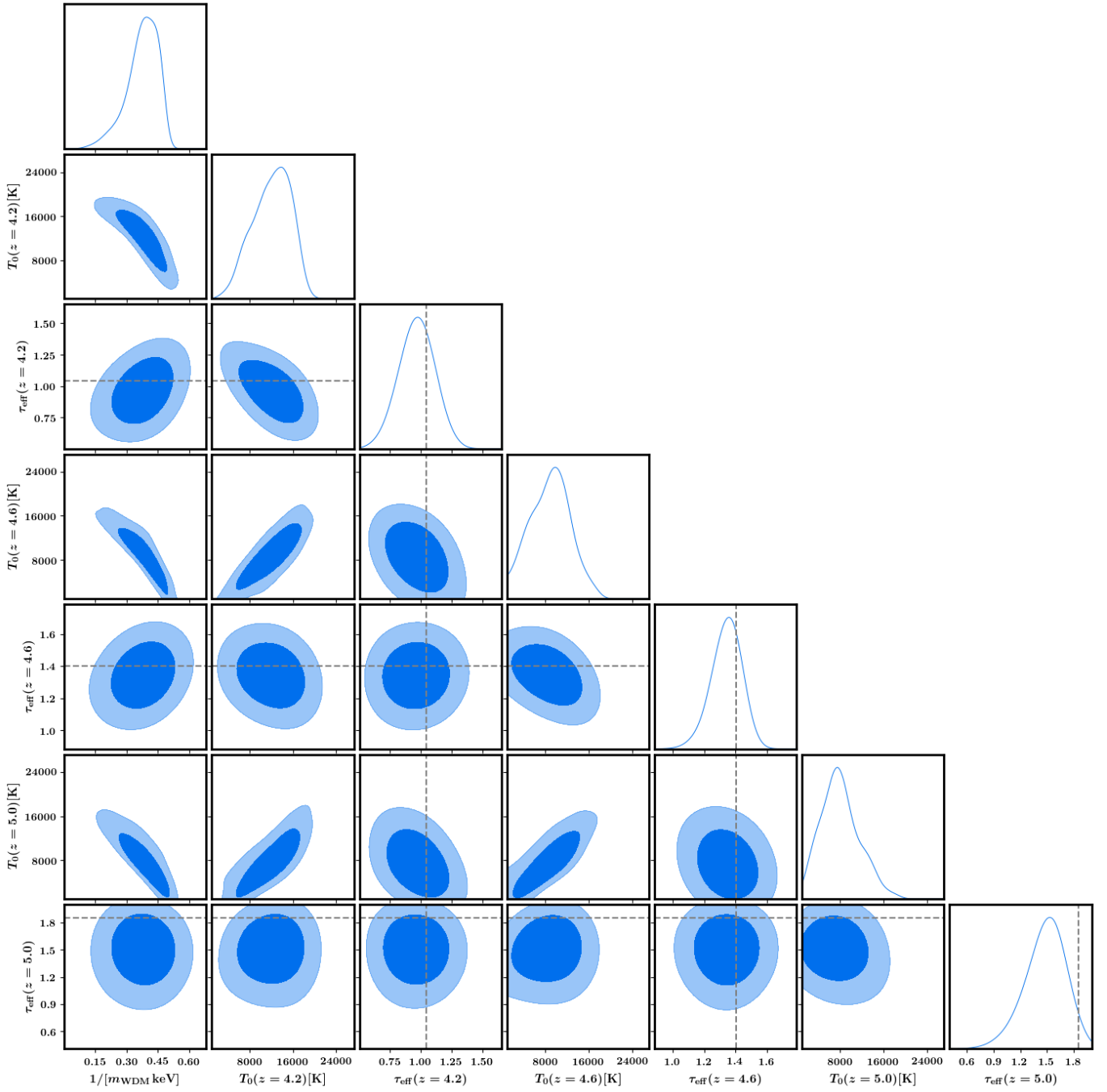
Figure 5.24: Confidence regions between the WDM mass m_{WDM} and the IGM mean temperature T_0 and effective optical depth at redshifts $z = 4.2, 4.6, 5.0$. The dashed grey lines are the τ_{eff} as measured in [151]
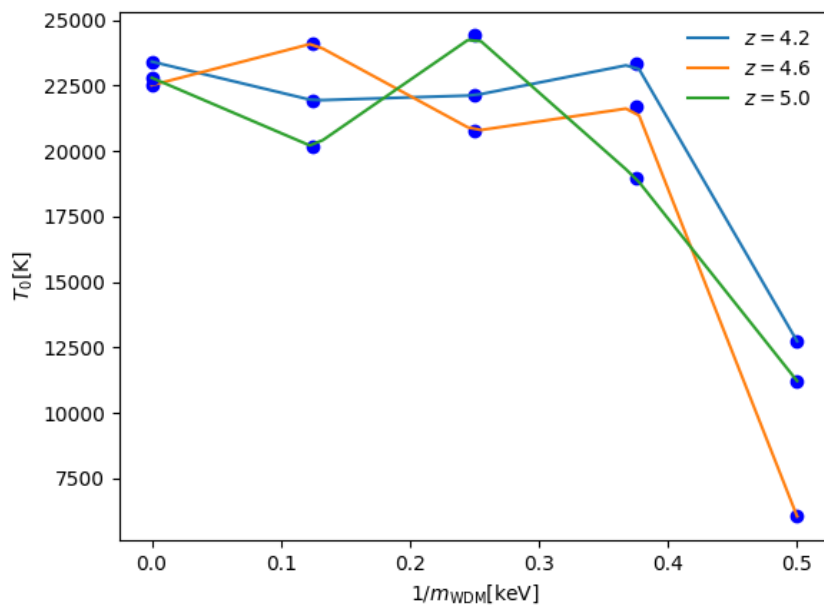
Figure 5.25: The 2-σ level upper limit on T_0 as a function of the inverse of m_{WDM}. The upper limit has been estimated separately for each redshift interval, with fixing m_{WDM}. We only show the interval $0 \leq m_{\mathrm{WDM}} \leq 0.5$, that is within the 2-$\sigma$ limit of the analysis of Figure 5.23. This result is in substantial agreement with the global fit that we have shown in Figure 5.23.
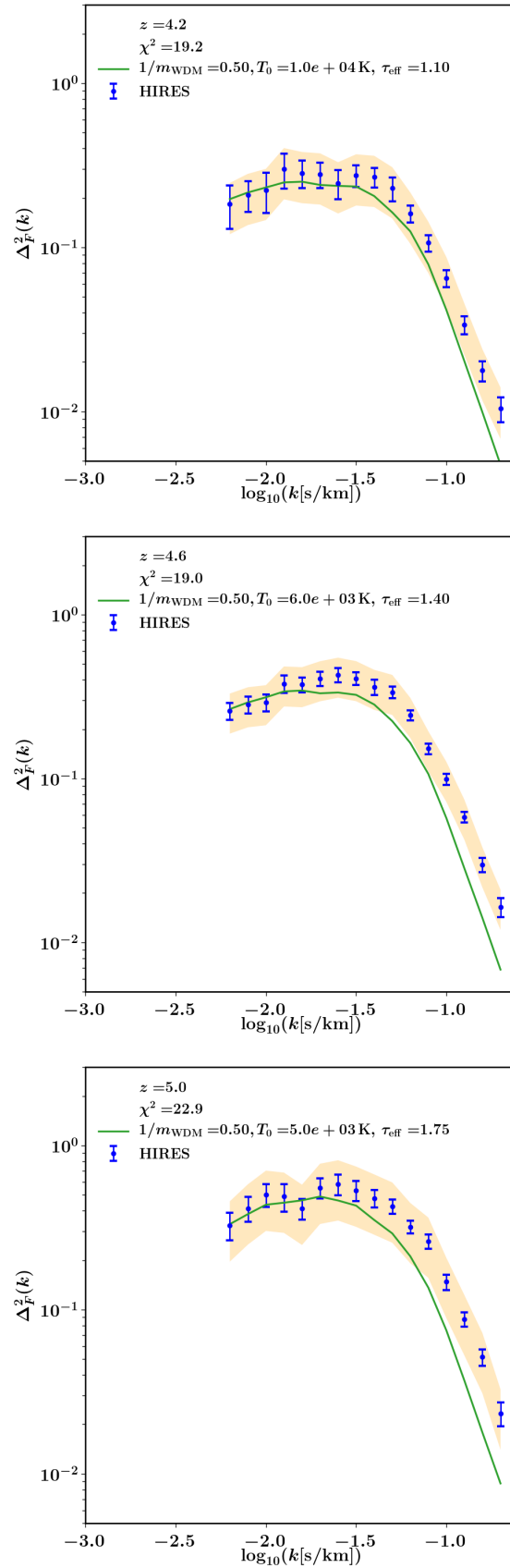
Figure 5.26: An example of cosmology with $m_{\mathrm{WDM}} = 2\,\mathrm{keV}$ and thermal history that is compatible within 2-σ with the observed flux power spectra. Orange shaded region shows the uncertainty in the measurement of FPS due to the sampling variance.

5.8 Conclusions

The power spectrum of the transmission in the Lyman-α forest (the flux power spectrum, FPS), exhibits a suppression on scales smaller than $\lambda_{\min} = 1/k_{\max} \sim 30$ km s^{-1}. Several physical effects may contribute to this observed cutoff: (*i*) Doppler broadening resulting from the finite temperature T_0 of the IGM, (*ii*) Jeans smoothing due to the finite pressure of the gas, and (*iii*) Dark Matter free streaming; these suppress power below scales λ_b, λ_p and λ_{FS}, respectively. We have shown in Section 5.7 that, when λ is expressed in velocity units, λ_b and λ_p are independent of redshift z for a given value of T_0, whereas $\lambda_{\mathrm{DM}} \propto (1 + z)^{1/2}$. This means that any smoothing of the density field due to WDM free-streaming will be most easily observable at high-redshift, and the observed FPS may provide constraints on the nature of the Dark Matter [150, 235–237], and possible be a 'WDM smoking gun'.

In our work we tried to answer two questions:

- Does the observed cutoff in the FPS favour Cold or Warm Dark Matter, or can both models provide acceptable fits to the existing data?

- Are the WDM models with large λ_{FS} that were previously excluded allowed if one considers a less restrictive thermal history?

To answer these questions we run a set of cosmological hydrodynamical simulations at very high resolution, varying λ_b, λ_p and λ_{FS} independently. We then compute mock spectra that mimic observational limitations (noise, finite spectral resolution and finite sample size), and compare the mock FPS to the observed FPS.

We demonstrate that all three effects (*i.e.* Doppler broadening, Jeans smoothing and DM free-streaming) yield a cutoff in the FPS that resembles the observed cutoff. Of course in reality all three effects will contribute at some level. In particular, Doppler broadening and Jeans smoothing both depend on the temperature T_0 of the IGM, and so always work together.

To answer the two questions posed above, we have tried to fit the observed FPS at redshifts $z = 5$ and 5.4 with (*i*) a CDM model (which has $\lambda_{\mathrm{DM}} = 0$), varying T_0 and the thermal history, and (*ii*) the particular case of a resonantly produced sterile neutrino WDM model (characterised by the mass of the particle, $m_{\mathrm{DM}}c^2 = 7$ keV, and the Lepton asymmetry parameter L_6, [153]), varying L_6, T_0 and the thermal history.

In addition to motivations based on particle physics (see *e.g.* [54]) our particular choice of WDM particle is motivated by the fact that: (*i*) its decay may have been observed as a 3.5 keV X-ray line in galaxies and clusters of galaxies [55–57], (*ii*) it produces galactic (sub)structures compatible with observations [136, 137], and (*iii*) it is apparently ruled out by the observed FPS [225].

