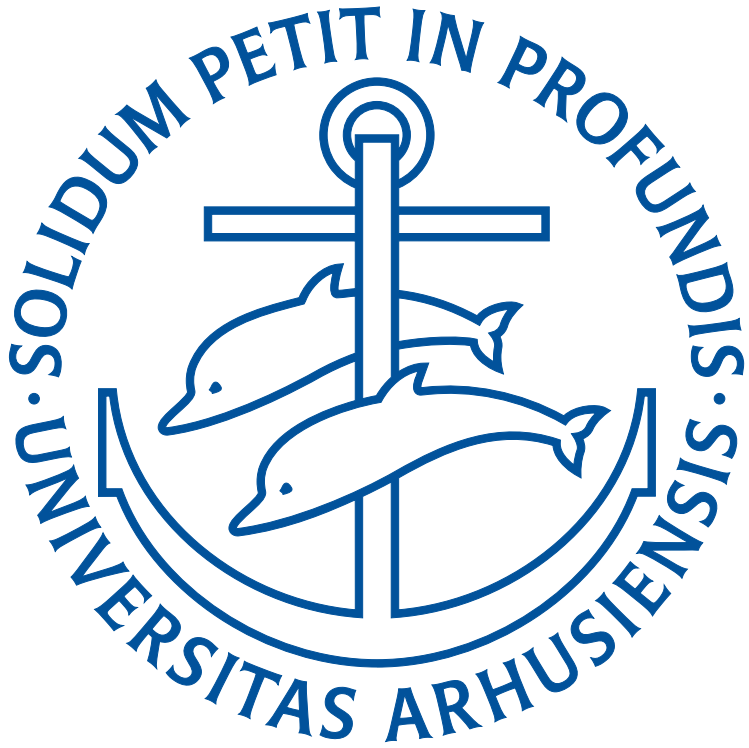
Dark Relics in Cosmology

from production to prediction

Thomas Tram

PhD Dissertation



Supervisor: Prof. Steen Hannestad
July 2012

Department of Physics and Astronomy
Faculty of Science
Aarhus University
Denmark

Dark Relics in Cosmology

from production to prediction

A dissertation presented to the faculty of Science and Technology of Aarhus University in partial fulfillment of the requirements for the PhD degree.

by
Thomas Tram
July 2012

ACKNOWLEDGEMENTS

I would like to thank my advisor, Steen Hannestad, for granting me this opportunity to study Cosmology. Even after all the frustrating hours of debugging code, I am still deeply fascinated by this young and dynamic field of physics! I will also thank Steen for all our discussions about physics and cosmology, but also for encouraging non-physics discussions at lunch and in the Friday bar. This is something I have learned to value.

I would also very much like to thank Julien Lesgourgues for being an excellent host and collaborator during my year at CERN. He taught me not to be ashamed of my fascination with numerical methods, and working with Julien on hils Boltzmann code `CLASS` was a privilege. I also enjoyed working with Diego Blas, Emanuele Castorina and Benjamin Audren on different `CLASS`-related projects while I was there.

During my first four years of studies I benefited greatly from being part of a group of physics-students from my year who like me were enthusiastic about mathematics. Our discussions and the peer-pressure to learn more was very motivating for me and I thank them for that.

I would also like to thank my family for encouragement and support during both studies and research. This is much appreciated! I must also thank my girlfriend Gunvor Thinggaard Koldste for making me happy every single day. Thank you for listening to all my complaints about my codes (when they are not working), and for sharing my interest in physics - I am lucky to have you.

I would very much like to thank Jan Hamann for a careful proof-reading of this thesis, and for physics discussions before and during the preparation of the thesis. Thanks to Anne-Louise Gadsbølle for taking me out for a run almost everyday during the last week of writing, and thanks to her and Ole Søe for keeping me company during the last stressful days of writing. You helped me endure.

Finally, many thanks to the cosmology group in Aarhus for making me enjoy what I do. This is due to past members, Anders Basbøll, Martin Sloth, Troels Haugbølle, Jacob Brandbyge, Anna Sejersen Riis, Katrine Skovbo, as well as current members Jan Hamann, Ole Eggers Bjælde, Tobias Basse, Christian Schultz, Rasmus Sloth Hansen and Io Odderskov.

Summary

The Universe started out in a very hot and dense state 13.7×10^9 years ago, and it has expanded and cooled since then. Cosmology aims at explaining how our Universe has evolved from this Big Bang to what it is today. The evolution is governed by the fundamental laws of physics, so we have a synergistic relation: By observing the Universe today we may hope to learn something about the laws of physics, and investigating the laws of physics may tell us how the Universe evolved.

The scientific papers which form part of this thesis, aim at answering the following three questions:

1. How is dark matter produced in models with dark force mediators, and will the distortion of the cosmic microwave background (CMB) spectrum due to dark matter annihilations be visible? *Relativistic cross section can change by a factor of a few, and some part of the parameter space is ruled out. However, anisotropy bounds are stronger.*

2. How do we implement massive neutrinos in a linear Boltzmann code efficiently, and to what extent is it possible to model the neutrinos as a fluid with anisotropic stress? *The sampling of the momentum space can be optimised, and progress was made on the fluid approximation.*

3. How are sterile neutrinos (if they exist) produced in the early Universe, and can we block the production by a large lepton asymmetry? *The production can indeed be heavily suppressed by a large lepton asymmetry.*

These papers are enclosed in part III. Some background theory is derived in part I, while part II is devoted to the numerical methods that made the papers possible. They are hard to cover in any depth in the actual papers, but since they represent a great deal of the time I spent during my PhD, I feel that this is a good opportunity to describe them.

Resumé (Summary in Danish)

For 13,7 milliarder år siden befandt Universet sig i en meget varm og tæt tilstand, siden har det udvidet sig og blevet afkølet.

Kosmologi forsøger at forklare, hvordan vores Univers har udviklet sig siden Big Bang og frem til i dag. Udviklingen afhænger af de fundamentale naturlove, så der er en synergy: Ved at observere Universet i dag kan vi håbe at lære noget om de fundamentale naturlove, og ved at undersøge de fysiske love kan vi lære noget om Universets udvikling.

De videnskabelige artikler som udgør en del af denne afhandling sigter mod at besvare følgende tre spørgsmål:

1. Hvordan bliver mørkt stof produceret i modeller med mørke kraftbærere, og vil forandringerne i frekvensspektret for den kosmiske mikrobølge baggrund (CMB) kunne ses? *Det relativistiske tværsnit kan ændre sig med en faktor på et par stykker, og dele af parameterrummet kan udelukkes. Men grænserne fra CMB-anisotropierne er stærkere.*

2. Hvordan implementerer vi neutrinoer med masse effektivt i en lineær Boltzmann-kode, og i hvor høj grad kan vi approksimere neutrinoer som et fluid med anisotropt stress? *Valget af punkter i impulsrummet kan optimeres, og der blev gjort fremskridt med hensyn til fluid-approksimationen.*

3. Hvordan er sterile neutrinoer (hvis de eksisterer) produceret i det tidlige Univers, og kan vi forhindre dem i at blive produceret ved at introducere en stor lepton asymmetri? *Det viser sig at være muligt at blokere produktionen med en stor lepton asymmetri.*

Disse artikler er inkluderet i del III. I del I er der udledt noget af den teori der ligger til grund for artiklerne, mens del II er dedikeret de numeriske metoder der har gjort artiklerne mulige. Metoderne er svære at dække i detalje i faktiske artikler, men da de repræsenterer en væsentlig del af mit samlede tidsforbrug i løbet af min PhD, så følte jeg at dette er en god mulighed for at beskrive dem.

Outline and publication list

During my PhD I have worked on several topics that may seem unrelated. This is also the reason for my title "Dark Relics in Cosmology", which is really very general: Any particle species that is at most weakly coupled and is produced in the early Universe falls into this category. I will use this section to explain what the motivation was for each different topic to give a more coherent picture of the thesis.

Dark matter and Sommerfeld enhancement

I started out investigating indirect detection of dark matter, and specifically the possibility of Sommerfeld enhancement in light of the positron excesses observed in satellite experiments such as EGRET and PAMELA. It was claimed that the strongest bounds on these models came from energy injection during recombination but we wanted to consider the effect from distortions of the energy spectrum. It was clear that any analysis aiming at calculating the effect of the boost factor on the actual annihilation cross-section in a halo today, would have to take the freeze-out process into account. This work is included as chapter 7 and reference [1].

Stiff integrator

I worked on momentum dependant neutrino freeze-out for a while before I came to CERN to work with Dr. Julien Lesgourgues. During that time I got interested in how to solve large systems of stiff differential equations efficiently. I was encouraged to implement my solver, `ndf15`, into `CLASS`, the Boltzmann solver that Dr. Lesgourgues was working on, because he was interested in improving the tight-coupling-approximation. Using `ndf15` it was possible to evolve the perturbation equations without the tight-coupling approximation, which then provides an ideal reference for testing different approximation schemes. This was my primary contribution to this work, which is why I only included this paper as a reference [2].

Massive neutrinos in `CLASS`

After having applied `ndf15` to `CLASS`, we began working on the implementation of massive neutrinos in `CLASS`. This is much more complicated than the massless neutrino case since the mass prevents us from integrating out the momentum dependence in the distribution function. However, since each

momentum mode is uncoupled, we are free to take any distribution of momenta that we like. I realised that this freedom could be used to optimise the evaluation of integrals over the neutrino[1] distribution functions.

We also wanted to accommodate very non-thermal distributions without any other input from the user than the distribution function itself. Specifically, we did not want any hard-coded cutoffs in the momentum distribution, and we wanted the code to automatically locate and sample peaks in the distribution.

At the same time as I was implementing the quadrature scheme and the massive neutrino perturbation equations, Dr. Lesgourgues found a wonderful approximation scheme for massless neutrinos. It amounts to modelling the massless neutrinos as a fluid with anisotropic pressure and finding an evolution equation for the shear. I tried to adapt the approximation scheme to the massive neutrino case, but it was not as easy. The main problem is that the pressure perturbation is a free parameter in the massive case, and it is much harder to find the correct/optimal evolution equation for the shear. This work is included as chapter 8 and reference [3].

Quantum Kinetic Equations

The past year has seen a growing interest in Dark Radiation, and especially sterile neutrinos as explanation. This interest is driven by CMB measurements hinting towards $N_{\text{eff}} > 3$, a deficit in the neutrino flux from nuclear reactors compared to calculations, as well as a few other anomalies in neutrino oscillation experiments. We felt that it would be important to calculate the degree of thermalisation for a sterile species using the full Quantum Kinetic Equations (QKE). This was also the only way that we could include the effect of a non-zero lepton number in a fully consistent way. This work is included as chapter 9 and reference [4].