Fig. 5.22 shows how the HIRES data is compatible with CDM and SN cosmologies if we choose relatively late reionisation model (*LateR* of [253]) so that λ_p is small and $T_0 \approx (7 - 8) \cdot 10^3$ K as predicted by this model. Both the assumed late reionisation redshift,

and the relatively low value of T_0, are reasonable and consistent with expectations and previous work. Crucially, a WDM model with $L_6 = 12$ and the same late redshift of reionisation also provides an acceptable fit to the data, provided $T_0 \leqslant 7000$ K. With such a low value of λ_b and λ_p, the FPS cutoff is mostly due to WDM free streaming.

From this comparison we conclude that the observed suppression in the FPS can be explained by thermal effects in CDM model but also by the free-streaming in a WDM model: current data do not strongly favour either possibility. We also find a reasonable fit for a WDM model that was previously ruled out by [150] and [235–237]. Our present analysis differs in a number of ways:

1. We vary the thermal history of the IGM within the allowed observational limits as discussed by [210, 253]. The previous works modeled the UVB according to [211]. The latter scenario is known to reionise the Universe too early with respect to current observations [210], plausibly overestimating λ_p.

2. We did not use any assumptions about the evolution $T_0(z)$ but inferred ranges of T_0 at $z = 5.0$ and $z = 5.4$ based on theoretical considerations and limits inferred from the Lyman-α data (see also [234]).

We also reconsidered the impact of peculiar velocities ('redshift space distortions'), which were claimed to affect the appearance of a cutoff at the smallest scales [246], but found these not to be important at the much higher resolution of our simulations.

We also demonstrated that spatial fluctuations in temperature, which are expected to be present close to reionisation, may dramatically affect the FPS. Spatial variations in T_0 can dramatically increase the amplitude of the FPS at the scale of the imposed fluctuations, effectively decoupling the large-scale and small-scale FPS. Unfortunately, this means that a model without fluctuations in T_0 will yield incorrect constraints on parameters if such fluctuations *are* present in the data. Interestingly, the nuisance caused by fluctuations in T_0 may actually be rather helpful if the cutoff in the FPS is in fact due to WDM, since in that case there would be no spatial fluctuations in the location of the cutoff – and the evolution with redshift of the cutoff would follow $\lambda_{\mathrm{DM}} \propto (1 + z)^{1/2}$.

Moving away from Lyman-α and studying the small-scale Universe in the H I 21-cm line during the 'Dark Ages' [265] instead is currently almost science fiction, but ultimately may be the most convincing way of determining once and for all whether most of the Dark Matter in the Universe is warm or cold.

In the meantime, we have produced a new robust constraint on WDM from the most recent QSO spectra available from HIRES at high redshift. We have assumed a conservative thermal history for the IGM that is still consistent with reionization, and that gives the minimal size of the intergalactic structures. We have assumed that the intergalactic medium is optically thin, and that reionization happens uniformly in all the space. We aim to discuss

a more realistic reionization scenario in an upcoming publication. We have neglected the effect of undetected metals that may increase the power on the smallest scales, implying a weaker cutoff.

Appendix

5.A   Convergence of the simulations in box-size
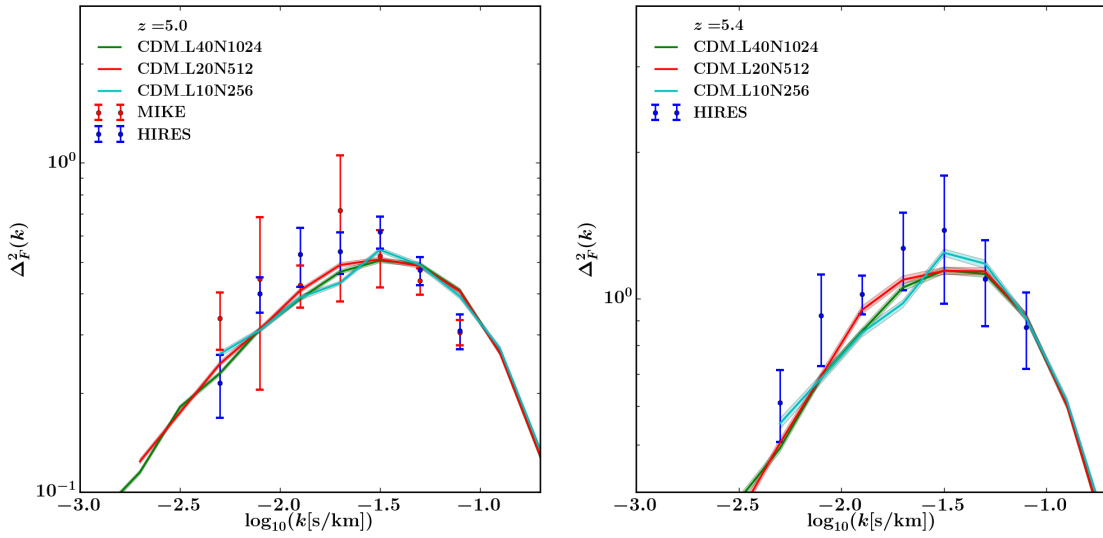
Figure 5.27: Study of the box-size needed in the numerical simulations to resolve the smallest scales probed by the HIRES and MIKE data samples. We show the FPS at $z = 5.0$ and $z = 5.4$ for three simulations without UVB and different box-sizes, yet same resolution. We have imposed a uniform temperature $T = 25\,\mathrm{K}$ in the post-processing of the spectra. We have applied the resolution of the HIRES spectrograph to the spectra, but we have excluded the effect of noise on the spectra. The FPS are normalized to the nominal observed optical depth of the observed spectra. The red solid line has a box-size $L = 10\,\mathrm{Mpc}/h$, the green solid line $L = 20\,\mathrm{Mpc}/h$, and the orange solid line $L = 40\,\mathrm{Mpc}/h$. The FPS for the case of $L = 10$, $L = 20\,\mathrm{Mpc}/h$, and $L = 40\,\mathrm{Mpc}/h$ agree with each other.

We have investigated the convergence of the FPS in box-size of the simulation with constant resolution. In section 5.B we have concluded that we need at least a number of particles $N = 1024^3$ and a boxsize $L = 20\,\mathrm{Mpc}/h$ to resolve the smallest scales reached by the data. Because we do not have the computing power to run a simulation with $L = 40\,\mathrm{Mpc}/h$ with this maximal resolution, we consider three simulations with $L = 10, 20, 40\,\mathrm{Mpc}/h$ and half the resolution. In this limit, we show in Figure 5.27 that the $L = 20\,\mathrm{Mpc}/h$ is sufficient to resolve the scales we intend to study.

162

5.B Numerical convergence

Before comparing the mock FPS to the observed FPS, we investigate to what extent the mock FPS is converged, both in terms of resolution and box size; the latter discussion can be found in Appendix 5.A. The gas temperature in our simulations that were performed without an imposed UVB is very low, and the gas distribution itself is not numerically converged at any of our resolutions. The effect of that on the FPS is shown in Fig. 5.15. For an imposed $T - \rho$ relation with $(T_0, \gamma) = (25\text{K}, 1)$, the CDM FPS does show a cutoff at small scales, but the value of $k_{\max}$ increases with increasing particle count, N. The value of $k_{\max}$ for $N = 896^3$ and $N = 1024^3$ is nearly identical (see Fig. 5.15). We run our main analysis with the box size $L = 20 \, \text{Mpc}/h$ and $N = 1024^3$ of both DM and gas particles, the corresponding scale $k_{\max,\text{sim}}$ is therefore much larger than k_s.

Our resolution is higher than used previously [150] as the latter work was interested in hotter thermal histories – IGM with the temperature $T_0 \sim 10000 - 20000 \, \text{K}$ with a non-negligible thermal smoothing. Note that [150] also recongnized that $N = 512^3$ with $L = 20 \, \text{Mpc}/h$ resolution is insufficient, but they applied a correcting factor to all power spectra. This factor was calibrated with a single simulation with $N = 896^3$, $L = 20 \, \text{Mpc}/h$. We instead rely on the intrinsic convergence of our simulations in the range of available data.

5.C Effect of noise

We investigate the effect of noise on the FPS. In our implementation of the noise, we have considered a Gaussian noise, with amplitude independent of flux or wavelength. In a spectrum from a bright quasar, the S/N is expected to increase with the flux. Because we have considered a S/N that is constant with flux and matches the S/N measured at the continuum level, we are likely underestimating the effect of noise in our analysis.

In some of the previous works on FPS in the Lyman-α forest, in particular [150, 235], the effect of noise on the flux PS is encoded with the application of a correction, that only depends on the chosen S/N and the redshift (in particular see Figure 16 in [150]), and not on other parameters of the IGM, such as the τ_{eff}. We have investigated whether the effect of noise is independent of the level of ionization of the IGM. In Figure 5.28 we show explicitly that the ratio between the FPS computed in the cases with and without noise depends on the value of the τ_{eff}, and the difference becomes larger on the smallest scales. This example was computed for a CDM simulation without, and with a uniform temperature $T = 2 \times 10^4 \, \text{K}$ imposed in post-processing. The effect of noise on the FPS is presumably being affected also by the temperature of examined spectra. Hence, we have resorted to including the effect of noise in our analysis by applying the noise to the spectra and then computing the resulting FPS.
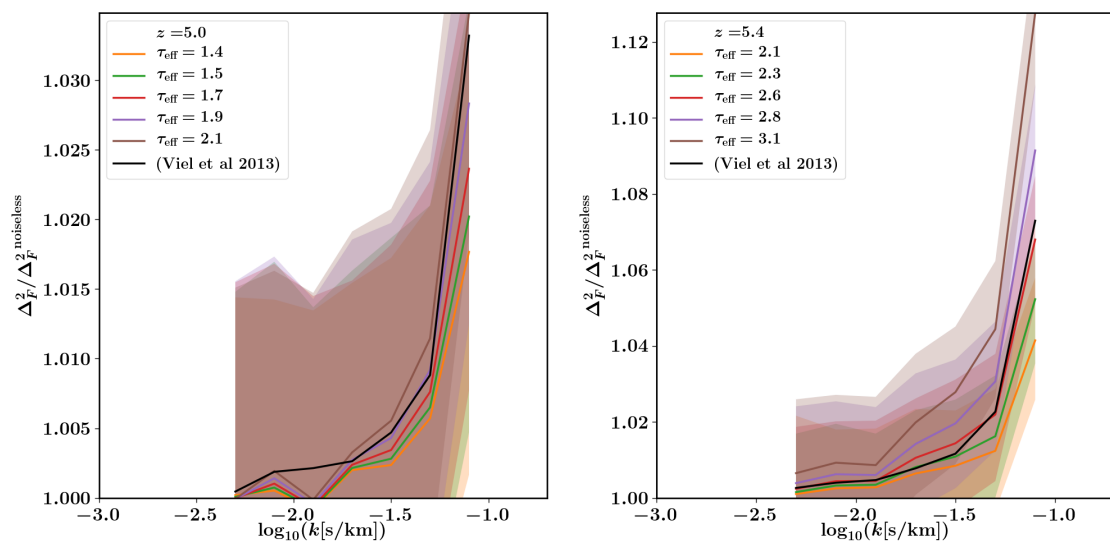
Figure 5.28: The effect of noise on FPS and its dependence from τ_{eff}. We show the ratio between the FPS computed with and without noise. We have considered a signal-to-noise ratio equal to 15, for a CDM simulation without UVB, and with imposed temperature in post-processing equal to $T = 2 \times 10^4$ K. The left (right) panel regards the redshift interval centered on $z = 5.0$ ($z = 5.4$). The solid lines refers to the mean of the ratio between the FPS computed with and without noise, the shaded region refer to the 1-σ uncertainty on the ratio. The black solid line is the correction for noise applied in [150], that is independent from τ_{eff}. We conclude that the effect of noise depends on τ_{eff}, and that accounting for noise only with a filter to the noiseless FPS is going to introduce a bias in the final estimate of the temperature.

5.D   Estimation of mean flux uncertainties

Available measurements of mean flux at high redshifts are based on small samples of quasars. Data from [150] that we are using contains only 25 quasars with emission redshifts $4.48 \leq z_{\text{em}} \leq 6.42$. Other works like [261] provide mean flux measurements also for only ~ 10 redshift intervals above $z = 5$. Even though quoted mean flux errors for individual spectra can be as low as $\sim 1\%$, tiny sample sizes suggest that undersampling of the density distribution is occurring.

To estimate this sampling error, we studied the distribution of mean flux for populations of mock spectra drawn from one of our simulations. To closely replicate the setup of [150], from 1000 lines of sight of the length $20\,\text{Mpc}/h$ we prepared 142 l.o.s. of $140\,\text{Mpc}/h$ by random concatenation (roughly corresponding to $\Delta z = 0.4$ used in [150] to bin the observations).

Next, we drew 1000 samples of the sizes 1, 10 and 100. For each population, we computed the standard as well as maximal deviations to gauge the sampling bias: Table 5.4. We see that typical error for $N_{sample} = 10$ is of the order of $4 - 5\%$.

On the other hand, the typical continuum level uncertainty is estimated to be $\sim 20\%$

[150]. Hence, uncertainty is dominated by continuum error.

Table 5.4: Means and standard deviations for populations of $140\,\mathrm{Mpc}/h$ mock spectra. $\bar{F}$ denotes averaging inside a population while angular brackets $\langle\rangle$ denote ensemble average.

z	N_{sample}	$\langle\bar{F}\rangle$	Standard Deviation
	1	0.1136	$\pm 0.0163\ (\pm 14.3\%)$
5.4	10	0.1121	$\pm 0.0056\ (\pm 5.0\%)$
	100	0.1121	$\pm 0.0008\ (\pm 0.7\%)$
	1	0.2086	$\pm 0.0247\ (\pm 11.8\%)$
5.0	10	0.2070	$\pm 0.0078\ (\pm 3.8\%)$
	100	0.2065	$\pm 0.0011\ (\pm 0.5\%)$

5.E Data analysis

Mock spectra are rescaled to match the corresponding experimental values of effective optical depth using the following procedure:

Algorithm 1 Rescaling of transmitted flux

1: **function** RESCALEFLUXES(spectra, τ_{exp})
2:     $\tau \leftarrow -\log\langle spectra\rangle$
3:     **while** $|\tau - \tau_{exp}| > 10^{-2}$ **do**
4:         $x \leftarrow \tau_{exp}/\tau$
5:         $spectra \leftarrow spectra^{x}$
6:         $\tau \leftarrow -\log\langle spectra\rangle$
    return spectra

This algorithm is a linear rescaling of the optical depth $\tau_{eff} = -\log\langle F\rangle$ that is complicated by exponentiation: $\langle e^{\tau}\rangle \neq e^{\langle\tau\rangle}$.

Next, we apply Gaussian filtering to simulate the point-spread function of the instruments and rebin the spectra according to pixel sizes.

From rescaled and filtered spectra the flux contrast is computed $\delta_F(v) = \frac{F(v)}{\langle F\rangle} - 1$ and transformed into momentum space using FFT. Flux power spectrum is computed as follows:

$$\Delta_F^2(k) = \frac{1}{\pi}kP_F(k) \tag{5.E.1}$$

$$P_F(k) = V\left\langle|\tilde{\delta}(k)|^2\right\rangle \tag{5.E.2}$$

$$\tilde{\delta}(k) = \frac{1}{V}\int dv\, e^{-ikv}\delta(v) \tag{5.E.3}$$

$$\delta(v) = \frac{F(v) - \langle F\rangle}{\langle F\rangle} \tag{5.E.4}$$

where V is the velocity interval over which the flux is given.

Flux is sampled in a linearly spaced velocity interval (v_{min}, v_{max}) with a step dv. To get the Fourier transform of the flux, we will use a Discrete Fourier Transform implemented by Fast Fourier Transform algorithm.

$$\tilde{\delta}(k) = \frac{1}{V} \int dv e^{-ikv} \delta(v) \tag{5.E.5}$$

$$\tilde{\delta}(k) \approx \frac{1}{V} \sum_{m=1}^{N} dv e^{-ik(v_{min}+mdv)} \delta(v_{min} + mdv) \tag{5.E.6}$$

$$= \frac{1}{V} dv e^{-ikv_{min}} \sum_{m=1}^{N} e^{-ikmdv} \delta(v_{min} + mdv) \tag{5.E.7}$$

Values of k are as well linearly sampled with a step $dk = \frac{2\pi}{Ndv} = \frac{2\pi}{V}$

$$\tilde{\delta}(ndk) \approx \frac{1}{V} dv e^{-in\,dk\,v_{min}} \sum_{m=1}^{N} e^{-in\,dk\,m\,dv} \delta(v_{min} + mdv) \tag{5.E.8}$$

$$= \frac{1}{V} dv e^{-in\,dk\,v_{min}} \sum_{m=1}^{N} e^{-2\pi i \frac{nm}{N}} \delta(v_{min} + mdv) \tag{5.E.9}$$

The sum is exactly the Discrete Fourier Transform while the rest of the expression is the phase factor by which one has to multiply the output of the FFT to obtain the Fourier image.

Since the implementation of this is non-trivial and specific to a programming language, we present the Python code we use in Algorithm 2.

For computed power spectra we estimate the error bars on the mean using the bootstrap technique.

Finally, power spectra are averaged over logarithmically-spaced k-bins corresponding to the binning in data from [150]. For visual clarity we extend the number of bins to include also smaller scales.

Algorithm 2 Computation of flux power spectrum (Python)

```python
# Given:
# spectra - array of spectra
# N - length of each spectrum
# V - velocity interval
import numpy as np
dv = V / N

freqs = np.fft.fftfreq(N) * (2. * np.pi / dv)
idx = np.argsort(freqs)
idx = idx[freqs[idx] >= 0]
freqs = freqs[idx] # select positive frequencies

mean = np.mean(spectra)

delta = (spectra - mean) / mean
fourier = np.fft.fft(delta)[idx] * dv / V
power_spectra = np.abs(fourier)**2 * V
dimless_power_spectra = power_spectra * freqs / np.pi
```

Bibliography

[1] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *Phys. Rev.* **D98** (2018) 030001.

[2] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett.* **B716** (2012) 1 [`1207.7214`].

[3] CMS collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, *Phys. Lett.* **B716** (2012) 30 [`1207.7235`].

[4] ATLAS collaboration, *Combined measurements of Higgs boson production and decay using up to* $80\,fb^{-1}$ *of proton-proton collision data at* $\sqrt{s} = 13$ *TeV collected with the ATLAS experiment*, [`1909.02845`].

[5] F. Bezrukov and M. Shaposhnikov, *Why should we care about the top quark Yukawa coupling?*, *J. Exp. Theor. Phys.* **120** (2015) 335 [`1411.1923`].

[6] CDF collaboration, *The CDF Detector: An Overview*, *Nucl. Instrum. Meth.* **A271** (1988) 387.

[7] D0 collaboration, *The D0 Detector*, *Nucl. Instrum. Meth.* **A338** (1994) 185.

[8] L. Evans and P. Bryant, *LHC Machine*, *Journal of Instrumentation* **3** (2008) S08001.

[9] F. Zwicky, *Die Rotverschiebung von extragalaktischen Nebeln*, *Helv. Phys. Acta* **6** (1933) 110.

[10] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [`1502.01589`].

[11] Y. Sofue and V. Rubin, *Rotation curves of spiral galaxies*, *Ann. Rev. Astron. Astrophys.* **39** (2001) 137 [`astro-ph/0010594`].

[12] W. J. G. de Blok, F. Walter, E. Brinks, C. Trachternach, S.-H. Oh and R. C. Kennicutt, Jr., *High-Resolution Rotation Curves and Galaxy Mass Models from THINGS*, *Astron. J.* **136** (2008) 2648 [`0810.2100`].

[13] F. Lelli, S. S. McGaugh and J. M. Schombert, *SPARC: Mass Models for 175 Disk Galaxies with Spitzer Photometry and Accurate Rotation Curves*, *arXiv* (2016) [`1606.09251`].

[14] M. Korsaga, B. Epinat, P. Amram, C. Carignan, P. Adamczyk and A. Sorgho, *GHASP: an Hα kinematical survey of spiral galaxies – XIII. Distribution of luminous and dark matter in spiral and irregular nearby galaxies using Hα and HI rotation curves and WISE photometry*, [`1909.08408`].

168

[15] J. T. Kleyna, M. I. Wilkinson, N. W. Evans and G. Gilmore, *Ursa Major: A Missing low-mass CDM halo?*, *Astrophys. J.* **630** (2005) L141 [`astro-ph/0507154`].

[16] M. G. Walker, M. Mateo, E. W. Olszewski, J. Penarrubia, N. W. Evans and G. Gilmore, *A Universal Mass Profile for Dwarf Spheroidal Galaxies*, *Astrophys. J.* **704** (2009) 1274 [`0906.0341`].

[17] S. Giodini, L. Lovisari, E. Pointecouteau, S. Ettori, T. H. Reiprich and H. Hoekstra, *Scaling relations for galaxy clusters: properties and evolution*, *Space Sci. Rev.* **177** (2013) 247 [`1305.3286`].

[18] J. Silk et al., *Particle Dark Matter: Observations, Models and Searches*. Cambridge Univ. Press, Cambridge, 2010, 10.1017/CBO9780511770739, http://www.cambridge.org/uk/catalogue/catalogue.asp?isbn=9780521763684.

[19] C. S. Frenk, A. E. Evrard, S. D. M. White and F. J. Summers, *Galaxy dynamics in clusters*, *Astrophys. J.* **472** (1996) 460.

[20] H. Hoekstra, *Weak gravitational lensing*, *Proc. Int. Sch. Phys. Fermi* **186** (2014) 59 [`1312.5981`].

[21] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, [`1807.06209`].

[22] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu et al., *Properties of galaxies reproduced by a hydrodynamic simulation*, Nature **509** (2014) 177.

[23] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu et al., *Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 1518.

[24] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns et al., *The EAGLE project: simulating the evolution and assembly of galaxies and their environments*, *Mon. Not. Roy. Astron. Soc.* **446** (2015) 521.

[25] V. Springel, R. Pakmor, A. Pillepich, R. Weinberger, D. Nelson, L. Hernquist et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676.

[26] T. D. Brandt, *Constraints on MACHO Dark Matter from Compact Stellar Systems in Ultra-Faint Dwarf Galaxies*, *Astrophys. J.* **824** (2016) L31 [`1605.03665`].

[27] H. Niikura et al., *Microlensing constraints on primordial black holes with the Subaru/HSC Andromeda observation*, [`1701.02151`].

[28] S. Tremaine and J. E. Gunn, *Dynamical Role of Light Neutral Leptons in Cosmology*, *Phys. Rev. Lett.* **42** (1979) 407.

[29] J. Miralda-Escude, *A test of the collisional dark matter hypothesis from cluster lensing*, *Astrophys. J.* **564** (2002) 60 [`astro-ph/0002050`].

[30] D. Gorbunov, A. Khmelnitsky and V. Rubakov, *Is gravitino still a warm dark matter candidate?*, *JHEP* **12** (2008) 055 [`0805.2836`].

[31] A. Boyarsky, O. Ruchayskiy and D. Iakubovskyi, *A Lower bound on the mass of Dark Matter particles*, *JCAP* **0903** (2009) 005 [`0808.3902`].

[32] V. Domcke and A. Urbano, *Dwarf spheroidal galaxies as degenerate gas of free fermions*, *JCAP* **1501** (2015) 002 [`1409.3167`].