List of publications

[1] S. Hannestad and T. Tram, "Sommerfeld Enhancement of DM Annihilation: Resonance Structure, Freeze-Out and CMB Spectral Bound," *JCAP* **1101** (2011) 016, `arXiv:1008.1511 [astro-ph.CO]`.

[2] D. Blas, J. Lesgourgues, and T. Tram, "The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes," *JCAP* **1107** (2011) 034, `arXiv:1104.2933 [astro-ph.CO]`.

[3] J. Lesgourgues and T. Tram, "The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics," *JCAP* **1109** (2011) 032, `arXiv:1104.2935 [astro-ph.CO]`.

[4] S. Hannestad, I. Tamborra, and T. Tram, "Thermalisation of light sterile neutrinos in the early universe," *JCAP* **1207** (2012) 025, `arXiv:1204.5861 [astro-ph.CO]`.

[1] In the paper [3], we use the term non-cold dark relics (`ncdm`), but most of the time I will just refer to such species as massive neutrinos.

Contents

Part I

Cosmology

Chapter 1

Introduction

It was once believed that the physical laws valid on Earth were different from the laws governing the Heavens and the Cosmos. That was changed by Isaac Newton when he published his famous *"Philosophiae Naturalis Principia Mathematica"* [133, 134] in 1687.

The instantaneous action-at-a-distance of Newton's law of gravitation was a theoretical problem, and it was also in conflict with special relativity. This led Einstein to develop the general theory of relativity in 1917, and now it was possible to construct consistent theories describing our Universe. Another important step was the developments of particle physics during the 20^{th} leading to the Standard Model.

The observations of type 1a supernovae in the 1990's, the large scale structure surveys like 2dF and finally the observation of the CMB temperature anisotropies by balloon- and satellite-based experiments has made cosmology into what it is today: a testing ground for particle physics beyond the Standard Model.

1.1 Notation and conventions

The standard system of units is the SI-system, which defines 7 base units: metre (m), kilogram (kg), second (s), ampere (A), Kelvin (K), candela (cd) and mole (mol). We restrict ourselves to the first 5 units, and impose the 4 identities $\hbar = c = k_B = \epsilon_0 = 1$, where $\hbar$ is the reduced Planck constant, c is the speed of light in vacuum, k_B is the Boltzmann constant and ϵ_0 is the vacuum permitivity. This allows us to re-express any unit that can be derived from the first 5 SI base units in terms of a unit of our choice. As is usual in particle physics, we take GeV as our fundamental unit, and we will not write $\hbar$, c, k_B or ϵ_0 in any expression in this thesis.

We use the sign convention $(-,+,+,+)$ for a general metric $g_{\mu\nu}$ and for the flat metric $\eta_{\mu\nu}$. Greek indices run from 0 to 4, while latin indices runs from 1 to 3. Apart from the sign convention for the metric, we choose $G_{\mu\nu} = +8\pi G T_{\mu\nu}$ and $R_{\mu\nu} = +R^{\alpha}_{\mu\alpha\nu}$. This convention is widely used.

1.2 Four forces of nature

Four fundamental forces have been discovered so far: the strong force, the weak force, electromagnetism and gravity. The strong force is mediated by a massless spin-1 particle, called the gluon, and it is responsible for binding quarks into mesons and hadrons. Through the exchange of mesons, a small residual part of the strong force also binds nucleons into atomic nuclei.

The weak force is responsible for the decay of the neutron and for beta decay of radioactive nuclei. It is mediated by three massive spin-1 particles, the Z-boson and the $W^{\pm}$-bosons, and the mass of these bosons (91.1876GeV and 80.385GeV respectively) explains the weakness (or short range) of this force.

The electromagnetic force is mediated by a massless spin-1 particle, called the photon, exactly like the gluon in the strong force. However, contrary to gluons, photons are not confined, so the range of the electromagnetic interaction is infinite. The electromagnetic force is responsible for the formation of atoms and molecules, it makes nuclei unstable and it transfers energy from the Sun to the Earth, just to name a few examples.

Gravity is not like any of the other forces. It couples to anything that has mass or energy, and it is exceedingly weak: the gravitational attraction between a positron and an electron is more than 42 orders of magnitude weaker than the Coulomb attraction. The only reason that we have to care about gravity at all, is due to electric charge neutrality. The positive and negative charges are balanced to a very high degree in objects such as planets and galaxies, so gravity becomes the most important force at that scale.

1.3 Standard Models

The Standard Model of Particle Physics

The strong force, the weak force and the electromagnetic force are described by what is known as the Standard Model. It is a gauge theory with the gauge group $SU(3) \times SU(2)_L \times U(1)$, meaning that the force carriers are the generators of these groups and that all the matter content in the model (the particles) fall into different representations of this group. There are 8 generators of $SU(3)$ corresponding to the 8 gluons, 3 generators of $SU(2)$ corresponding to W^+, W^0 and W^-, and one generator of $U(1)$ corresponding to the B^0-boson. The $SU(2)_L \times U(1)$ symmetry is spontaneously broken by a vacuum expectation value of the Higgs field, thereby giving masses to the W^+ and W^- bosons. The remaining W^0 and B^0 bosons form a massive and a massless linear combination, which is the Z^0-boson and the photon respectively. The matter content of the Standard Model falls in different representations of the gauge group.

The Standard Model of Cosmology

The Universe has expanded from a very hot and dense initial state 13.7×10^9 years ago. To quantify the expansion, we introduce the scale factor a which measures the size of the Universe at a given time. The expansion rate depends

on the amount of radiation, the amount of matter and dark matter and the amount of dark energy. A combination of different observations has shown that we live in a flat Universe containing about 73% dark energy, 23% cold dark matter and 5% baryonic matter. The dark energy component has negative pressure, and for this reason it accelerates the expansion. The amount of radiation today is very small, only about $6 \times 10^{-3}\%$, but it once dominated the Universe. This is due to the scaling of the energy density with scale factor: for matter, the energy density is just inversely proportional to volume, i.e. $\propto a^{-3}$. The energy density in dark energy is constant[1] while radiation scale as a^{-4}. This scaling can be loosely understood as the combination of dilution of photons and red-shifting of individual photons by the expansion. Using our rough numbers, we can calculate the scale factor of equality, when matter and radiation had equal energy density:

$$28\% a_{\mathrm{eq}}^{-3} \sim 6 \times 10^{-3}\% a_{\mathrm{eq}}^{-4} \Rightarrow a_{\mathrm{eq}} \sim 2 \times 10^{-4}. \tag{1.1}$$

This means that when the Universe was less than ten thousand times smaller than today, it was completely dominated by radiation.

1.4 Beyond the Standard Models

I will briefly mention some of the most important problems in particle physics and cosmology. Any effort in cosmology or particle physics should aim at bringing us closer to answers to these problems. While only a few of these problems are directly related to the work I have done, I will list them all for completeness.

The cosmological constant problem

According to quantum field theory, any field would contribute of the order $\sim \Lambda^4$ to the energy of the vacuum, where Λ is the cutoff in the theory. Assuming the cutoff is the Planck scale m_{Pl}, one finds $\rho_{\mathrm{vac.}} \sim m_{\mathrm{Pl}}^4 \sim 10^{76}\mathrm{GeV}^4$, which is 122 orders of magnitude larger than the observed value of $\rho_{\mathrm{vac.}} \simeq 3 \times 10^{-47}\mathrm{GeV}^4$. This is known as a fine-tuning problem, because it requires us to put in a counter term which is tuned to give this low value. Nothing prevents us from doing so, but we believe that a more general theory should be able to explain this phenomenon.

Supersymmetry, the symmetry between fermions and bosons, would cancel any contribution to the cosmological constant. However, it is clearly broken at energy scales below $\sim 10^3\mathrm{GeV}$, so the best we could hope for is reducing the cutoff to this value, which would still give a contribution 59 orders of magnitude larger than the observed value. This is clearly something that we do not understand, and it is likely that we need a theory of quantum gravity to explain this phenomenon. Some implementation of the holographic principle might work, since it would limit the amount of possible micro states in the Universe.

[1] This is true in the Standard Model of cosmology, but there are many proposed models where dark energy is allowed to evolve.

The baryon asymmetry

The ratio of baryons to photons, $\eta_b = n_b/n_\gamma = 6.2 \times 10^{-10}$, is much larger than what the Standard Model predicts when starting from a baryon symmetric state. Thus, some mechanism is needed to create this asymmetry which is responsible for all baryonic matter in the Universe. Given a lepton asymmetry, the Standard Model can generate a baryon asymmetry by processes known as sphalerons. This scenario is called leptogenesis, and it can be realised in see-saw models that simultaneously explain the smallness of the neutrino mass. The new required physics in these models often resides at the energy scale of $\sim 10^{10}$GeV, so it is usually hard to probe these models using present experiments.

The gauge hierarchy problem

A scalar particle such as the Higgs boson will receive radiative mass corrections from any particle it couples to, and the mass corrections are quadratically divergent. For instance, the mass corrections from a fermions and a heavy complex scalar field are [140]

$$\Delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda^2 + \cdots \qquad \text{(Fermions)}, \tag{1.2}$$

$$\Delta m_H^2 = \frac{|\lambda_f|^2}{16\pi^2}\Lambda^2 - 2m_s^2 \ln\left(\frac{\Lambda}{m_s}\right) + \cdots \qquad \text{(Complex scalar)}, \tag{1.3}$$

where Λ is the cutoff. This is a fine-tuning problem, since we may believe that the very large positive and negative corrections from different parts of the theory cancel out to give us the observed mass of 126GeV. One approach to this problem is to invent a symmetry that cancels these contributions exactly from some energy scale. This is what happens in low energy supersymmetry. Another idea is to somehow lower the Planck scale to the weak scale, which is the approach in some models of extra dimensions.

Dark Matter

Dark matter makes up about 22% of the energy density in the Universe today, and its existence has been established beyond reasonable doubt by many different observations. If the dark matter species has a thermal distribution which is the case for the standard production scenario, the mass of the particle is bounded from below by observations of the large scale structure (LSS) of the universe: A light particle would form less matter at small scales because of free-streaming, and observations gives the limit $m_\xi > 1$keV. An upper limit is much harder to establish, but the lack of micro-lensing events from dark matter 'particles' in the Magellanic clouds gives the constraint $m_x < 10^{-7} M_\odot \sim 10^{57}$GeV [141].