[33] S. D. M. White, C. S. Frenk and M. Davis, *Clustering in a neutrino-dominated universe*, *Astrophys. J.* **274** (1983) L1.

[34] B. W. Lee and S. Weinberg, *Cosmological Lower Bound on Heavy Neutrino Masses*, *Phys. Rev. Lett.* **39** (1977) 165.

[35] T. Marrodn Undagoitia and L. Rauch, *Dark matter direct-detection experiments*, *J. Phys.* **G43** (2016) 013001 [`1509.08767`].

[36] M. Battaglieri et al., *US Cosmic Visions: New Ideas in Dark Matter 2017: Community Report*, in *U.S. Cosmic Visions: New Ideas in Dark Matter College Park, MD, USA, March 23-25, 2017*, 2017, [`1707.04591`], http://lss.fnal.gov/archive/2017/conf/fermilab-conf-17-282-ae-ppd-t.pdf.

[37] T. Lin, *Dark matter models and direct detection*, *PoS* **333** (2019) 009 [`1904.07915`].

[38] G. Brooijmans et al., *Les Houches 2011: Physics at TeV Colliders New Physics Working Group Report*, in *Proceedings, 7th Les Houches Workshop on Physics at TeV Colliders: Les Houches, France, May 30-June 17, 2011*, pp. 221–463, 2012, [`1203.1488`], http://lss.fnal.gov/archive/preprint/fermilab-conf-12-924-t.shtml.

[39] WORKING GROUP 3 collaboration, *Beyond the Standard Model Physics at the HL-LHC and HE-LHC*, [`1812.07831`].

[40] J. Beacham et al., *Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report*, [`1901.09966`].

[41] J. Alimena et al., *Searching for long-lived particles beyond the Standard Model at the Large Hadron Collider*, [`1903.04497`].

[42] J. Preskill, M. B. Wise and F. Wilczek, *Cosmology of the Invisible Axion*, *Phys. Lett.* **B120** (1983) 127.

[43] G. D. Martinez, J. S. Bullock, M. Kaplinghat, L. E. Strigari and R. Trotta, *Indirect Dark Matter Detection from Dwarf Satellites: Joint Expectations from Astrophysics and Supersymmetry*, *JCAP* **0906** (2009) 014 [`0902.4715`].

[44] A. V. Maccio, S. Paduroiu, D. Anderhalden, A. Schneider and B. Moore, *Cores in warm dark matter haloes: a Catch 22 problem*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 1105 [`1202.1282`].

[45] V. Springel, *The Cosmological simulation code GADGET-2*, *Mon. Not. Roy. Astron. Soc.* **364** (2005) 1105 [`astro-ph/0505010`].

[46] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu et al., *Properties of galaxies reproduced by a hydrodynamic simulation*, *Nature* **509** (2014) 177 [`1405.1418`].

[47] M. R. Lovell, C. S. Frenk, V. R. Eke, A. Jenkins, L. Gao and T. Theuns, *The properties of warm dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **439** (2014) 300 [`1308.1399`].

[48] S. Bose, W. A. Hellwing, C. S. Frenk, A. Jenkins, M. R. Lovell, J. C. Helly et al., *The COpernicus COmplexio: Statistical Properties of Warm Dark Matter Haloes*, *Mon. Not. Roy. Astron. Soc.* **455** (2016) 318 [`1507.01998`].

[49] S. Alekhin et al., *A facility to Search for Hidden Particles at the CERN SPS: the SHiP physics case*, *Rept. Prog. Phys.* **79** (2016) 124201 [`1504.04855`].

[50] A. D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32.

[51] T. Asaka and M. Shaposhnikov, *The nuMSM, dark matter and baryon asymmetry of the universe*, *Phys. Lett.* **B620** (2005) 17 [`hep-ph/0505013`].

[52] T. Asaka, S. Blanchet and M. Shaposhnikov, *The nuMSM, dark matter and neutrino masses*, *Phys. Lett.* **B631** (2005) 151 [`hep-ph/0503065`].

[53] S. Eijima, M. Shaposhnikov and I. Timiryasov, *Parameter space of baryogenesis in the νMSM*, *JHEP* **07** (2019) 077 [`1808.10833`].

[54] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens and O. Ruchayskiy, *Sterile Neutrino Dark Matter*, *Prog. Part. Nucl. Phys.* **104** (2019) 1 [`1807.07938`].

[55] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein and S. W. Randall, *Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters*, *Astrophys. J.* **789** (2014) 13 [`1402.2301`].

[56] A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi and J. Franse, *Unidentified Line in X-Ray Spectra of the Andromeda Galaxy and Perseus Galaxy Cluster*, *Phys. Rev. Lett.* **113** (2014) 251301 [`1402.4119`].

[57] A. Boyarsky, J. Franse, D. Iakubovskyi and O. Ruchayskiy, *Checking the dark matter origin of 3.53 keV line with the Milky Way center*, *Phys. Rev. Lett.* **115** (2015) 161301 [`1408.2503`].

[58] O. Urban, N. Werner, S. W. Allen, A. Simionescu, J. S. Kaastra and L. E. Strigari, *A Suzaku Search for Dark Matter Emission Lines in the X-ray Brightest Galaxy Clusters*, *Mon. Not. Roy. Astron. Soc.* **451** (2015) 2447 [`1411.0050`].

[59] D. Iakubovskyi, E. Bulbul, A. R. Foster, D. Savchenko and V. Sadova, *Testing the origin of 3.55 keV line in individual galaxy clusters observed with XMM-Newton*, [`1508.05186`].

[60] J. Franse et al., *Radial Profile of the 3.55 keV line out to R_{200} in the Perseus Cluster*, *Astrophys. J.* **829** (2016) 124 [`1604.01759`].

[61] O. Ruchayskiy, A. Boyarsky, D. Iakubovskyi, E. Bulbul, D. Eckert, J. Franse et al., *Searching for decaying dark matter in deep XMMNewton observation of the Draco dwarf spheroidal*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1390 [`1512.07217`].

[62] A. Neronov, D. Malyshev and D. Eckert, *Decaying dark matter search with NuSTAR deep sky observations*, *Phys. Rev.* **D94** (2016) 123504 [`1607.07328`].

[63] K. Perez, K. C. Y. Ng, J. F. Beacom, C. Hersh, S. Horiuchi and R. Krivonos, *Almost closing the MSM sterile neutrino dark matter window with NuSTAR*, *Phys. Rev.* **D95** (2017) 123002 [1609.00667].

[64] N. Cappelluti, E. Bulbul, A. Foster, P. Natarajan, M. C. Urry, M. W. Bautz et al., *Searching for the 3.5 keV Line in the Deep Fields with Chandra: the 10 Ms observations*, *Astrophys. J.* **854** (2018) 179 [1701.07932].

[65] S. Riemer-Srensen, *Constraints on the presence of a 3.5 keV dark matter emission line from Chandra observations of the Galactic centre*, *Astron. Astrophys.* **590** (2016) A71 [1405.7943].

[66] D. Malyshev, A. Neronov and D. Eckert, *Constraints on 3.55 keV line emission from stacked observations of dwarf spheroidal galaxies*, *Phys. Rev.* **D90** (2014) 103506 [1408.3531].

[67] S. Riemer-Srensen et al., *Dark matter line emission constraints from NuSTAR observations of the Bullet Cluster*, *Astrophys. J.* **810** (2015) 48 [1507.01378].

[68] M. E. Anderson, E. Churazov and J. N. Bregman, *Non-Detection of X-Ray Emission From Sterile Neutrinos in Stacked Galaxy Spectra*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 3905 [1408.4115].

[69] XQC collaboration, *Searching for keV Sterile Neutrino Dark Matter with X-ray Microcalorimeter Sounding Rockets*, *Astrophys. J.* **814** (2015) 82 [1506.05519].

[70] T. Tamura, R. Iizuka, Y. Maeda, K. Mitsuda and N. Y. Yamasaki, *An X-ray Spectroscopic Search for Dark Matter in the Perseus Cluster with Suzaku*, *Publ. Astron. Soc. Jap.* **67** (2015) 23 [1412.1869].

[71] N. Sekiya, N. Y. Yamasaki and K. Mitsuda, *A Search for a keV Signature of Radiatively Decaying Dark Matter with Suzaku XIS Observations of the X-ray Diffuse Background*, *Publ. Astron. Soc. Jap.* (2015) [1504.02826].

[72] HITOMI collaboration, *Hitomi constraints on the 3.5 keV line in the Perseus galaxy cluster*, *Astrophys. J.* **837** (2017) L15 [1607.07420].

[73] S. Dodelson and L. M. Widrow, *Sterile-neutrinos as dark matter*, *Phys. Rev. Lett.* **72** (1994) 17 [hep-ph/9303287].

[74] K. Bondarenko, A. Boyarsky, D. Gorbunov and O. Ruchayskiy, *Phenomenology of GeV-scale Heavy Neutral Leptons*, *JHEP* **11** (2018) 032 [1805.08567].

[75] P. Bodenheimer, *Depletion of Deuterium and Beryllium during Pre-Main Evolution*, *Astrophys. J.* **144** (1966) 103.

[76] R. I. Epstein, J. M. Lattimer and D. N. Schramm, *The origin of deuterium*, Nature **263** (1976) 198.

[77] G. Gamow, *The origin of elements and the separation of galaxies*, *Phys. Rev.* **74** (1948) 505.

[78] T. Otsuka, *Exotic nuclei and nuclear forces*, *Phys. Scripta* **T152** (2013) 014007.

[79] C. Pitrou, A. Coc, J.-P. Uzan and E. Vangioni, *Precision Big Bang Nucleosynthesis with the*

New Code PRIMAT, in *15th International Symposium on Origin of Matter and Evolution of the Galaxies (OMEG15) Kyoto, Japan, July 2-5, 2019*, 2019, [`1909.12046`].

[80] D. H. Wilkinson, *ANALYSIS OF NEUTRON BETA DECAY*, *Nucl. Phys.* **A377** (1982) 474.

[81] M. Pospelov and J. Pradler, *Metastable GeV-scale particles as a solution to the cosmological lithium problem*, *Phys. Rev.* **D82** (2010) 103514 [`1006.4172`].

[82] S. Dodelson, *Modern Cosmology*. Academic Press, Amsterdam, 2003, http://www.slac.stanford.edu/spires/find/books/www?cl=QB981:D62:2003.

[83] L. Kawano, *Let's go: Early universe. 2. Primordial nucleosynthesis: The Computer way*. 1992, http://inspirehep.net/search?p=FERMILAB-PUB-92-004-A.

[84] E. Lisi, S. Sarkar and F. L. Villante, *The big bang nucleosynthesis limit on N(neutrino)*, *Phys. Rev.* **D59** (1999) 123520 [`hep-ph/9901404`].

[85] L. Mendoza and C. J. Hogan, *A java calculator of standard big bang nucleosynthesis*, [`astro-ph/9904334`].

[86] A. Arbey, *AlterBBN: A program for calculating the BBN abundances of the elements in alternative cosmologies*, *Comput. Phys. Commun.* **183** (2012) 1822 [`1106.1363`].

[87] R. Consiglio, P. F. de Salas, G. Mangano, G. Miele, S. Pastor and O. Pisanti, *PArthENoPE reloaded*, *Comput. Phys. Commun.* **233** (2018) 237 [`1712.04378`].

[88] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *Chin. Phys.* **C40** (2016) 100001.

[89] Y. I. Izotov and T. X. Thuan, *The Primordial Abundance of ^{4}He: Evidence for Non-Standard Big Bang Nucleosynthesis*, *Astrophys. J.* **710** (2010) L67 [`1001.4440`].

[90] Y. I. Izotov, T. X. Thuan and N. G. Guseva, *A new determination of the primordial He abundance using the He i λ10830 emission line: cosmological implications*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) 778 [`1408.6953`].

[91] F. Iocco, G. Mangano, G. Miele, O. Pisanti and P. D. Serpico, *Primordial Nucleosynthesis: from precision cosmology to fundamental physics*, *Phys. Rept.* **472** (2009) 1 [`0809.0631`].

[92] R. Cooke and M. Fumagalli, *Measurement of the primordial helium abundance from the intergalactic medium*, *Nat. Astron.* **2** (2018) 957 [`1810.06561`].

[93] R. J. Cooke, M. Pettini and C. C. Steidel, *One Percent Determination of the Primordial Deuterium Abundance*, *Astrophys. J.* **855** (2018) 102 [`1710.11129`].

[94] R. I. Epstein, J. M. Lattimer and D. N. Schramm, *The Origin of deuterium*, *Nature* **263** (1976) 198.

[95] B. E. Wood, J. L. Linsky, G. Hebrard, G. M. Williger, H. W. Moos and W. P. Blair, *Two new low Galactic D/H measurements from FUSE*, *Astrophys. J.* **609** (2004) 838 [`astro-ph/0403606`].

[96] T. M. Bania, R. T. Rood and D. S. Balser, *The cosmological density of baryons from observations of 3He+ in the Milky Way*, *Nature* **415** (2002) 54.

[97] K. A. Olive, D. N. Schramm, S. T. Scully and J. W. Truran, *Low mass stars and the He-3 problem*, *Astrophys. J.* **479** (1997) 752 [`astro-ph/9610039`].

[98] M. Spite and F. Spite, *Lithium abundance at the formation of the Galaxy*, Nature **297** (1982) 483.

[99] L. Sbordone et al., *The metal-poor end of the Spite plateau. 1: Stellar parameters, metallicities and lithium abundances*, *Astron. Astrophys.* **522** (2010) A26 [`1003.4510`].

[100] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, *ArXiv e-prints* (2018) [`1807.06209`].

[101] R. H. Cyburt, B. D. Fields, K. A. Olive and T.-H. Yeh, *Big Bang Nucleosynthesis: 2015*, *Rev. Mod. Phys.* **88** (2016) 015004 [`1505.01076`].

[102] M. Kawasaki and T. Moroi, *Electromagnetic cascade in the early universe and its application to the big bang nucleosynthesis*, *Astrophys. J.* **452** (1995) 506 [`astro-ph/9412055`].

[103] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P. D. Serpico, *Relic neutrino decoupling including flavor oscillations*, *Nucl. Phys.* **B729** (2005) 221 [`hep-ph/0506164`].

[104] A. D. Dolgov, S. H. Hansen, G. Raffelt and D. V. Semikoz, *Cosmological and astrophysical bounds on a heavy sterile neutrino and the KARMEN anomaly*, *Nucl. Phys.* **B580** (2000) 331 [`hep-ph/0002223`].

[105] A. D. Dolgov, S. H. Hansen, G. Raffelt and D. V. Semikoz, *Heavy sterile neutrinos: Bounds from big bang nucleosynthesis and SN1987A*, *Nucl. Phys.* **B590** (2000) 562 [`hep-ph/0008138`].

[106] O. Ruchayskiy and A. Ivashko, *Restrictions on the lifetime of sterile neutrinos from primordial nucleosynthesis*, *JCAP* **1210** (2012) 014 [`1202.2841`].

[107] M. Kawasaki, P. Kernan, H.-S. Kang, R. J. Scherrer, G. Steigman and T. P. Walker, *Big bang nucleosynthesis constraints on the tau-neutrino mass*, *Nucl. Phys.* **B419** (1994) 105.

[108] S. Hannestad, *What is the lowest possible reheating temperature?*, *Phys. Rev.* **D70** (2004) 043506 [`astro-ph/0403291`].

[109] B. Gough, *GNU Scientific Library Reference Manual - Third Edition*. Network Theory Ltd., 3rd ed., 2009, https://dl.acm.org/citation.cfm?id=1538674.

[110] C. Pitrou, A. Coc, J.-P. Uzan and E. Vangioni, *Precision big bang nucleosynthesis with improved Helium-4 predictions*, *Phys. Rept.* **754** (2018) 1 [`1801.08023`].

[111] E. Grohs, G. M. Fuller, C. T. Kishimoto, M. W. Paris and A. Vlasenko, *Neutrino energy transport in weak decoupling and big bang nucleosynthesis*, *Phys. Rev.* **D93** (2016) 083522 [`1512.02205`].

[112] E. P. Hubble, *Extragalactic nebulae*, *Astrophys. J.* **64** (1926) 321.

[113] H. Shapley and A. Ames, *A Survey of the External Galaxies Brighter than the Thirteenth Magnitude*, *Annals of Harvard College Observatory* **88** (1932) 41.

[114] E. Hubble, *The Distribution of Extra-Galactic Nebulae*, *Astrophys. J.* **79** (1934) 8.

[115] M. Colless et al., *The 2dF Galaxy Redshift Survey: Final data release*, [`astro-ph/0306581`].

[116] J. R. Gott, III, M. Juric, D. Schlegel, F. Hoyle, M. Vogeley, M. Tegmark et al., *A map of the universe*, *Astrophys. J.* **624** (2005) 463 [`astro-ph/0310571`].

[117] L. S. Sparke and J. S. Gallagher, *Galaxies in the universe: an introduction*. Cambridge Univ. Press, 2010.

[118] F. Fontanot, G. De Lucia, P. Monaco, R. S. Somerville and P. Santini, *The Many Manifestations of Downsizing: Hierarchical Galaxy Formation Models confront Observations*, *Mon. Not. Roy. Astron. Soc.* **397** (2009) 1776 [`0901.1130`].

[119] BOSS collaboration, *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Releases 10 and 11 Galaxy samples*, *Mon. Not. Roy. Astron. Soc.* **441** (2014) 24 [`1312.4877`].

[120] E. Aubourg et al., *Cosmological implications of baryon acoustic oscillation measurements*, *Phys. Rev.* **D92** (2015) 123516 [`1411.1074`].

[121] DES collaboration, *Dark Energy Survey Year 1 Results: A Precise H0 Estimate from DES Y1, BAO, and D/H Data*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 3879 [`1711.00403`].

[122] Ya. B. Zeldovich, *Gravitational instability: An Approximate theory for large density perturbations*, *Astron. Astrophys.* **5** (1970) 84.

[123] M. Crocce, S. Pueblas and R. Scoccimarro, *Transients from initial conditions in cosmological simulations*, *Monthly Notices of the Royal Astronomical Society* **373** (2006) 369381.

[124] J. Makino and T. Saitoh, *Astrophysics with GRAPE*, *PTEP* **2012** (2012) 01A303.

[125] J. Barnes and P. Hut, *A Hierarchical 0 Nlogn Force Calculation Algorithm*, *Nature* **324** (1986) 446.

[126] M. J. Berger and P. Colella, *Local Adaptive Mesh Refinement for Shock Hydrodynamics*, *Journal of Computational Physics* **82** (1989) 64.

[127] R. A. Gingold and J. J. Monaghan, *Smoothed particle hydrodynamics: theory and application to non-spherical stars.*, *Mon. Not. Roy. Astron. Soc.* **181** (1977) 375.

[128] J. J. Monaghan, *Smoothed particle hydrodynamics*, *Ann. Rev. Astron. Astrophys.* **30** (1992) 543.

[129] N. Katz, D. H. Weinberg and L. Hernquist, *Cosmological simulations with TreeSPH*, *Astrophys. J.Suppl.* **105** (1996) 19 [`astro-ph/9509107`].

[130] V. Springel et al., *Simulating the joint evolution of quasars, galaxies and their large-scale distribution*, *Nature* **435** (2005) 629 [`astro-ph/0504097`].

[131] V. Springel, C. S. Frenk and S. D. M. White, *The large-scale structure of the Universe*, *Nature* **440** (2006) 1137 [`astro-ph/0604561`].

[132] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [`1707.03397`].

[133] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy and M. Viel, *Lyman-α constraints on warm and on warm-plus-cold dark matter models*, J. Cosmology Astropart. Phys. **5** (2009) 012 [`0812.0010`].

[134] A. Schneider, R. E. Smith and D. Reed, *Halo mass function and the free streaming scale*, *Mon. Not. Roy. Astron. Soc.* **433** (2013) 1573 [`1303.0839`].

[135] R. E. Angulo, O. Hahn and T. Abel, *The warm dark matter halo mass function below the cut-off scale*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) 3337 [`1304.2406`].

[136] M. R. Lovell, S. Bose, A. Boyarsky, S. Cole, C. S. Frenk, V. Gonzalez-Perez et al., *Satellite galaxies in semi-analytic models of galaxy formation with sterile neutrino dark matter*, *Mon. Not. Roy. Astron. Soc.* **461** (2016) 60 [`1511.04078`].

[137] M. R. Lovell, S. Bose, A. Boyarsky, R. A. Crain, C. S. Frenk, W. A. Hellwing et al., *Properties of Local Group galaxies in hydrodynamical simulations of sterile neutrino dark matter cosmologies*, *Mon. Not. Roy. Astron. Soc.* **468** (2017) 4285 [`1611.00010`].

[138] R. Li, C. S. Frenk, S. Cole, L. Gao, S. Bose and W. A. Hellwing, *Constraints on the identity of the dark matter from strong gravitational lenses*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 363 [`1512.06507`].

[139] D. Erkal, V. Belokurov, J. Bovy and J. L. Sanders, *The number and size of subhalo-induced gaps in stellar streams*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 102 [`1606.04946`].

[140] S. Tremaine and J. E. Gunn, *Dynamical role of light neutral leptons in cosmology*, *Physical Review Letters* **42** (1979) 407.

[141] A. V. Maccio, S. Paduroiu, A. Schneider and B. Moore, *Cores in warm dark matter haloes: a Catch 22 problem*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 1105 [`1202.1282`].

[142] S. Shao, L. Gao, T. Theuns and C. S. Frenk, *The phase-space density of fermionic dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **430** (2013) 2346 [`1209.5563`].

[143] J. F. Navarro, V. R. Eke and C. S. Frenk, *The cores of dwarf galaxy haloes*, *Mon. Not. Roy. Astron. Soc.* **283** (1996) L72 [`astro-ph/9610187`].

[144] F. Governato, C. Brook, L. Mayer, A. Brooks, G. Rhee, J. Wadsley et al., *Bulgeless dwarf galaxies and dark matter cores from supernova-driven outflows*, Nature **463** (2010) 203 [`0911.2237`].

[145] A. Pontzen and F. Governato, *How supernova feedback turns dark matter cusps into cores*, *Mon. Not. Roy. Astron. Soc.* **421** (2012) 3464 [`1106.0499`].

[146] K. A. Oman, J. F. Navarro, A. Fattahi, C. S. Frenk, T. Sawala, S. D. M. White et al., *The unexpected diversity of dwarf galaxy rotation curves*, *Mon. Not. Roy. Astron. Soc.* **452** (2015) 3650 [`1504.01437`].

[147] R. Adhikari et al., *A White Paper on keV Sterile Neutrino Dark Matter*, *JCAP* **1701** (2017) 025 [`1602.04816`].

[148] U. Seljak, A. Makarov, P. McDonald and H. Trac, *Can sterile neutrinos be the dark matter?*, *Phys. Rev. Lett.* **97** (2006) 191303 [`astro-ph/0602430`].

[149] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese and A. Riotto, *Can sterile neutrinos be ruled out as warm dark matter candidates?*, *Phys. Rev. Lett.* **97** (2006) 071301 [`astro-ph/0605706`].

[150] M. Viel, G. D. Becker, J. S. Bolton and M. G. Haehnelt, *Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data*, Phys. Rev. D **88** (2013) 043502 [`1306.2314`].