The dark matter particle should be at most weakly interacting, should be stable on cosmological time-scales, and it should come together with a production mechanism that can explain the amount of dark matter we observe today. The solution to the gauge hierarchy problem could give us a nice dark matter candidate known as the WIMP, which is an acronym for Weakly

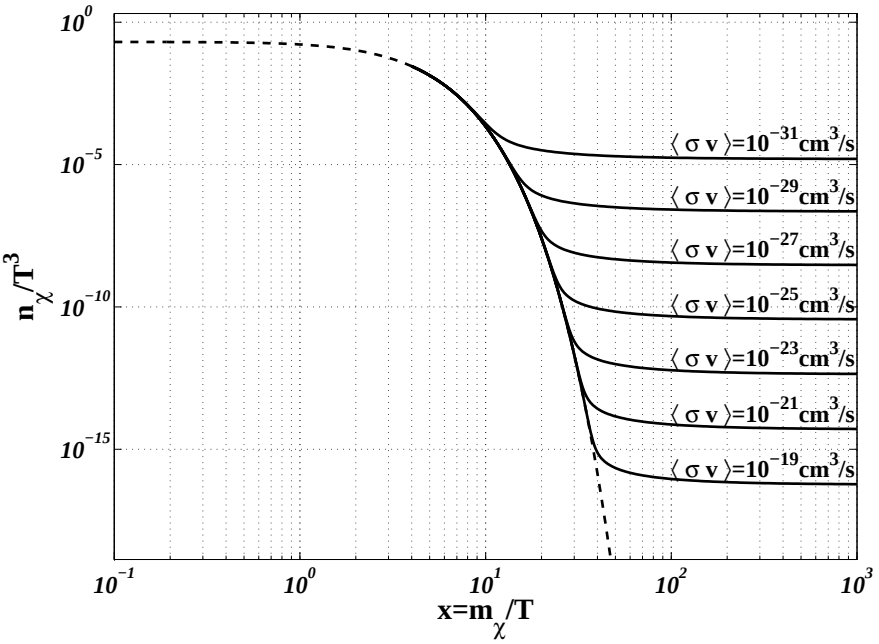


Figure 1.1: The freeze-out process for a particle species coupled to the plasma through s-wave annihilations for different thermally averaged cross sections. The dashed curve is the equilibrium number density, and the number density tracks the equilibrium density until a point $x \simeq 20$.

Interacting Massive Particle. The argument for the WIMP is that once we insert new physics at the TeV-scale to protect the Higgs-mass, we usually also needs to protect the standard model from baryon number violating operators that would lead to (too fast) proton decay. This symmetry, which is called R-parity in SUSY-models, also guarantees that the lightest of these new particles is stable. The production mechanism is also given: such a particle with weak gauge interaction would be produced thermally through freeze-out in roughly the right amount. This coincidence is called the WIMP miracle, but there are many other candidates. Figure 1.1 shows the abundance of a particle species today for different thermally averaged cross sections.

The strong CP problem

The strong CP problem is again a fine-tuning problem, this time in Quantum Chromo Dynamics (QCD), the $SU(3)$ part of the Standard Model. The problem is a certain $\bar{\theta}$-term in the QCD Lagrangian that breaks CP-symmetry in the quark sector. Broken CP-symmetry in the QCD-sector would induce an electric dipole moment on the neutron which has not been observed. The non-detection constrain this parameter $\bar{\Theta}$ to be less than 10^{-10} while we would expect $\bar{\Theta}$ to be of order 1. The offending term is [12]:

$$\mathcal{L}_{\text{QCD}} = \mathcal{L}_{\text{Perturbative}} + \bar{\Theta}\frac{g^2}{32\pi^2}G^{a\mu\nu}\tilde{G}_{a_\mu\nu}, \tag{1.4}$$

$$\bar{\Theta} = \Theta_{\text{vac.}} + \text{Arg}\left(\det\left(\mathcal{M}\right)\right), \tag{1.5}$$

where $\mathcal{M}$ is the quark mixing matrix. Usually $\mathcal{M}$ is written such that the determinant is real, and it is understood that the phase of the determinant has been absorbed into the $\bar{\Theta}$-term as shown above. $\Theta_{\text{vac.}}$ is a parameter due to non-trivial nature of the QCD vacuum.

The smallness of $\bar{\Theta}$ can be explained in an elegant way by introducing a global $U(1)_{\text{PQ}}$-symmetry, known as the Peccei-Quinn symmetry, which is spontaneously broken at some high energy scale. The scalar-field which is charged under this symmetry will acquire a vacuum expectation value such that $\bar{\Theta} = 0$. The spontaneous breaking of $U(1)_{\text{PQ}}$ leads to a light pseudo-Nambu-Goldstone boson known as the Axion. The Axion mass and its couplings are suppressed by $1/f_{\text{PQ}}$ where f_{PQ} is the energy scale where the Peccei-Quinn symmetry is broken. There are three distinct Axion production mechanisms in the early Universe: Axionic string decay, thermal production and misalignment production. The first one is suppressed by inflation as we shall see later, the second one could result in extra radiation, while Axions produced by the third mechanism are a cold dark matter candidate. Note that the lower mass bound mentioned early does not apply here, because the production in non-thermal.

Initial conditions for Big Bang

The temperature of the Cosmic Microwave Background (CMB) radiation is 2.725K in all directions, and the anisotropies are at the level of 10^{-5}. This might not sound too weird if the different patches of sky had been in thermal contact at some point, but in the standard Big Bang theory this was not the case. In this case it is hard to understand how thousands of causally disconnected patches of sky just happens to have the same temperature.

Another problem is the flatness of the Universe. From the Friedmann equation (2.9) we observe that the energy in curvature scales as $\propto a^{-2}$. Since observations show that "curvature energy" is completely negligible compared to matter density today, this must have been even more true in the early Universe - the Universe must have evolved from an initial state which was extremely flat and homogeneous. Both these problems can be solved if the early Universe went through an inflationary phase where it expanded exponentially. Such a phase can be realised by introducing a heavy scalar field known as the inflaton. Inflation also gives an explanation for the small anisotropies, because they can be generated as quantum fluctuations in the inflaton-field.

Unwanted relics

Looking at figure 1.1 we may wonder what happens to particles interacting even more weakly than a WIMP. If they are stable on cosmological time scales, they would dominate the Universe today in contradiction with observations. This is a problem for the gravitino, the superpartner of the graviton, in some SUSY-models.

Another kind of relic is topological defects such as magnetic monopoles and cosmological strings. They should produced in the different[2] phase-transitions

[2]The phase-transition producing magnetic monopoles is the hypothetical breaking of a the grand unified gauge group down to the $SU(3) \times SU(2)_L \times U(1)$ gauge group of the Standard Model.

that happens when the Universe cools. Such defects should be abundant, but they have not been observed.

This problem can also be cured by inflation. Depending on the energy scale of inflation compared to the energy scale of the phase-transition, inflation offers two ways out: If inflation happens before the phase-transition, the field governing the phase-transition becomes homogeneous at large scales thereby drastically reducing the number density of defects. If inflation happens after the phase transition, the exponential expansion quickly dilutes the produced defects, which again explains the non-observation.

Chapter 2

General Relativity

General Relativity (GR) is the relativistic theory of gravitation. It relates the local geometry of space-time to the energy content in the Universe in the form of the energy-momentum tensor.

2.1 Formalism in GR

Field equations

The effect of gravity on the Standard Model particles in the Universe is governed by the field equations of general relativity, which relates the local geometry of space-time in the form of the metric $g_{\mu\nu}$ to the energy-momentum tensor $T^{\mu\nu}$ of matter. The famous equation due to Einstein is

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu} \equiv \frac{8\pi}{m_{\mathrm{Pl}}^2}, \tag{2.1}$$

where $R_{\mu\nu}$ is the Ricci tensor and we have defined the Planck mass $m_{\mathrm{Pl}} = 1.2209 \times 10^{19}\mathrm{GeV}$. The Ricci tensor is a contraction of the Riemann curvature tensor, and it can be written in terms of the Christoffel symbols:

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} + \Gamma^{\alpha}_{\beta\alpha}\Gamma^{\beta}_{\mu\nu} - \Gamma^{\alpha}_{\beta\nu}\Gamma^{\beta}_{\mu\alpha}, \tag{2.2}$$

where $\phi_{,\alpha}$ denotes the derivative of ϕ with respect to x^{α}, i.e. $\phi_{,\alpha} \equiv \frac{\partial\phi}{\partial x^{\alpha}}$. The Christoffel symbols themselves can be calculated from the formula

$$\Gamma^{\mu}_{\alpha\beta} = \frac{g^{\mu\nu}}{2}\left[g_{\alpha\nu,\beta} + g_{\beta\nu,\alpha} - g_{\alpha\beta,\nu}\right]. \tag{2.3}$$

Action for GR

The field equations of GR can be derived from an action, and the action can be constructed using effective field theory arguments. We write a Lagrangian containing all operators which are compatible with the symmetry, which in the case of GR is diffeomorphism invariance. In vacuum, the lagrangian must then be a linear combination of curvature invariants with constant coefficients,

$$S = \int \mathrm{d}^4x^{\mu}\sqrt{-g}c_1R + c_2R^2 + c_3R_{\mu\nu}R^{\mu\nu} + \cdots, \tag{2.4}$$

where the square root of $-g$, the determinant of the metric, keeps the integral iteself invariant. Inspection of equation (2.2) and (2.3), reveals that R and $R_{\mu\nu}$ has mass-dimension M^2, since they are second derivatives of the metric. If we are interested in situations where $\sqrt{R} \ll m_{\rm Pl}$, we can truncate the expansion at the first term. This action is then called the Einstein-Hilbert action, and the variation of this with respect to the metric $g_{\mu\nu}$ gives the field equations (2.1). The constant c_1 can be found from comparing the weak field approximation of GR to Newton's law of gravitation, and it is $c_1 = m_{\rm Pl}^2/(16\pi)$.

2.2 The homogeneous and isotropic Universe

FLRW metric

The Cosmological Principle states that our vantage point in the Universe is not special in any way. This can be stated formally as a requirement on the spatial part of the metric: it must describe a homogeneous and isotropic space. Only three types of spaces fall in this category: three-dimensional Euclidean space, a three-dimensional sphere of constant positive curvature and a three-dimensional hyperbolic space of constant negative curvature. This was realised by Howard Percy Robertson and Arthur Geoffrey Walker [135–138], and the general form of the full 4 dimensional metric is given by

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x\nu = dt^2 - a^2(t)\left(\frac{dr^2}{1-Kr^2} + r^2 d\Omega^2\right), \tag{2.5}$$

where K is 0, 1 or -1 for Euclidean, spherical and hyperbolic space respectively. In these coordinates we have $x^\mu = (t, r, \theta, \phi)$. This metric is known as the FLRW-metric, after Alexander Friedmann and Georges Lemaître in addition to Robertson and Walker.

Perfect fluids

As a first approximation, we will assume all the energy content in the universe to be of the form of perfect fluids. That restricts the energy-momentum tensor $T^{\mu\nu}$ of each species to depend on only two parameters, the energy density ρ and the pressure p. In the rest frame of the fluid chosen to be locally flat, the energy-momentum tensor reads

$$T^\mu{}_\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}. \tag{2.6}$$

The general covariant form of $T^{\mu\nu}$ for a perfect fluid is then given by

$$T^{\mu\nu} = pg^{\mu\nu} + (p+\rho)u^\mu u^\nu, \tag{2.7}$$

where the 4-velocity has been normalised such that $u^\mu u^\nu g_{\mu\nu} = -1$. This expression matches equation (2.6) in the fluid rest frame, and since it is also general covariant, it is the correct expression.