[151] E. Boera, G. D. Becker, J. S. Bolton and F. Nasir, *Revealing Reionization with the Thermal History of the Intergalactic Medium: New Constraints from the Lyα Flux Power Spectrum*, *Astrophys. J.* **872** (2019) 101 [`1809.06980`].

[152] P. Bode, J. P. Ostriker and N. Turok, *Halo formation in warm dark matter models*, *Astrophys. J.* **556** (2001) 93 [`astro-ph/0010389`].

[153] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy and M. Viel, *Realistic sterile neutrino dark matter with keV mass does not contradict cosmological bounds*, *Phys. Rev. Lett.* **102** (2009) 201304 [`0812.3256`].

[154] L. Covi, H.-B. Kim, J. E. Kim and L. Roszkowski, *Axinos as dark matter*, *JHEP* **05** (2001) 033 [`hep-ph/0101009`].

[155] K.-Y. Choi, J. E. Kim and L. Roszkowski, *Review of axino dark matter*, *J. Korean Phys. Soc.* **63** (2013) 1685 [`1307.3330`].

[156] T. Moroi, H. Murayama and M. Yamaguchi, *Cosmological constraints on the light stable gravitino*, *Phys. Lett.* **B303** (1993) 289.

[157] X.-D. Shi and G. M. Fuller, *A New dark matter candidate: Nonthermal sterile neutrinos*, *Phys. Rev. Lett.* **82** (1999) 2832 [`astro-ph/9810076`].

[158] M. Laine and M. Shaposhnikov, *Sterile neutrino dark matter as a consequence of nuMSM-induced lepton asymmetry*, *JCAP* **0806** (2008) 031 [`0804.4543`].

[159] K. Kashiyama and M. Oguri, *Detectability of Small-Scale Dark Matter Clumps with Pulsar Timing Arrays*, [`1801.07847`].

[160] J. S. Bullock and M. Boylan-Kolchin, *Small-Scale Challenges to the ΛCDM Paradigm*, *Ann. Rev. Astron. Astrophys.* **55** (2017) 343 [`1707.04256`].

[161] B. Moore, S. Ghigna, F. Governato, G. Lake, T. R. Quinn, J. Stadel et al., *Dark matter substructure within galactic halos*, *Astrophys. J.* **524** (1999) L19 [`astro-ph/9907411`].

[162] A. A. Klypin, A. V. Kravtsov, O. Valenzuela and F. Prada, *Where are the missing Galactic satellites?*, *Astrophys. J.* **522** (1999) 82 [`astro-ph/9901240`].

[163] A. Fattahi, J. F. Navarro, T. Sawala, C. S. Frenk, L. V. Sales, K. Oman et al., *The cold dark*

matter content of Galactic dwarf spheroidals: no cores, no failures, no problem,
[1607.06479].

[164] R. A. Flores and J. R. Primack, *Observational and theoretical constraints on singular dark matter halos*, *Astrophys. J.* **427** (1994) L1 [astro-ph/9402004].

[165] B. Moore, *Evidence against dissipation-less dark matter from observations of galaxy haloes*, *Nature* **370** (1994) 629.

[166] M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, *Too big to fail? The puzzling darkness of massive Milky Way subhaloes*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) L40 [1103.0007].

[167] M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, *The Milky Way's bright satellites as an apparent failure of LCDM*, *Mon. Not. Roy. Astron. Soc.* **422** (2012) 1203 [1111.2048].

[168] S. Mashchenko, J. Wadsley and H. M. P. Couchman, *Stellar Feedback in Dwarf Galaxy Formation*, *Science* **319** (2008) 174 [0711.4803].

[169] J. Wang, C. S. Frenk, J. F. Navarro, L. Gao and T. Sawala, *The missing massive satellites of the Milky Way*, *Mon. Not. Roy. Astron. Soc.* **424** (2012) 2715 [1203.4097].

[170] T. Sawala et al., *The APOSTLE simulations: solutions to the Local Group's cosmic puzzles*, *Mon. Not. Roy. Astron. Soc.* **457** (2016) 1931 [1511.01098].

[171] Q. Guo, A. P. Cooper, C. Frenk, J. Helly and W. A. Hellwing, *The Milky Way system in Λ cold dark matter cosmological simulations*, *Mon. Not. Roy. Astron. Soc.* **454** (2015) 550 [1503.08508].

[172] K. A. Oman, J. F. Navarro, L. V. Sales, A. Fattahi, C. S. Frenk, T. Sawala et al., *Missing dark matter in dwarf galaxies?*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 3610 [1601.01026].

[173] G. Despali, S. Vegetti, S. D. M. White, C. Giocoli and F. C. van den Bosch, *Modelling the line-of-sight contribution in substructure lensing*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 5424 [1710.05029].

[174] K. T. Inoue, R. Takahashi, T. Takahashi and T. Ishiyama, *Constraints on warm dark matter from weak lensing in anomalous quadruple lenses*, *Mon. Not. Roy. Astron. Soc.* **448** (2015) 2704 [1409.1326].

[175] D. Gilman, S. Birrer, T. Treu, C. R. Keeton and A. Nierenberg, *Probing the nature of dark matter by forward modelling flux ratios in strong gravitational lenses*, *Mon. Not. Roy. Astron. Soc.* **481** (2018) 819 [1712.04945].

[176] S. Birrer, A. Amara and A. Refregier, *Lensing substructure quantification in RXJ1131-1231: A 2 keV lower bound on dark matter thermal relic mass*, *JCAP* **1705** (2017) 037 [1702.00009].

[177] EUCLID THEORY WORKING GROUP collaboration, *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **16** (2013) 6 [1206.1225].

[178] L. Amendola et al., *Cosmology and fundamental physics with the Euclid satellite*, *Living Rev. Rel.* **21** (2018) 2 [1606.00180].

[179] L. Gao and T. Theuns, *Lighting the Universe with filaments*, *Science* **317** (2007) 1527 [`0709.2165`].

[180] L. Gao, T. Theuns and V. Springel, *Star forming filaments in warm dark matter models*, *Mon. Not. Roy. Astron. Soc.* **450** (2015) 45 [`1403.2475`].

[181] S. Cantalupo, F. Arrigoni-Battaia, J. X. Prochaska, J. F. Hennawi and P. Madau, *A cosmic web filament revealed in Lyman-α emission around a luminous high-redshift quasar*, Nature **506** (2014) 63 [`1401.4469`].

[182] P. Madau, A. Meiksin and M. J. Rees, *21 Centimeter Tomography of the Intergalactic Medium at High Redshift*, *Astrophys. J.* **475** (1997) 429 [`astro-ph/9608010`].

[183] S. Tulin and H.-B. Yu, *Dark Matter Self-interactions and Small Scale Structure*, [`1705.02358`].

[184] XENON collaboration, *First Dark Matter Search Results from the XENON1T Experiment*, *Phys. Rev. Lett.* **119** (2017) 181301 [`1705.06655`].

[185] PANDAX-II collaboration, *Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment*, *Phys. Rev. Lett.* **119** (2017) 181302 [`1708.06917`].

[186] N. Y. Gnedin and L. Hui, *Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium*, *Mon. Not. Roy. Astron. Soc.* **296** (1998) 44 [`arXiv:astro-ph/9706219`].

[187] L. Hui and N. Y. Gnedin, *Equation of state of the photoionized intergalactic medium*, *Mon. Not. Roy. Astron. Soc.* **292** (1997) 27 [`astro-ph/9612232`].

[188] F. Haardt and P. Madau, *Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background*, *Astrophys. J.* **461** (1996) 20 [`astro-ph/9509093`].

[189] T. Theuns, A. Leonard, G. Efstathiou, F. R. Pearce and P. A. Thomas, *Pˆ3M-SPH simulations of the Lyalpha forest*, *Mon. Not. Roy. Astron. Soc.* **301** (1998) 478 [`astro-ph/9805119`].

[190] P. R. Upton Sanderbeck, A. D'Aloisio and M. J. McQuinn, *Models of the thermal evolution of the intergalactic medium after reionization*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1885 [`1511.05992`].

[191] G. Giesen, J. Lesgourgues, B. Audren and Y. Ali-Haimoud, *CMB photons shedding light on dark matter*, *JCAP* **1212** (2012) 008 [`1209.0247`].

[192] V. Poulin, P. D. Serpico and J. Lesgourgues, *Dark Matter annihilations in halos and high-redshift sources of reionization of the universe*, *JCAP* **1512** (2015) 041 [`1508.01370`].

[193] V. Poulin, J. Lesgourgues and P. D. Serpico, *Cosmological constraints on exotic injection of electromagnetic energy*, *JCAP* **1703** (2017) 043 [`1610.10051`].

[194] SDSS collaboration, *Evidence for reionization at Z 6: Detection of a Gunn-Peterson trough in a Z = 6.28 quasar*, *Astron. J.* **122** (2001) 2850 [`astro-ph/0108097`].

[195] J. E. Gunn and B. A. Peterson, *On the Density of Neutral Hydrogen in Intergalactic Space.*, *Astrophys. J.* **142** (1965) 1633.

[196] D. K. Hazra and G. F. Smoot, *Witnessing the reionization history using Cosmic Microwave Background observation from Planck*, *JCAP* **1711** (2017) 028 [`1708.04913`].

[197] THE HISTORY OF REIONIZATION ARRAY collaboration, *A Roadmap for Astrophysics and Cosmology with High-Redshift 21 cm Intensity Mapping*, [`1907.06440`].

[198] I. Pris et al., *The Sloan Digital Sky Survey quasar catalog: tenth data release*, *Astron. Astrophys.* **563** (2014) A54 [`1311.4870`].

[199] P. J. Francis, P. C. Hewett, C. B. Foltz, F. H. Chaffee, R. J. Weymann and S. L. Morris, *A High Signal-to-Noise Ratio Composite Quasar Spectrum*, *Astrophys. J.* **373** (1991) 465.

[200] T. Theuns, A. Leonard, G. Efstathiou, F. R. Pearce and P. A. Thomas, *P**3M-SPH simulations of the lyman-alpha forest*, *Mon. Not. Roy. Astron. Soc.* **301** (1998) 478 [`astro-ph/9805119`].

[201] F. Haardt and P. Madau, *Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background*, *Astrophys. J.* **461** (1996) 20 [`astro-ph/9509093`].

[202] R. A. C. Croft, *Ionizing Radiation Fluctuations and Large-Scale Structure in the Ly alpha Forest*, *Astrophys. J.* **610** (2004) 642 [`astro-ph/0310890`].

[203] P. McDonald, U. Seljak, R. Cen, P. Bode and J. P. Ostriker, *Physical effects on the Lyα forest flux power spectrum: damping wings, ionizing radiation fluctuations and galactic winds*, *Mon. Not. Roy. Astron. Soc.* **360** (2005) 1471 [`astro-ph/0407378`].

[204] G. D. Becker, F. B. Davies, S. R. Furlanetto, M. A. Malkan, E. Boera and C. Douglass, *Evidence for Large-Scale Fluctuations in the Metagalactic Ionizing Background Near Redshift Six*, *ArXiv e-prints* (2018) [`1803.08932`].

[205] S. E. I. Bosman, X. Fan, L. Jiang, S. Reed, Y. Matsuoka, G. Becker et al., *New constraints on Lyman-α opacity with a sample of 62 quasarsat z ¿ 5.7*, *Mon. Not. Roy. Astron. Soc.* **479** (2018) 1055 [`1802.08177`].

[206] Planck Collaboration, R. Adam, N. Aghanim, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck intermediate results. XLVII. Planck constraints on reionization history*, A&A **596** (2016) A108 [`1605.03507`].

[207] D. J. Mortlock, S. J. Warren, B. P. Venemans, M. Patel, P. C. Hewett, R. G. McMahon et al., *A luminous quasar at a redshift of z = 7.085*, Nature **474** (2011) 616 [`1106.6088`].

[208] F. B. Davies, J. F. Hennawi, E. Bañados, Z. Lukić, R. Decarli, X. Fan et al., *Quantitative Constraints on the Reionization History from the IGM Damping Wing Signature in Two Quasars at z ¿ 7*, *ArXiv e-prints* (2018) [`1802.06066`].

[209] J. Schaye, R. A. Crain, R. G. Bower and al., *The EAGLE project: simulating the evolution and assembly of galaxies and their environments*, *Mon. Not. Roy. Astron. Soc.* **446** (2015) 521 [`1407.7040`].

[210] J. Oñorbe, J. F. Hennawi and Z. Lukić, *Self-consistent Modeling of Reionization in Cosmological Hydrodynamical Simulations*, *Astrophys. J.* **837** (2017) 106 [`1607.04218`].

[211] F. Haardt and P. Madau, *Modelling the UV/X-ray cosmic background with CUBA*, in *Clusters of Galaxies and the High Redshift Universe Observed in X-rays*, D. M. Neumann and J. T. V. Tran, eds., ArXiv astro-ph/0106018, 2001, [`arXiv:astro-ph/0106018`].

[212] R. Scoccimarro, L. Hui, M. Manera and K. C. Chan, *Large-scale bias and efficient generation of initial conditions for nonlocal primordial non-Gaussianity*, Phys. Rev. D **85** (2012) 083002 [`1108.5512`].

[213] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [`1502.01589`].

[214] M. Viel, J. Schaye and C. M. Booth, *The impact of feedback from galaxy formation on the Lyman α transmitted flux*, *Mon. Not. Roy. Astron. Soc.* **429** (2013) 1734 [`1207.6567`].

[215] A. Lewis, A. Challinor and A. Lasenby, *Efficient computation of CMB anisotropies in closed FRW models*, *Astrophys. J.* **538** (2000) 473 [`astro-ph/9911177`].

[216] J. Ghiglieri and M. Laine, *Improved determination of sterile neutrino dark matter spectrum*, *JHEP* **11** (2015) 171 [`1506.06752`].

[217] T. Venumadhav, F.-Y. Cyr-Racine, K. N. Abazajian and C. M. Hirata, *Sterile neutrino dark matter: Weak interactions in the strong coupling epoch*, *Phys. Rev.* **D94** (2016) 043515 [`1507.06655`].

[218] A. H. Pawlik, A. Rahmati, J. Schaye, M. Jeon and C. Dalla Vecchia, *The Aurora radiation-hydrodynamical simulations of reionization: calibration and first results*, *Mon. Not. Roy. Astron. Soc.* **466** (2017) 960 [`1603.00034`].

[219] T. Theuns, M. Viel, S. Kay, J. Schaye, R. F. Carswell and P. Tzanavaris, *Galactic Winds in the Intergalactic Medium*, *Astrophys. J.* **578** (2002) L5 [`astro-ph/0208418`].

[220] R. P. C. Wiersma, J. Schaye and B. D. Smith, *The effect of photoionization on the cooling rates of enriched, astrophysical plasmas*, *Mon. Not. Roy. Astron. Soc.* **393** (2009) 99 [`0807.3748`].

[221] G. Altay and T. Theuns, *URCHIN: a reverse ray tracer for astrophysical applications*, *Mon. Not. Roy. Astron. Soc.* **434** (2013) 748 [`1304.4235`].

[222] A. A. Meiksin, *The physics of the intergalactic medium*, *Reviews of Modern Physics* **81** (2009) 1405 [`0711.3358`].

[223] S. H. Hansen, J. Lesgourgues, S. Pastor and J. Silk, *Constraining the window on sterile neutrinos as warm dark matter*, *Mon. Not. Roy. Astron. Soc.* **333** (2002) 544 [`astro-ph/0106108`].

[224] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese and A. Riotto, *Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest*, *Phys. Rev.* **D71** (2005) 063534 [`astro-ph/0501562`].

[225] J. Baur, N. Palanque-Delabrouille, C. Yèche, A. Boyarsky, O. Ruchayskiy, É. Armengaud et al., *Constraints from Ly-α forests on non-thermal dark matter including resonantly-produced sterile neutrinos*, J. Cosmology Astropart. Phys. **12** (2017) 013 [`1706.03118`].

[226] J. Miralda-Escude and M. J. Rees, *Tests for the minihalo model of the Lyman-alpha forest*, *Mon. Not. Roy. Astron. Soc.* **260** (1993) 617.

[227] N. Y. Gnedin and L. Hui, *Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium*, *Mon. Not. Roy. Astron. Soc.* **296** (1998) 44 [`astro-ph/9706219`].

[228] A. Garzilli, T. Theuns and J. Schaye, *The broadening of Lyman-α forest absorption lines*, *Mon. Not. Roy. Astron. Soc.* **450** (2015) 1465 [`1502.05715`].

[229] A. Rorai, R. F. Carswell, M. G. Haehnelt, G. D. Becker, J. S. Bolton and M. T. Murphy, *A new measurement of the intergalactic temperature at z 2.55-2.95*, *Mon. Not. Roy. Astron. Soc.* **474** (2018) 2871 [`1711.00930`].

[230] T. Theuns, J. Schaye and M. G. Haehnelt, *Broadening of QSO Lyα forest absorbers*, *Mon. Not. Roy. Astron. Soc.* **315** (2000) 600 [`astro-ph/9908288`].

[231] SDSS collaboration, *The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey*, *Astrophys. J.Suppl.* **203** (2012) 21 [`1207.7137`].

[232] J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville and M. Viel, *Lyman-alpha forests cool warm dark matter*, *JCAP* **1608** (2016) 012 [`1512.01981`].

[233] U. Seljak, A. Makarov, P. McDonald and H. Trac, *Can sterile neutrinos be the dark matter?*, *Phys. Rev. Lett.* **97** (2006) 191303 [`astro-ph/0602430`].

[234] A. Garzilli, A. Boyarsky and O. Ruchayskiy, *Cutoff in the Lyman α forest power spectrum: warm IGM or warm dark matter?*, *Phys. Lett.* **B773** (2017) 258 [`1510.07006`].

[235] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, S. Cristiani, G. D. Becker et al., *New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data*, Phys. Rev. D **96** (2017) 023522 [`1702.01764`].

[236] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton and G. D. Becker, *First constraints on fuzzy dark matter from Lyman-α forest data and hydrodynamical simulations*, *Phys. Rev. Lett.* **119** (2017) 031302 [`1703.04683`].

[237] R. Murgia, V. Iri and M. Viel, *Novel constraints on noncold, nonthermal dark matter from Lyman-α forest data*, *Phys. Rev.* **D98** (2018) 083540 [`1806.08371`].

[238] G. D. Becker, J. S. Bolton, M. G. Haehnelt and W. L. W. Sargent, *Detection of extended He II reionization in the temperature evolution of the intergalactic medium*, *Mon. Not. Roy. Astron. Soc.* **410** (2011) 1096 [`1008.2622`].

[239] A. P. Calverley, G. D. Becker, M. G. Haehnelt and J. S. Bolton, *Measurements of the UV background at 4.6 ¡ z ¡ 6.4 using the quasar proximity effect*, *Mon. Not. Roy. Astron. Soc.* **412** (2011) 2543 [`1011.5850`].

[240] G. D. Becker, M. Rauch and W. L. W. Sargent, *The Evolution of Optical Depth in the Ly-alpha Forest: Evidence Against Reionization at z 6*, *Astrophys. J.* **662** (2007) 72 [`astro-ph/0607633`].

[241] J. Baur, N. Palanque-Delabrouille, C. Yeche, A. Boyarsky, O. Ruchayskiy, . Armengaud et al., *Constraints from Ly-α forests on non-thermal dark matter including resonantly-produced sterile neutrinos*, *JCAP* **1712** (2017) 013 [`1706.03118`].

[242] N. Kaiser, *Clustering in real space and in redshift space*, *Mon. Not. Roy. Astron. Soc.* **227** (1987) 1.

[243] N. Kaiser and J. A. Peacock, *Power-spectrum analysis of one-dimensional redshift surveys*, *Astrophys. J.* **379** (1991) 482.

[244] C. McGill, *The redshift projection. I - Caustics and correlation functions*, *Mon. Not. Roy. Astron. Soc.* **242** (1990) 428.

[245] R. Scoccimarro, *Redshift-space distortions, pairwise velocities and nonlinearities*, *Phys. Rev.* **D70** (2004) 083007 [`astro-ph/0407214`].

[246] V. Desjacques and A. Nusser, *Redshift distortions in one-dimensional power spectra*, *Mon. Not. Roy. Astron. Soc.* **351** (2004) 1395 [`astro-ph/0401544`].

[247] H. J. Mo, Y. P. Jing and G. Borner, *Analytical approximations to the low-order statistics of dark matter distributions*, *Mon. Not. Roy. Astron. Soc.* **286** (1997) 979 [`astro-ph/9607143`].

[248] J. Miralda-Escudé and M. J. Rees, *Reionization and thermal evolution of a photoionized intergalactic medium.*, *Mon. Not. Roy. Astron. Soc.* **266** (1994) 343.

[249] T. Abel and M. G. Haehnelt, *Radiative Transfer Effects during Photoheating of the Intergalactic Medium*, *Astrophys. J.* **520** (1999) L13 [`astro-ph/9903102`].

[250] J. Miralda-Escude and J. P. Ostriker, *What produces the ionizing background at large redshift?*, *Astrophys. J.* **350** (1990) 1.

[251] M. McQuinn, *Constraints on X-ray emissions from the reionization era*, *Mon. Not. Roy. Astron. Soc.* **426** (2012) 1349 [`1206.1335`].

[252] L. Hui and Z. Haiman, *The Thermal Memory of Reionization History*, *Astrophys. J.* **596** (2003) 9 [`astro-ph/0302439`].

[253] J. Oñorbe, J. F. Hennawi, Z. Lukić and M. Walther, *Constraining Reionization with the z = 5-6 Ly alpha Forest Power Spectrum: The Outlook after Planck*, *Astrophys. J.* **847** (2017) 63 [`1703.08633`].

[254] P. Jakobsen, A. Boksenberg, J. M. Deharveng, P. Greenfield, R. Jedrzejewski and F. Paresce, *Detection of intergalactic ionized helium absorption in a high-redshift quasar*, Nature **370** (1994) 35.

[255] J. Schaye, T. Theuns, M. Rauch, G. Efstathiou and W. L. W. Sargent, *The thermal history of the intergalactic medium**, *Mon. Not. Roy. Astron. Soc.* **318** (2000) 817 [`astro-ph/9912432`].

[256] P. La Plante, H. Trac, R. Croft and R. Cen, *Helium Reionization Simulations. II. Signatures of Quasar Activity on the IGM*, *Astrophys. J.* **841** (2017) 87 [`1610.02047`].

[257] D. Syphers and J. M. Shull, *Hubble Space Telescope/Cosmic Origins Spectrograph Observations of the Quasar Q0302-003: Probing the He II Reionization Epoch and QSO Proximity Effects*, *Astrophys. J.* **784** (2014) 42.

[258] P. McDonald, J. Miralda-Escude, M. Rauch, W. L. W. Sargent, T. A. Barlow and R. Cen, *A measurement of the temperature-density relation in the intergalactic medium using a new Lyman-alpha absorption line fitting method*, *Astrophys. J.* **562** (2001) 52 [`astro-ph/0005553`].

[259] A. Lidz, C. A. Faucher-Giguere, A. Dall'Aglio, M. McQuinn, C. Fechner, M. Zaldarriaga et al., *A Measurement of Small Scale Structure in the 2.2 ¡ z ¡ 4.2 Lyman-alpha Forest*, *Astrophys. J.* **718** (2010) 199 [`0909.5210`].