Friedmann equations

It is convenient to calculate the Einstein tensor $G_{\mu\nu}$ using the diagonal metric (2.5). One finds

$$G_{00} = 3\left(\frac{\dot{a}}{a}\right)^2 + 3\frac{K}{a^2}, \tag{2.8}$$

where the dot denotes derivative with respect to time. Inserting G_{00} and $T_{00} = \rho$ into the Einstein equation (2.1) yields the Friedmann equation

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{K}{a^2} = \frac{8\pi G}{3}\rho. \tag{2.9}$$

The Einstein equation has one more independent equation for $a(t)$, but since we have calculated $G_{\mu\nu}$ in spherical coordinates we cannot simply equate the spatial components of $G_{\mu\nu}$ to the spatial components of (2.6). However, since the trace $g_{\mu\nu}F^{\mu_\nu}$ is an invariant, we have $G_{\mu\nu}g^{\mu\nu} = -\rho + 3p$. Alternatively, we can calculate the Jacobian for the transformation to spherical coordinates and express the energy-momentum tensor in spherical coordinates. Both approaches yield the same equation as they should, and it is called the acceleration equation

$$\ddot{a} = -\frac{4\pi G}{3}(\rho + 3p)a. \tag{2.10}$$

Energy conservation

The conservation of the energy-momentum tensor

$$T^{\mu\nu}{}_{;\nu} \equiv T^{\mu\nu}{}_{,\nu} + \Gamma^{\mu}_{\alpha\nu}T^{\alpha\nu} + \Gamma^{\nu}_{\alpha\nu}T^{\mu\alpha} = 0, \tag{2.11}$$

follows from the second Bianchi identity for the Riemann curvature tensor. It leads to the conservation equation

$$\dot{\rho} + \frac{\dot{a}}{a}(\rho + 3p) = 0, \tag{2.12}$$

but since it is a consequence of the Einstein equation (2.1), it is in fact not an independent equation. It is easily derived by differentiating equation (2.9) with respect to time and using equation (2.10) to eliminate $\ddot{a}$. Together with an equation of state $p(\rho)$, any two of equations (2.9), (2.10) and (2.12) yields a closed system.

2.3 Inhomogeneities

Perturbing the metric

We assumed that the energy distribution in the Universe was homogeneous and isotropic, which enabled us to derive the Friedmann equations. In reality the Universe started out with small fluctuations, which we think were generated by quantum fluctuations in the inflaton field in the very early Universe[1]. In

[1]The quantum fluctuations could have been generated by another field which was subdominant during inflation. Such a field is called a curvaton.

principle, we would have to abandon the Friedmann equations and go back to Einstens field equations for the metric $g_{\mu\nu}$. However, since the deviations from homogeneity and isotropy are small, we can consider the metric to be a small perturbation of the FLRW-metric. Assuming a flat geometry, we have

$$\mathrm{d}s^2 = (\bar{g}_{\mu\nu} + \delta g_{\mu\nu})\,\mathrm{d}x^\mu \mathrm{d}x^\nu, \tag{2.13}$$

where $\bar{g}_{\mu\nu}$ is the FLRW-metric and $\delta g_{\mu\nu}$ is the perturbation.

$\delta g_{\mu\nu}$ has 16 components, but since it is symmetric in indices we have only 10 independent degrees of freedom in $\delta g_{\mu\nu}$. The metric perturbation $\delta g_{\mu\nu}$ can be decomposed into four 3-scalars, two transverse (divergence-less) 3-vectors and a symmetric, transverse and trace-free 3-tensor. The advantage of doing this splitting, is that the evolution of each type of perturbations decouple to first order in perturbation theory. We take the first scalar to be the time-time component of the metric

$$\delta g_{00} = -2a^2\psi, \tag{2.14}$$

and then we decompose the space-time components δg_{0i} using the Helmholtz theorem

$$\delta g_{0i} = a^2\left(B_{,i} + S_i\right), \tag{2.15}$$

where S_i is transverse, i.e. $S^i_{,i} = 0$. The 9 remaining components[2] of δg_{ij}, are decomposed as [142]:

$$\delta g_{ij} = a^2\left(-2\phi\delta_{ij} + 2E_{,ij} + F_{i,j} + F_{j,i} + h_{ij}\right), \tag{2.16}$$

where F_i is again transverse. Counting the degrees of freedom, we have 4 from the scalars ψ, ϕ, B and E, 4 from the transverse vectors S_i and F_i, and 2 from h_{ij}. (A symmetric 3×3 matrix has 6 independent components, but 1+3 constraints are imposed by $h^i_i = 0$ and $h^i_{j,i} = 0$.) This gives 10 degrees of freedom, the same that we started out with.

Gauge freedom

Now we must face the issue (or blessing) of gauge freedom. General covariance implies that the physics is invariant with respect to change of coordinates. However, a change of coordinates could affect some of our degrees of freedom. Let us consider an infinitesimal coordinate transformation

$$x^\mu \to \tilde{x}^\mu = x^\mu + \xi^\mu. \tag{2.17}$$

The effect of the coordinate transformation can be stated as a transformation law for the field perturbation $\delta g_{\mu\nu}$ as

$$\delta g_{\mu\nu} \to \delta g_{\mu\nu,\gamma}\xi^\gamma - g_{\mu\nu}\xi^\mu{}_{,\gamma} - g_{\mu\gamma}\xi^\gamma{}_{,\nu}. \tag{2.18}$$

The coordinate transformation ξ^μ can be decomposed in the same way as we did $g_{\mu\nu}$ before. Take ξ^0 to be a 3-scalar, and let us do the Helmholtz decomposition of the remaining part,

$$\xi^i = \xi^i_T + \zeta^{,i}, \tag{2.19}$$

[2]There are 9 components, but only 6 independent degrees of freedom because of the symmetry.

where the vector ξ^i_T is transverse. Writing out $\delta\tilde{g}_{\mu\nu}$ in terms of ξ^0, ζ and ξ^i_T enables us to find the transformation laws for ψ, ϕ, B, E, F_i, S_i and h_{ij}. It turns out, that only ξ^0 and ζ contribute to the gauge-transformations of the scalars ψ, ϕ, B and E, while the vectors F_i and S_i transform according to ξ^i_T. h_{ij} does not transform and is thus gauge-invariant. One could hope that two of the scalar degrees of freedom in the metric could be eliminated by a clever choice of ζ and ξ^0. This is indeed possible as we shall see now.

Fixing the gauge

We have two different methods of dealing with this gauge-dependence: Either we form gauge-invariant physical combinations of our degrees of freedom, or we fix the gauge by using our freedom in choosing ξ^μ. We will do the latter. Considering only scalar perturbations, the metric before gauge-fixing is

$$\mathrm{d}s^2 = a^2\left[-(1+2\psi)d\tau^2 + 2B_{,i}\mathrm{d}x^i d\tau + ((1-2\phi)\delta_{ij} - 2E_{,ij})\,\mathrm{d}x^i\mathrm{d}x^j\right], \quad (2.20)$$

where τ denotes conformal time defined by $dt = a(\tau)d\tau$. A popular choice is the conformal Newtonian gauge, where we put $E = B = 0$ in equation (2.20). Another choice is the synchronous gauge, which is defined such that all perturbations are in the spatial part of the metric. Considering equation (2.20), this amounts to setting $\psi = B = 0$. We find the metrics

$$\mathrm{d}s^2 = a^2\left[-(1+2\psi)d\tau^2 + (1-2\phi)\,\delta_{ij}\mathrm{d}x^i\mathrm{d}x^j\right], \qquad \text{Newtonian} \quad (2.21)$$

$$\mathrm{d}s^2 = a^2\left[-d\tau^2 + ((1-2\phi)\delta_{ij} - 2E_{,ij})\,\mathrm{d}x^i\mathrm{d}x^j\right], \qquad \text{Synchronous} \quad (2.22)$$

where we are again only considering scalar perturbations. Using either metric, we can construct the perturbation to the Einstein tensor $\delta G_{\mu\nu}$. We will be working in the synchronous gauge for the remaining part of this chapter.

Perturbed energy momentum tensor

The energy momentum is a symmetric tensor, so like the metric it has 10 independent components. Following Ma & Bertschinger [61] we write those as

$$T^0{}_0 = \bar{T}^0{}_0 + \delta T^0{}_0 = -(\bar{\rho} + \delta\rho), \quad (2.23)$$

$$T^0{}_i = \bar{T}^0{}_i + \delta T^0{}_i = (\bar{\rho} + \bar{p})\,v_i, \quad (2.24)$$

$$T^i{}_j = \bar{T}^i{}_j + \delta T^i{}_j = (\bar{p} + \delta p)\,\delta^i{}_j + \Sigma^i{}_j, \quad (2.25)$$

where $\Sigma^i{}_j$ is traceless and $\bar{\rho}$ and $\bar{p}$ are the mean energy density and pressure of the fluid. $v^i \equiv \mathrm{d}x^i/d\tau$ is the coordinate velocity, so it can be decomposed into a transverse vector component and its divergence, θ. $\Sigma^i{}_j$ is traceless, but according to equation (2.16), it contains one more scalar degree of freedom which is the anisotropic stress, σ. Going to Fourier space, these two scalar degrees of freedom can be expressed as

$$(\bar{\rho} + \bar{p})\,\theta \equiv ik^j\delta T^0_j, \quad (2.26)$$

$$(\bar{\rho} + \bar{p})\,\sigma \equiv -\left(\hat{k}_i\hat{k}_j - \frac{1}{3}\delta_{ij}\right)\Sigma^i{}_j. \quad (2.27)$$

Continuity and Euler equation

The conservation of the energy-momentum tensor (2.11) applies for each uncoupled species. Defining the energy perturbation $\delta \equiv \delta\rho/\bar{\rho}$, we apply the conservation equation in Fourier space to the perturbed energy-momentum tensor. This gives the continuity and Euler equation of the species in the synchronous gauge [61]:

$$\delta' = -(1+w)\left(\theta + \frac{h'}{2}\right) - 3\frac{a'}{a}\left(\frac{\delta p}{\delta\rho} - w\right)\delta, \tag{2.28}$$

$$\theta' = -\frac{a'}{a}(1-3w)\,\theta - \frac{w'}{1+w}\theta + \frac{\frac{\delta p}{\delta\rho}}{1+w}k^2\delta - k^2\sigma, \tag{2.29}$$

where h and η are linear combinations of the metric scalar degrees of freedom in the synchronous gauge such that h is the trace of the spatial part of δg_{ij}. The prime denotes derivatives with respect to conformal time τ. For most species, $\delta p/\delta\rho$ is given by the adiabatic sound speed squared, and therefore it is only a function of the equation of state parameter w. This means that δp is not an independent perturbation. For massive neutrinos, this is not true, and the deviations of $\delta p/\delta\rho$ from the adiabatic sound speed can be very significant as we show in chapter 8.