[260] J. S. Bolton, G. D. Becker, S. Raskutti, J. S. B. Wyithe, M. G. Haehnelt and W. L. W. Sargent, *Improved measurements of the intergalactic medium temperature around quasars: possible evidence for the initial stages of He II reionization at z = 6*, *Mon. Not. Roy. Astron. Soc.* **419** (2012) 2880.

[261] G. D. Becker, J. S. Bolton, P. Madau, M. Pettini, E. V. Ryan-Weber and B. P. Venemans, *Evidence of patchy hydrogen reionization from an extreme Lyα trough below redshift six*, *Mon. Not. Roy. Astron. Soc.* **447** (2015) 3402 [`1407.4850`].

[262] J. Schaye, *Model-independent Insights into the Nature of the Ly alpha Forest and the Distribution of Matter in the Universe*, *Astrophys. J.* **559** (2001) 507 [`arXiv:astro-ph/0104272`].

[263] C. A. Faucher-Giguere, A. Lidz, M. Zaldarriaga and L. Hernquist, *A New Calculation of the Ionizing Background Spectrum and the Effects of HeII Reionization*, *Astrophys. J.* **703** (2009) 1416 [`0901.4554`].

[264] F. Haardt and P. Madau, *Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background*, *Astrophys. J.* **746** (2012) 125 [`1105.2039`].

[265] J. R. Pritchard and A. Loeb, *21 cm cosmology in the 21st century*, *Reports on Progress in Physics* **75** (2012) 086901 [`1109.6012`].

Samenvatting

De kwantumtheorie van de elektrodynamica en pogingen om de subatomaire fysica fenomenologisch te verklaren, hebben samen geleid tot de ontwikkeling van de meest geavanceerde beschrijving van de deeltjesfysica tot nu toe. Het Standaardmodel van de deeltjesfysica verenigt de modellen van de elektromagnetische, zwakke en sterke interacties in een krachtig en elegant theoretisch kader. Desondanks is vandaag vastgesteld dat het Standaardmodel moet worden uitgebreid om de fenomenen buiten het Standaardmodel (BSM) te verklaren: donkere materie, materie-antimaterie asymmetrie van het heelal en neutrino oscillaties.

Nieuwe deeltjes en interacties die nodig zijn voor de BSM-fysica zijn tot nu toe in deeltjesfysica-experimenten niet ontdekt. De moeilijkheid van directe detectie van deze deeltjes ligt in de enorme parameterruimte van de mogelijke kandidaten. Daarom kan data afkomstig van kosmologische en astrofysische observaties van onschatbare waarde zijn voor laboratoriumexperimenten.

In dit proefschrift onderzoeken wij twee methoden om BSM kandidaten in te perken: door hun invloed op de zogenaamde Big Bang Nucleosynthese, en door waarneembare verschillen in de materieverdeling veroorzaakt door vrije stroming van donkere-materie deeltjes. We concentreren onze aandacht op die uitbreiding van het Standaardmodel, die tot doel heeft alle drie BSM fenomenen tegelijkertijd te verklaren: het *Neutrino Minimal Standard Model*. In deze uitbreiding worden drie extra *heavy neutral leptons* (of *steriele neutrino's*) toegevoegd aan het Standaardmodel. Één hiervan speelt de rol van donkere materie, terwijl de andere twee nodig zijn voor de verklaring van de materie-antimaterie asymmetrie en neutrino oscillaties. De donkere-materie kandidaat is een voorbeeld van Warme Donkere Materie, waarvan vrije stroming kan worden gedetecteerd in de "Lyman-α forest" spectra van verre quasars. De andere twee deeltjes hebben levensduren die relevant zijn voor Big Bang Nucleosynthese.

Heavy neutral leptons zijn vervallende deeltjes met neutrino-achtige interacties die de vorming van lichte elementen tijdens de Big Bang Nucleosynthese kunnen beïnvloeden. Dit gebeurt door middel van hun effect op de Hubble expansie en door hun verval in deeltjes, die zowel de spectra van gewone neutrino's deformeren of rechtstreeks wisselwerken met nucleonen. Steriele neutrino's met een massa hoger dan $\sim 100\,\mathrm{MeV}$ produceren ook kortlevende muonen en mesonen die ingewikkelde vervalketens op gang kunnen brengen. We leggen een methode voor om Big Bang Nucleosynthese numeriek te modelleren in de aanwezigheid van steriele neutrino's met massa's tot $\sim 1\,\mathrm{GeV}$, en stellen beperkingen aan

hun levensduur, uitgaande van verschillende koppelingsmechanismen.

Het waargenomen Lyman-α flux power spectrum (FPS) is onderdrukt op schalen kleiner dan $\sim 30\,\mathrm{km/s}$. Deze afkapping kan te wijten zijn aan de hoge temperatuur T_0 en druk p_0 van het absorberende gas of anders zou het de vrije stroom van donkere materie deeltjes in het vroege universum kunnen weerspiegelen. We voeren een reeks kosmologische hydrodynamische simulaties van zeer hoge resolutie uit, waarin we T_0, p_0 en de sterkte van de vrije stroming van donkere materie variëren, en vergelijken de FPS van gesimuleerde spectra met data. We tonen aan dat de FPS afkapping veroorzaakt kan worden door Koude Donkere Materie. Echter, het kan even goed worden verklaard, onder de aanname van dat de donkere materie bestaat uit $\sim 7\,\mathrm{keV}$ steriele neutrino's, in welk geval de afkapping voornamelijk te wijten is aan de vrije stroming van donkere materie. Hierdoor zijn de constricties voor de donkere materie kandidaten afhankelijk van een gedetailleerde kennis van de periode van reïonisatie. We laten zien hoe we krachtige beperkingen kunnen stellen aan warme donkere materie in het algemeen door over de thermische geschiedenis, die consistent zijn met waarnemingen, te marginaliseren.

Summary

Attempts to phenomenologically explain subatomic physics together with the well-rounded quantum theory of electrodynamics have culminated in the development of the most advanced description of particle physics to date. The Standard Model (SM) of particles unites the models of the electromagnetic, weak and strong interactions in a rigid and elegant theoretical framework. Nevertheless, today it is an established fact that the SM has to be extended to explain the so-called *Beyond the Standard Model* (BSM) phenomena: dark matter, matter-antimatter asymmetry of the Universe and neutrino flavour oscillations.

New particles and interactions necessary for BSM physics have so far evaded discovery in particle-physics experiments. The difficulty of direct detection lies in the huge parameter space of the possible candidates. Hence, data coming from the cosmological and astrophysical observation can provide invaluable directions for laboratory experiments.

In this thesis we explore two methods of constraining new-physics candidates: through their influence on the primordial nucleosynthesis and through observable differences in the matter distribution caused by free-streaming of the dark-matter particles. We concentrate our attention on the well-motivated extension of the SM that aims at explaining all 3 BSM problems at the same time: the *Neutrino Minimal Standard Model*. In this extension, there are 3 additional *heavy neutral leptons* (or *sterile neutrinos*), one of which plays the role of dark matter, while the other two are necessary for induction of matter-antimatter asymmetry and neutrino oscillations. The dark-matter candidate is an example of a Warm Dark Matter particle, the free-streaming of which might be detected in the Lyman-α forest spectra of distant quasars. The other two particles have lifetimes that make them relevant to the primordial nucleosynthesis.

Heavy sterile neutrinos are decaying particles with neutrino-like interactions that can influence the formation of light nuclei during primordial nucleosynthesis, both through their effect on the Hubble expansion rate and through the generation of particles that distort the spectra of SM neutrinos or interact directly with nucleons. Sterile neutrinos with masses above $\sim 100\,\mathrm{MeV}$ also produce short-lived muons and mesons that can trigger complicated decay chains. We present a method for numerical modelling the primordial nucleosynthesis in the presence of sterile neutrinos in the mass range up to $\sim 1\,\mathrm{GeV}$, and we put constraints on their lifetime assuming various coupling patterns.

The observed Lyman-α flux power spectrum (FPS) is suppressed on scales below $\sim 30\,\mathrm{km/s}$. This cutoff could be due to the high temperature T_0 and pressure p_0 of the

absorbing gas or, alternatively, it could reflect the free streaming of dark-matter particles in the early universe. We perform a set of very high resolution cosmological hydrodynamic simulations in which we vary T_0, p_0 and the amplitude of the free streaming of dark-matter, and compare the FPS of mock spectra to the data. We demonstrate that the FPS cutoff can be fitted assuming cold dark matter, but it can be equally well fitted assuming that the dark matter consists of ~ 7 keV sterile neutrinos in which case the cutoff is due primarily to the free-streaming dark matter. Consequently, the constraints on the dark-matter candidates depend on the detailed knowledge of the Epoch of Reionization. We demonstrate how to put robust constraints on general Warm Dark Matter by marginalizing over thermal histories consistent with observations.

List of publications

[1] A. Garzilli, A. Magalich, T. Theuns, C. S. Frenk, C. Weniger, O. Ruchayskiy and A. Boyarsky, "The Lyman-α forest as a diagnostic of the nature of the dark matter," Mon. Not. Roy. Astron. Soc., 489, 3456, 2018.

[2] A. Boyarsky, A. Garzilli, A. Magalich, O. Ruchayskiy, "How warm is too warm? Towards robust Lyman-α forest bounds on warm dark matter," [*in preparation for publication*] 2019.

[3] A. Magalich, N. Sabti, A. Filimonova, "Primordial nucleosynthesis constraints on sterile neutrinos with masses above pion mass," [*in preparation for publication*] 2019.

[4] M. Ovchynnikov, V. Syvolap, A. Magalich, O. Ruchayskiy, A. Boyarsky, "BBNSLOW. Primordial nucleosynthesis bounds on new physics without numerical codes," [*in preparation for publication*] 2019.

[5] A. Martini, A. Lutov, V. Gemmetto, A. Magalich, A. Cardillo, A. Constantin, V. Palchykov, M. Khayati, P. Cudré-Mauroux, A. Boyarsky, O. Ruchayskiy, D. Garlaschelli, P. De Los Rios, and, K. Aberer, "ScienceWISE: Topic Modeling over Scientific Literature Networks," 2016, *arXiv e-prints* [1612.07636].

Curriculum vitæ

I was born in Kyiv, Ukraine, on the 7th of August 1992. I have received my primary education at the Gymnasium No. 191, and my secondary education at the Kyiv Natural-Scientific Lyceum No. 145.

After finishing high school, I entered the Physics Department of the Taras Shevchenko National University of Kyiv in 2009. I have obtained my bachelor's degree in 2013 and started the Master's program at the Lorentz Institute for Theoretical Physics in Leiden. My Master thesis was titled "Constraints on sterile neutrinos lifetime from primordial nucleosynthesis". After graduation, I started my PhD studies under the supervision of Dr. Alexey Boyarsky at the Lorentz Institute.

During my PhD studies, I was a teaching assistant for the courses "Classical electrodynamics", "Topics in Theoretical Physics", "Particle Physics in the Early Universe" and "Effective Field Theory". I have visited a number of schools and conferences in the Netherlands, Italy, Switzerland, and Japan, where I presented my work.

After finishing my PhD studies, I would like to continue my academic career and stay in the field of theoretical physics.

Acknowledgements

I want to thank my supervisor Alexey Boyarsky for guidance during the years of my master's studies and graduate school.

I am also thankful to my second promotor Ana Achúcarro and my collaborators, Oleg Ruchayskiy, Antonella Garzilli, Nashwan Sabti, Anastasiia Filimonova and to my colleagues, Kyrylo Bondarenko, Anastasia Sokolenko, for company and countless hours of work together. Artem Ivashko was very kind to help me start my research in cosmology.

The support and patience of my fiancée Larysa have made the completion of this work possible. I am grateful to my parents, Angelica and Anatolii, and my sister, Tatiana, for bringing me up and for the dear home back in Ukraine.

The Lorentz Institute has been a hospitable environment all these years.