If we know the anisotropic stress perturbation σ, equation (2.28) and (2.29) governs the evolution of the perturbations for an uncoupled species. This is the case for cold dark matter, for which $\sigma = 0$, and for baryons after they have decoupled from the photons. In general however, we must solve the Boltzmann equation for each species.

2.4 Boltzmann equation

The distribution function

The classical evolution of a particle species is in general given by the Boltzmann equation

$$\frac{df}{dt} = C\left[f\right]. \tag{2.30}$$

We write the distribution function f in terms of the variables $(\tau, x^i, q, \hat{n})$ where q is the size of the comoving momentum and $\hat{n}$ is its direction. The Bolztmann equation then becomes

$$\frac{\partial f}{\partial\tau} + \frac{\mathrm{d}x^i}{\mathrm{d}\tau}\frac{\partial f}{\partial x^i} + \frac{q}{\tau}\frac{\partial f}{\partial q} + \frac{\mathrm{d}n_i}{\mathrm{d}\tau}\frac{\partial f}{\partial n_i} = \left(\frac{\partial f}{\partial\tau}\right)_{\text{Coll.}}. \tag{2.31}$$

We write the distribution function as

$$f\left(\tau, x^i, q, \hat{n}\right) = f_0(q)\left[1 + \Psi\left(\tau, x^i, q, \hat{n}\right)\right], \tag{2.32}$$

where the where background distribution $f_0(q)$ depends on the freeze-out process. For a thermally produced species, we have

$$f_0(q) = \frac{g_s}{(2\pi)^3}\frac{1}{e^{\epsilon/T_0} \pm 1} \tag{2.33}$$

where (+) is for fermions and (−) is for bosons. T_0 is the redshifted freeze-out temperature today, and ϵ is the energy given by $\epsilon \equiv \sqrt{q^2 + a^2 m^2}$. In `CLASS` and in chapter 8, we divide both the comoving momentum q and the mass m by T_0 to make them dimensionless. This does not change the equations of motions since it is just a rescaling of energy, but a factor of T_0^{-d} must be remembered in the integrals over q. We will not do the rescaling in this chapter.

Evolution equation for Ψ

Using the definition of $\Psi\left(\tau, x^i, q, \hat{n}\right)$ in equation (2.32), the Boltzmann equation (2.31) can be rewritten as an evolution equation for the perturbation $\Psi\left(\tau, k^i, q, \hat{n}\right)$, now written in Fourier space. In the synchronous gauge, it is [61]

$$\frac{\partial \Psi}{\partial \tau} + \mathrm{i}\frac{q}{\epsilon}(\boldsymbol{k}\cdot\hat{n})\Psi + \frac{\mathrm{d}\ln f_0}{\mathrm{d}\ln q}\left[\eta' - \frac{h' + 6\eta'}{2}(\hat{k}\cdot\hat{n})^2\right] = \frac{1}{f_0}\left(\frac{\partial f}{\partial \tau}\right)_{\text{Coll.}} \tag{2.34}$$

For a massless neutrino we have $\epsilon = q$, and the momentum dependence in equation (2.34) can be integrated out. The same is true for photons because the linearised collision operator is independent of momentum. We will now consider the case of massive neutrinos.

Because the Boltzmann equation (2.34) depends on $\hat{n}$ only through its angle with $\boldsymbol{k}$, the perturbation Ψ will stay axial-symmetric if it was so initially. Assuming an axial-symmetric initial state, we can expand the perturbation in Legendre series

$$\Psi\left(\tau, x^i, q, \hat{n}\right) = \sum_0^\infty (-1)^l (2l+1)\Psi_l\left(\tau, q, \boldsymbol{k}\right) P_l\left(\hat{k}\cdot\hat{n}\right). \tag{2.35}$$

The perturbations of the energy-momentum tensor are given in terms of Ψ by

$$\delta\rho = a^{-4}\int q^2 \mathrm{d}q\mathrm{d}\Omega\epsilon f_0(q)\Psi, \tag{2.36a}$$

$$\delta p = \frac{1}{3}a^{-4}\int q^2 \mathrm{d}q\mathrm{d}\Omega\frac{q^2}{\epsilon} f_0(q)\Psi, \tag{2.36b}$$

$$\delta T^0_i = a^{-4}\int q^2 \mathrm{d}q\mathrm{d}\Omega q n_i f_0(q)\Psi, \tag{2.36c}$$

$$\Sigma^i_{\ j} = a^{-4}\int q^2 \mathrm{d}q\mathrm{d}\Omega\epsilon\left(n_i n_j - \frac{1}{3}\delta_{ij}\right) f_0(q)\Psi. \tag{2.36d}$$

The Boltzmann hierarchy

Using the definitions of the scalar perturbations θ and σ in equation (2.26) and (2.27), we can insert the Legendre expansion of Ψ into equations (2.36) and perform the angular part of the integrals. Because of the orthogonality of

the Legendre polynomials, only one multipole survives in each integral:

$$\delta\rho = 4\pi a^{-4} \int q^2 \mathrm{d}q \epsilon f_0(q) \Psi_0, \tag{2.37a}$$

$$\delta p = \frac{4\pi}{3} a^{-4} \int q^2 \mathrm{d}q \frac{q^2}{\epsilon} f_0(q) \Psi_0, \tag{2.37b}$$

$$(\bar{\rho} + \bar{p})\,\theta = 4\pi k a^{-4} \int q^2 \mathrm{d}q q f_0(q) \Psi_1, \tag{2.37c}$$

$$(\bar{\rho} + \bar{p})\,\sigma = \frac{8\pi}{3} a^{-4} \int q^2 \mathrm{d}q \frac{q^2}{\epsilon} f_0(q) \Psi_2. \tag{2.37d}$$

Inserting the Legendre expansion into the Boltzmann equation (2.34) in the synchronous gauge finally yields the Boltzmann hierarchy

$$\Psi_0' = -\frac{qk}{\epsilon}\Psi_1 + \frac{1}{6}h' \frac{\mathrm{d}\ln f_0}{\mathrm{d}\ln q}, \tag{2.38a}$$

$$\Psi_1' = \frac{qk}{3\epsilon}\left(\Psi_0 - 2\Psi 2\right), \tag{2.38b}$$

$$\Psi_2' = \frac{qk}{5\epsilon}\left(2\Psi_1 - 3\Psi_3\right) - \left(\frac{1}{15}h' + \frac{2}{5}\eta'\right)\frac{\mathrm{d}\ln f_0}{\mathrm{d}\ln q}, \tag{2.38c}$$

$$\Psi_l' = \frac{qk}{(2l+1)\epsilon}\left(l\Psi_{l-1} - (l+1)\Psi_{l+1}\right), \qquad l \geq 3. \tag{2.38d}$$

This set of equations is the starting point of chapter 8.

Chapter 3

Sterile neutrinos

Sterile neutrinos are spin-$\frac{1}{2}$ fermions that has no standard model interactions; they are not charged under the $SU(3) \times SU(2)_L \times U(1)$ gauge group of the standard model. But in our definition of a sterile neutrino we will also make the requirement, that at least one active neutrino has a non-zero probability of oscillating into a sterile neutrino. In this chapter I will use α as a flavour index and i to index mass eigenstates.

3.1 Neutrino oscillations

In this section I will discuss oscillations among ordinary active neutrinos. This will set the stage for the more complicated formalism introduced in section 3.2.

Neutrino oscillations in vacuum

A neutrino in flavour eigenstate $|\nu^\alpha\rangle$ at time t can be written as a sum over mass eigenstates $|\nu^i\rangle$

$$|\nu^\alpha\rangle_t = \sum_i U_{\alpha i} e^{-\mathrm{i}E_i t} |\nu^i\rangle, \tag{3.1}$$

where $U_{\alpha i}$ is a unitary matrix that diagonalises the mass matrix

$$U^\dagger \mathcal{M} U = m^2_{\text{diag}} = \text{diag}\left[(m_1^2,\ m_2^2,\ \ldots)\right]. \tag{3.2}$$

We can expand the energy E_i in the small mass m,

$$E_i = \sqrt{p^2 + m^2} \simeq p + \frac{m_i^2}{2p} \simeq p + \frac{m_i^2}{2E}. \tag{3.3}$$

Taking e^{-ipt} outside, equation (3.1) can now be written as

$$|\nu^\alpha\rangle_t \simeq e^{-\mathrm{i}pt} U e^{-\mathrm{i}\frac{m^2_{\text{diag}}}{2E}t} U^\dagger U |\nu^i\rangle, \tag{3.4}$$

$$\simeq e^{-ipt} U e^{-\mathrm{i}\frac{m^2_{\text{diag}}}{2E}t} U^\dagger U |\nu^i\rangle, \tag{3.5}$$

$$\simeq e^{-ipt} e^{-\mathrm{i}\frac{1}{2E}\mathcal{M}t} |\nu^\beta\rangle, \tag{3.6}$$

showing that $|\nu^\alpha\rangle$ is a solution to the Schrödinger equation

$$i\frac{\mathrm{d}}{\mathrm{d}t}|\nu^\alpha\rangle_t \simeq \frac{1}{2E}\mathcal{M}|\nu^\alpha\rangle. \tag{3.7}$$

Two generation mixing

In the two generation mixing case, we may take U to be a rotation matrix

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{3.8}$$

In this case we can write

$$\mathcal{M} = U m^2_{\text{diag}} U^\dagger = \frac{m_1^2 + m_2^2}{2} + \frac{m_2^2 - m_1^2}{2} \begin{pmatrix} -\cos 2\theta & \sin 2\theta \\ \sin 2\theta & \cos 2\theta \end{pmatrix}, \tag{3.9}$$

and inserting $\mathcal{M}$ from equation (3.9) into equation (3.6) then yields the useful result:

$$|\nu^\alpha\rangle_t = \begin{pmatrix} \cos\frac{\delta m^2}{4E}t - i\sin\frac{\delta m^2}{4E}t\cos 2\theta & -i\sin\frac{\delta m^2}{4E}t\sin 2\theta \\ -i\sin\frac{\delta m^2}{4E}t\sin 2\theta & \cos\frac{\delta m^2}{4E}t + i\sin\frac{\delta m^2}{4E}t\cos 2\theta \end{pmatrix}, \tag{3.10}$$

where we have defined $\delta m^2 \equiv m_2^2 - m_1^2$ and discarded a phase factor. The probability for an initial flavour state of type $|\nu^e\rangle = (1,0)^T$ to be observed in the flavour state $|\nu^\mu\rangle = (0,1)^T$ after a time t can then be found from the matrix elements in equation (3.10). We find:

$$P_{\nu^e \to \nu^\mu} = |\langle \nu^\mu | \nu^e \rangle|^2 \tag{3.11}$$

$$= \sin^2 2\theta \sin^2\left[\frac{\delta m^2}{4E}t\right]. \tag{3.12}$$

To get a better feeling for the scales involved, it is instructive to reinstate factors of c and $\hbar$ and express the result in SI-units. Using the well known relation [energy] = [mass]c^2 and the time-energy relation [energy] $\times$ [time] = $\hbar$, we make the argument in the sine dimensionless:

$$P_{\nu^e \to \nu^\mu} = \sin^2 2\theta \sin^2 \frac{\delta m^2 c^4}{4E\hbar}t \tag{3.13}$$

$$= \sin^2 2\theta \sin^2 \frac{\delta m^2 c^3 L}{4E\hbar} \tag{3.14}$$

$$\simeq \sin^2 2\theta \sin^2\left[1.267 \frac{\delta m^2}{\text{eV}^2}\frac{L}{\text{km}}\frac{\text{GeV}}{E}\right]. \tag{3.15}$$

MSW effect

The last section did not include any effect of the medium. Inside the Sun, in supernovae and in the early Universe, the medium can not be ignored. We will now consider the effect of the solar medium on the propagation of electron neutrinos. Since the Sun is filled with electrons, there is a contribution to the electron-neutrino energy given by $\sqrt{2}G_F n_e$, where n_e is the electron number density. Adding the contribution to the e-e part of $\mathcal{M}$ in equation (3.7) yields the Hamiltonian

$$\mathcal{H} = \begin{pmatrix} -\frac{\delta m^2}{4E}\cos 2\theta + \sqrt{2}G_F n_e & \frac{\delta m^2}{4E}\sin 2\theta \\ \frac{\delta m^2}{4E}\sin 2\theta & \frac{\delta m^2}{4E}\cos 2\theta \end{pmatrix}. \tag{3.16}$$

The eigenvalues $\lambda_{1/2}$ of $\mathcal{H}$ are

$$4E\lambda_{1/2} = A \mp \sqrt{(A - \delta m^2 \cos 2\theta)^2 + (\delta m^2) \sin^2 2\theta}, \tag{3.17}$$

where $A = 2\sqrt{2}EG_F n_e$. We denote the eigenvectors corresponding to these eigenvalues $|\tilde{\nu}^1\rangle$ and $|\tilde{\nu}^2\rangle$. The change of basis is unitary, so we can still find some rotation matrix such that

$$\begin{pmatrix} \nu^e \\ \nu^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_m & \sin\theta_m \\ -\sin\theta_m & \cos\theta_m \end{pmatrix} \begin{pmatrix} \tilde{\nu}^1 \\ \tilde{\nu}^2 \end{pmatrix} \tag{3.18}$$

The effect of the medium can then be described as an effective in-medium mixing angle θ_m defined by

$$\cos(2\theta_m) = \frac{-A + \delta m^2 \cos 2\theta_m}{\sqrt{(A - \delta m^2 \cos 2\theta)^2 + (\delta m^2)^2 \sin^2 2\theta}}, \tag{3.19}$$

and it is evident that a resonance occurs if $A = A_{\text{crit.}} \equiv \delta m^2 \cos 2\theta$. An electron-neutrino produced at $A(n_e) > A_{\text{crit.}}$ and propagating into a region $A(n_e) < A_{\text{crit.}}$, can be completely converted to muon-neutrinos in the case of a small vacuum mixing angle. This is called the MSW effect and this could have solved the solar neutrino problem, had the mixing angle turned out to be small.

3.2 Quantum Kinetic Equations

The simple approach of the previous section can not be used if we are interested in a statistical ensemble of neutrinos coupled to a background plasma. If we were not interested in neutrino oscillations, we should just solve the Boltzmann equation for neutrino distribution function. However, if we are interested in both oscillations and scattering, we must solve a Boltzmann-like equation for the density matrices [67]. This system of equations is called the Quantum Kinetic Equations (QKE).

The $1+1$ scenario

We will now consider the system of chapter 9 consisting of one active and one sterile neutrino. The density matrices are Hermitian, so they have 4 independent degrees of freedom. They are conveniently expressed in terms of Pauli-matrices $\boldsymbol{\sigma}$:

$$\rho = \frac{1}{2} f_0 (P_0 + \boldsymbol{P} \cdot \boldsymbol{\sigma}), \qquad \overline{\rho} = \frac{1}{2} f_0 (\overline{P}_0 + \overline{\boldsymbol{P}} \cdot \boldsymbol{\sigma}), \tag{3.20}$$

where $\boldsymbol{P}$ and $\overline{\boldsymbol{P}}$ are real vectors and $f_0 = (e^x + 1)^{-1}$ is the un-normalised Fermi-Dirac distribution with zero chemical potential for comoving momentum

$x = p/T$. The evolution equations for P_0 and $\boldsymbol{P}$ are given by

$$\frac{\mathrm{d}P_0}{\mathrm{d}t} = \Gamma\left[\frac{f_{\mathrm{eq}}}{f_0} - \frac{1}{2}\left(P_0 + P_z\right)\right], \tag{3.21a}$$

$$\frac{\mathrm{d}P_x}{\mathrm{d}t} = -V_z P_y, \tag{3.21b}$$

$$\frac{\mathrm{d}P_y}{\mathrm{d}t} = V_z P_x - V_x P_z, \tag{3.21c}$$

$$\frac{\mathrm{d}P_z}{\mathrm{d}t} = V_x P_y, \tag{3.21d}$$

where the potentials V_x and V_z are [94]

$$V_x = \frac{\delta m_s^2}{2xT}\sin 2\theta_s, \tag{3.22a}$$

$$V_z = V_0 + V_1 + V_L, \tag{3.22b}$$

$$V_0 = -\frac{\delta m_s^2}{2xT}\cos 2\theta_s, \tag{3.22c}$$

$$V_1^{(a)} = -\frac{7\pi^2}{45\sqrt{2}}\frac{G_F}{M_Z^2}xT^5\left[n_{\nu_a} + n_{\bar{\nu}_a}\right]g_a, \tag{3.22d}$$

$$V_L = \frac{2\sqrt{2}\zeta(3)}{\pi^2}G_F T^3 L^{(a)}. \tag{3.22e}$$

In these potentials we have introduced the number density of neutrinos n_{ν_a} and anti-neutrinos $n_{\bar{\nu}_a}$, which are normalised such that they are 1 in thermal equilibrium. For muon- and tau neutrinos we have no charged current (CC) interactions, because muons and taus have annihilated at the temperatures we are concerned about. Electrons and positrons remains so for electron neutrinos we have charged current interactions in addition to neutral current (NC) interactions. This is take into account by g_a. We have:

$$g_e = 1 + \frac{4}{\cos^2\theta_W\left(n_{\nu_e} + n_{\bar{\nu}_e}\right)}, \qquad \text{CC+NC} \tag{3.23}$$

$$g_{\mu,\tau} = 1, \qquad \text{NC} \tag{3.24}$$

Lepton asymmetry

The effective lepton number $L^{(a)}$ is in general a combination of the different lepton numbers L_f, see equation (9.12)-(9.14). The effective lepton number in our system evolves only due to L_{ν_a}, which changes because of active-sterile oscillations. We approximate the scattering kernel by an effective equilibration rate $\Gamma = C_a G_F^2 x T^5$, where $C_e \simeq 1.27$ and $C_{\mu,\tau} = 0.92$. Unfortunately this breaks lepton number conservation, so to circumvent this problem we derive an evolution equation for L which does not explicitly depend on Γ. From the

definition of L_f we find:

$$L^{(a)} = \frac{2}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 \left(\rho_{aa} - \bar{\rho}_{aa}\right), \tag{3.25}$$

$$= \frac{1}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 f_0 \left(P_0 - \bar{P}_0 + P_z - \bar{P}_z\right) \Rightarrow \tag{3.26}$$

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \frac{1}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 f_0 \left(\frac{\mathrm{d}P_z}{\mathrm{d}t} - \frac{\mathrm{d}\bar{P}_z}{\mathrm{d}t}\right), \tag{3.27}$$

$$= \frac{1}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 f_0 V_x \left(P_y - \bar{P}_y\right). \tag{3.28}$$

The damping term D is given by $D \simeq \Gamma$ to a good approximation. This is the term responsible for destroying coherence and it is easy to see that it appears as a damping factor in the equations for the off-diagonal elements parametrised by P_x and P_y. The equilibrium distributions f_{eq} and $\bar{f}_{\mathrm{eq}}$ are the Fermi-Dirac distributions with a finite chemical potential ξ calculated from the neutrino lepton number. ξ is given by the equation

$$L^{(a)}_{\mathrm{eq}} = \frac{1}{4\zeta(3)} \int_0^\infty \mathrm{d}x\ x^2 \left[\frac{1}{1+e^{x-\xi}} - \frac{1}{1+e^{x+\xi}}\right] = \frac{1}{12\zeta(3)} \left(\pi^2\xi + \xi^3\right). \tag{3.29}$$

The real root of this equation can be expressed entirely using real functions. It is given by the Chebyshev cube root, and we find

$$\xi = \frac{-2\pi}{\sqrt{3}} \sinh\left(\frac{1}{3}\mathrm{arcsinh}\left[-\frac{18\sqrt{3}\zeta(3)}{\pi^3} L^{(a)}\right]\right). \tag{3.30}$$

MSW-like resonances

This system exhibits resonances as we saw with the MSW effect before. The resonance condition is $V_z = 0$ for particles and $\bar{V}_z = 0$ for anti-particles. Let us do a quick comparison between the two. Using $E \simeq p = xT$, the MSW resonance condition reads

$$2\sqrt{2}EG_F n_e - \delta m^2 \cos 2\theta = 0, \tag{3.31}$$

$$\sqrt{2}G_F n_e - \frac{\delta m^2}{2xT} \cos 2\theta = 0, \tag{3.32}$$

$$(\sim V_L) + V_0 = 0. \tag{3.33}$$

Here, n_e is the actual electron number density. In the early Universe, L may be small, but the second-order neutrino-neutrino scattering is important. Knowing where the resonances reside in momentum space is important, both for numerical work and for understanding the evolution of the system qualitatively. There are many signs involved: the sign of δm_s^2, the sign of L, and the sign of V_L which determines if it is a particle or anti-particle resonance. Too

keep track of these signs, we find the following notation convenient:

$$m \equiv \text{sign}[\delta m_s^2] \tag{3.34}$$

$$\ell = \begin{cases} \text{sign}[L^{(a)}] & \text{for particles} \\ -\text{sign}[L^{(a)}] & \text{for anti-particles} \end{cases} \tag{3.35}$$

The resonance conditions ($V_z = 0$ and $\bar{V}_z = 0$) can be written as

$$\widetilde{V_1} x^2 + \ell \left| V_L \right| x + \widetilde{V_0} = 0, \tag{3.36}$$

which has the solution

$$x_{\text{res}} = x_0 \left[A\ell \pm \sqrt{A^2 - m} \right] \equiv x_0 F_{\ell m}^{\pm}(A), x_0 \qquad = \sqrt{\frac{m\widetilde{V_0}}{\widetilde{V_1}}}, \tag{3.37}$$

$$A = \frac{|V_L|}{2\sqrt{m\widetilde{V_0}\widetilde{V_1}}}, \tag{3.38}$$

$$F_{\ell m}^{\pm}(A) = \left[A\ell \pm \sqrt{A^2 - m} \right]. \tag{3.39}$$

The physical solutions are $x_{\text{res}} \geq 0$, which is equivalent to $F_{\ell m}^{\pm}(A)$ because x_0 is positive. In the inverted hierarchy, we have $m = -1$ and only $F_{\pm 1,-1}^{+}(A)$ is positive. This is independent of A, and we have one resonance for particles and one for resonance. If $L > 0$, the anti-particle resonance is to the left of x_0 and the particle resonance is to the right. In the normal hierarchy, the resonance can only occur if we have an asymmetry. The positivity of $F_{\ell m}^{\pm}(A)$ requires us to take $F_{+1,+1}^{\pm}(A)$. Even then, we still require $A \geq 1$ for the square root to be real. In this case, both resonances belongs to either particles or anti-particles depending on the sign of L. If $L > 0$ they are both particle resonances, and they will work to remove the asymmetry.

Part II

Numerical methods

The importance of numerical methods

Good numerical methods are very important in cosmology today, but the details of them are often omitted in papers for clarity and to emphasise the physics. This limits the diffusion of good numerical methods and it makes it harder to reproduce and test results of others. I will take this chance to document some of the methods I have found useful. Chapter 4 describes methods for evaluating integrals which were especially useful for the project of chapter 8. Chapter 5 deals with the numerical solution of ordinary differential equations with emphasis on stiffness and it introduces two solvers that I have wrote, `ndf15` and `RADAU5`. Chapter 6 contains a bit about sparse methods for solving the linear equations in `ndf15` and `RADAU5`.

Stiffness occur frequently in cosmology because the Hubble time scale is long compared to time scales set by particle physics. One example is the freeze-out calculation of chapter (7). Initially, the dark matter species is kept in thermal equilibrium with the background plasma by weak interactions, but at some point it decouples. (See figure 1.1.) The evolution of the dark matter abundance in this regime is smooth on Hubble time scales, but even so a non-stiff solvers will take exceedingly small time steps[1]. This behaviour is explained in section 5.6.

The tight-coupling regime in Boltzmann codes are equivalent to the (somewhat simpler) dark matter case, and as we demonstrated in [2], a Runge-Kutta method takes many orders of magnitude more time steps than a stiff method. If you understand that the physics well, you will understand that you are following an equilibrium solution, and you may be able to approximate the evolution by not evolving the equilibrium solution.

[1] We can cheat by imposing our initial conditions just after the dark matter decouples, but we will introduce a small mistake in doing so. The correct approach is of course to take advantage of a stiff solver!

Chapter 4

Quadrature

Quadrature is the numerical evaluation of integrals.

4.1 Basics

The numerical estimation of a 1-dimensional integral can always be written as a finite sum

$$\int_a^b \mathrm{d}x f(x) \simeq \sum_{i=1}^{N} w_i f(x_i), \tag{4.1}$$

where x_i are the nodes and w_i are the weights. Usually it is not more expensive to evaluate $f(x)$ successively rather than once for all x_i, but note that this is not the case when the values $f_i \equiv f(x_i)$ are found by solving a differential equation for the coupled f_i's as is the case in chapter 8.

In the usual case then, it is possible to adaptively refine the integral by adding more points until the integral has converged. Adding more points can be done in two ways: either we can divide the interval into a number of pieces and use our method recursively, or we can choose a higher order method using more points. In both cases, we have to choose how to make use of our freedom in selecting the nodes x_i.

We have freedom in the choice of the nodes which we should take advantage of. If we are using an adaptive scheme, we could place the points such that they will coincide with the next iteration of the method. If we are using a N-point method, and we subdivide the interval into 2, we only need to calculate $f(x)$ on N new points compared to $2N$ points. However, we could also use our freedom to optimise our estimate of the integral with respect to a certain family of functions. This approach is known as Gaussian quadrature, and we will now consider a few different methods.

4.2 Gaussian quadrature rules

Let us consider an N-point method. It gives us $2N$ free parameters, the nodes and the weights. What is the optimal choice of these nodes and weights? We are often interested in integrals of the form

$$\int_a^b \mathrm{d}x f(x) = \int_a^b \mathrm{d}x W(x) g(x), \tag{4.2}$$

where $g(x)$ is well approximated by polynomials, and $W(x)$ is some weight function. Judged by the number of free parameters, it seems likely that a clever choice of N nodes and N weights should make it possible to integrate $f(x)$ exactly for $g(x)$ being a polynomial of degree $2N-1$ or less. Without loss of generality, we can restrict ourselves to the domain $[0;1]$. For the weight function $W(x)=1$, this is equivalent to solving the following non-linear algebraic system of equations:

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1^1 & x_2^1 & \cdots & x_N^1 \\ \vdots & & \ddots & \vdots \\ x_1^{2N-1} & x_2^{2N-1} & \cdots & x_N^{2N-1} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} \int_0^1 \mathrm{d}x x^0 W(x) \\ \int_0^1 \mathrm{d}x x^1 W(x) \\ \vdots \\ \int_0^1 \mathrm{d}x x^{2N-1} W(x) \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \\ \vdots \\ \frac{1}{2N-1} \end{bmatrix}$$

where we have exploited the linearity of the integral. This system can be solved using Newtons method for $N \lesssim 6$, but after that the system becomes too non-linear. However, there is a much better way of finding the nodes and weights for a given weight function. Given a weight function $W(x)$, we can define an inner product on the interval of interest by

$$\langle f|g\rangle \equiv \int_a^b \mathrm{d}x W(x) f(x) g(x), \tag{4.3}$$

and we search for a family of polynomials that have exactly one polynomial $p_j(x)$ of each order $j = 0, 1, \ldots$ and which are orthogonal with respect to the inner product (4.3). It then turns out, that the nodes for the N-point quadrature rule with respect to the weight function $W(x)$ are the N zero-points of $p_N(x)$, and the corresponding weights are [139]

$$w_i = \frac{\langle p_{N-1}|\rangle p_{N-1}\rangle}{p_{N-1}(x_i) p'_N(x_i)}, \tag{4.4}$$

where the polynomials have been assumed monic[1]. For the case of $W(x)=1$ and the domain $[-1;1]$, the polynomials are Legendre polynomials and the quadrature rule is called Gauss-Legendre quadrature. Another relevant case is $W(x) = x^\alpha \exp(-x)$ on the domain $[0;\infty]$, where the polynomials are Laguerre polynomials and the quadrature rule is named accordingly.

4.3 Adaptive methods

If we are in the situation where we can evaluate the integrand at will, it is advantageous to use an adaptive method. First off, we should have some measure of the error we are doing in using a specific quadrature rule. This is often accomplished by having a different quadrature rule which uses a subset of the points of the original quadrature rule, and the error is then estimated as some function of the difference between the two methods. If this error is larger than our goal, we can either try again using a higher order rule, or we can subdivide the interval and call the same method recursively on the sub-intervals.

[1] A monic polynomial is a polynomial where the coefficient of the highest power is 1.

A very good method is the 15-point Gauss-Kronrod quadrature method which has an embedded 7-point Gauss-Legendre rule. The 15-point formula is based on Stieltjes polynomials and the 7 point method is based on Legendre polynomials. An adaptive method based on this rule can be coded in a very compact and transparent way using recursion. Adaptive methods are efficient in resolving structure in the integrand, but as mentioned before, they are only available if the function evaluations can be done successively. Sec. 8.2 in chapter 8 explains how to take advantage of adaptive methods even in the case where nodes must be selected in advance.

The standard reference for numerical methods in physics is the excellent book by Press et al. [139], which is freely available online. However, one should not blindly trust any recommendation from the book: for instance, they strongly recommend Romberg integration and gives the example integral

$$\int_0^2 \mathrm{d}x x^4 \log\left(x + \sqrt{x^2+1}\right) = \frac{1}{75}\left[8 - 40\sqrt{5} + 480\sinh^{-1}(2)\right], \tag{4.5}$$

which is computed with a target relative error of 10^{-6}. The Romberg method (in double precision) converges after 17 calls to the integrand, and the relative error compared to the exact value is 3.1×10^{-8}. This is a decent performance, but note that it is a rather complex code involving a polynomial extrapolation. The above mentioned Gauss-Kronrod method converges in the first step after 15 calls to the integrand, and the embedded error estimate reports an error of 8.2×10^{-10}. The actual relative error is 2.0×10^{-14}, close to machine precision. This performance comes without any effort, since it is exceedingly easy to code. It is very robust since it does not depend on polynomial extrapolation, and it only evaluates points strictly inside the integration interval.

Chapter 5

Ordinary differential equations

In this section, I will describe some aspects of the numerical solution of ordinary differential equations (ODE).

5.1 Boundary conditions

There are two kinds of ODE problems, initial value problems (IVP) and boundary value problems (BVP). In an IVP, the solution of the ODE is known at an initial time t_{ini} and the system is evolved from that point on. In a BVP, parts of the solution may be known at each boundary point t_{ini} and t_{final}. This is the case for the thermal production of dark matter in chapter 7: we know the initial abundance at t_{ini}, since it is in thermal equilibrium, and we know the final value since it must match the observed abundance of dark matter, but the relativistic scattering cross section is not known. Any BVP can be solved by considering the n unknown parts of the solution at t_{ini} as a vector $\boldsymbol{x}$ of independent variables in the n-dimensional system of non-linear algebraic equations describing the difference between the known solution values at t_{final} and the result from solving an IVP with $\boldsymbol{x}$ in place of the unknown initial conditions. This is called a shooting method, and it works well for $n = 1$.

5.2 Implicit and explicit formulation

The most general formulation of an ODE is

$$\boldsymbol{F}\left(t, \boldsymbol{y}, \frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}\right) = 0. \tag{5.1}$$

Note that only a first order derivative is needed, since any higher order ODE can be transformed to a larger first order system by the identification of higher order derivatives as new components in $\boldsymbol{y}$. Equation (5.1) also describes differential algebraic equations (DAEs) if some of the equations in $\boldsymbol{F} = 0$ does not contain derivatives. If $\boldsymbol{F}\left(t, \boldsymbol{y}, \frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}\right)$ depends linearly on $\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t}$, equation (5.1) can be brought to the much simpler form

$$A(t)\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} - \boldsymbol{f}\left(t, \boldsymbol{y}\right) = 0, \tag{5.2}$$

and if $A(t)$ is invertible, it is possible to use either completely explicit methods or implicit methods. For the rest of this chapter, we will restrict ourselves even more, and only consider problems on the explicit form

$$\frac{\mathrm{d}\boldsymbol{y}}{\mathrm{d}t} = \boldsymbol{f}\left(t, \boldsymbol{y}\right), \tag{5.3}$$

with the initial value $\boldsymbol{y}\left(t_{\mathrm{ini}}\right) \equiv \boldsymbol{y}_{\mathrm{ini}}$, and "ODE" will refer to systems of the form (5.3).

5.3 First order methods

Forward Euler

The basic principle in solving an ODE is approximation of the derivative by a finite difference method. The simplest one approximates the derivative by a forward difference. Let $\boldsymbol{y}_n$ be the solution at time t_n, and define $h \equiv t_{n+1} - t_n$. The forward difference approximation to the derivative is then

$$\left.\frac{\mathrm{d}\boldsymbol{y}_n}{\mathrm{d}t}\right|_{t_n} \equiv f_n = \frac{\boldsymbol{y}_{n+1} - \boldsymbol{y}_n}{h} + \mathcal{O}\left(h\right), \tag{5.4}$$

which immediately leads to the forward Euler method:

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h f_n + \mathcal{O}\left(h^2\right). \tag{5.5}$$

This is an *explicit*, *first-order*, *one-step* method. Explicit refers to the fact that the Euler step (5.5) does not depend on $\boldsymbol{y}_{n+1}$ or equivalently f_{n+1}. The forward Euler method is first order, because the error after $N \sim (t_{\mathrm{final}} - t_{\mathrm{ini}})/h$ steps scales as $\mathcal{O}\left(h\right)$. It is a one-step method because it does not use the previously computed steps, $\boldsymbol{y}_{n-1}$, $\boldsymbol{y}_{n-2}$ etc. This method should never be used in practice, since it is very unstable as we will show later.

Backward Euler

The Euler method (5.5) builds upon the forward difference estimate (5.4). We could also have started with the backwards difference estimate

$$\left.\frac{\mathrm{d}\boldsymbol{y}_n}{\mathrm{d}t}\right|_{t_{n+1}} \equiv f_n = \frac{\boldsymbol{y}_{n+1} - \boldsymbol{y}_n}{h} + \mathcal{O}\left(h\right), \tag{5.6}$$

which leads to the backward Euler method

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h f_{n+1} + \mathcal{O}\left(h^2\right). \tag{5.7}$$

This expression resembles equation (5.5), but this is no longer an explicit method. Since f_{n+1} depends on $\boldsymbol{y}_{n+1}$, equation (5.7) is a system of algebraic equations that must be solved in each step. If f depends on $\boldsymbol{y}$ in a simple way, the solution may be known analytically. This is the case for the freeze-out calculation in chapter 7, where the system becomes a second-order algebraic equation. In the general case the algebraic system is nearly always solved by a modified version of Newton's method. Each time step in implicit methods are then much more expensive than in explicit methods, but as we shall see later, the pros of implicit methods often outweighs the cons.

5.4 Runge-Kutta methods

Runge-Kutta methods are one-step methods which may use a number of internal steps to compute the step. This number s is called the number of stages. Any Runge-Kutta method is defined by its Butcher tableau

$$\begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1s} \\ \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & \cdots & a_{ss} \\ \hline \boldsymbol{b}^T & b_1 & \cdots & b_s \end{array}, \tag{5.8}$$

and the rule for calculating $\boldsymbol{y}_{n+1}$:

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h \sum_{i=1}^{s} b_i k_i \tag{5.9a}$$

$$k_i = f\left(t_n + c_i h, \boldsymbol{y}_n + h \sum_{j=0}^{s} a_{ij} k_j \right). \tag{5.9b}$$

Inspection of equation (5.9) reveals that the explicit Runge-Kutta methods (ERK) is exactly those methods for which the matrix $A = [a_{ij}]$ is lower triangular. The forward and backward Euler methods, equation (5.5) and (5.7) are actually Runge-Kutta methods with the simple $[1 \times 1]$ A-matrices [0] and [1]. Their Butcher tableaus are

$$\begin{array}{c|c} 0 & 0 \\ \hline \boldsymbol{b}^T & 1 \end{array}, \qquad \begin{array}{c|c} 1 & 1 \\ \hline \boldsymbol{b}^T & 1 \end{array}. \tag{5.10}$$

There is an equivalent definition of the Runge-Kutta method (5.9) which is not so obvious. It is

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h \sum_{i=1}^{s} b_i f(t_n + c_i h, \boldsymbol{y}_n + \boldsymbol{z}_i) \tag{5.11a}$$

$$\boldsymbol{z}_i = h \sum_{j=1}^{s} a_{ij} f(t_n + c_i h, \boldsymbol{y}_n + \boldsymbol{z}_i). \tag{5.11b}$$

To see the equivalence, we define $\boldsymbol{F}$ by

$$\boldsymbol{F}(\boldsymbol{c}, \boldsymbol{x}) \equiv \begin{pmatrix} f(t_n + c_1 h, \boldsymbol{y}_n + x_1) \\ \vdots \\ f(t_n + c_s h, \boldsymbol{y}_n + x_s) \end{pmatrix}, \tag{5.12}$$

where we are considering $\boldsymbol{x}$ as a vector of the vectors ${x_i}^1$, and define $\boldsymbol{k}$ and $\boldsymbol{z}$ accordingly. We can now write the definition for $\boldsymbol{z}$ in equation (5.11) as

$$\boldsymbol{z} = hA\boldsymbol{F}\left(\boldsymbol{c}, \boldsymbol{z}\right), \tag{5.13}$$

[1] This can be made rigorous by defining the product space $\mathbb{R}^s \otimes \mathbb{R}^n$.

while the definition for $\boldsymbol{k}$ in equation (5.11) is equivalent to

$$\begin{aligned}\boldsymbol{k} &= \boldsymbol{F}\left(\boldsymbol{c}, hA\boldsymbol{k}\right)\\ hA\boldsymbol{k} &= hA\boldsymbol{F}\left(\boldsymbol{c}, hA\boldsymbol{k}\right),\end{aligned}$$

so it is clear that the two definitions are equivalent, and we have $\boldsymbol{z} \equiv hA\boldsymbol{k}$. It is easy to see that the two definitions also agree on $\boldsymbol{y}_{n+1}$ as they should. The number of possible Runge-Kutta methods are of course enormous, but not all of them are good. We will now discuss one of the best explicit methods.

5.5 Explicit Runge-Kutta

We will only consider embedded methods, which are methods that have a second set of coefficients $\hat{b_i}$, which gives an additional estimate of $\boldsymbol{y}_{n+1}$, which we call $\hat{\boldsymbol{y}}_{n+1}$. This estimate is usually one order higher or lower than the value used to evolve the system, and the difference between the two can be used to monitor the local truncation error (LTE). If the higher order method is used for stepping, the method is said to use local extrapolation. Numerical studies have shown such methods to be more efficient, and the 6-stage methods of order 5 are in general superior. Among the 6-stage 5(4) methods, the Dormand-Prince coefficients are arguably the best:

$$\begin{array}{c|ccccccc}
0 & & & & & & & \\
\frac{1}{5} & \frac{1}{5} & & & & & & \\
\frac{3}{10} & \frac{3}{40} & \frac{9}{40} & & & & & \\
\frac{4}{5} & \frac{44}{45} & -\frac{56}{15} & \frac{32}{9} & & & & \\
\frac{8}{9} & \frac{19372}{6561} & -\frac{25360}{2187} & \frac{64448}{6561} & -\frac{212}{729} & & & \\
1 & \frac{9017}{3168} & -\frac{355}{33} & \frac{46732}{5247} & \frac{49}{176} & -\frac{5103}{18656} & & \\
1 & \frac{35}{384} & 0 & \frac{500}{1113} & \frac{125}{192} & -\frac{2187}{6784} & \frac{11}{84} & \\
\hline
\boldsymbol{b}^T & \frac{35}{384} & 0 & \frac{500}{1113} & \frac{125}{192} & -\frac{2187}{6784} & \frac{11}{84} & \\
\hat{\boldsymbol{b}}^T & \frac{5179}{57600} & 0 & \frac{7571}{16695} & \frac{393}{640} & -\frac{92097}{339200} & \frac{187}{2100} & \frac{1}{40}
\end{array} \tag{5.14}$$

The method (5.14) is actually a 7-stage process, but note that the the rows a_{7j} and $\boldsymbol{b}^T$ are identical, so in the last step we are evaluating $f(t_{n+1}, \boldsymbol{y}_{n+1})$, which can be reused in the next step. This works because k_7 is not used to evolve the solution but only to evaluate the lower order estimate. The Dormand-Prince method has one additional feature, known as dense output. From the k_i's, one can construct a formula accurate to 4-th order which interpolates the function values between t_n and t_{n+1}. This is a very important feature in situations where the solution must be known at specific points, since the code can keep the step size at the largest possible size. Forcing a reduced step size to hit an output point exactly compromises the step size algorithm, since it works best when the step size makes the relative error close to the tolerance.

5.6 Stability analysis

The stability function

Some ODE systems are hard to solve, and that is often due to a phenomenon known as stiffness. It happens when we have vastly separated time scales in the problem, which happens often in cosmology since we usually have a particle physics time scale which is much shorter than the Hubble time scale on which we are interested in the solution. We can understand this issue from the simple linear differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \lambda t, \qquad y(t) = y(0)e^{\lambda t}. \tag{5.15}$$

One time scale is set by λ^{-1}, and the other is set by us. If λ is negative, the solution has decayed after a few times λ^{-1}, and the solution should be trivial. However, this system becomes stiff if we are interested in the evolution of y (or in general another component of the solution vector which depend on y) on times much longer than λ^{-1}. One Euler step of length h applied to the test equation (5.15) yields

$$\boldsymbol{y}_{n+1} = (1 + h\lambda)\boldsymbol{y}_n \equiv R^{\mathrm{FE}}(h\lambda)\boldsymbol{y}_n, \tag{5.16}$$

while one step using the backwards Euler method (5.7) can be written as

$$\boldsymbol{y}_{n+1} = \frac{1}{1 - h\lambda}\boldsymbol{y}_n \equiv R^{\mathrm{BE}}(h\lambda)\boldsymbol{y}_n, \tag{5.17}$$

where $R(z)$ is called the stability function of the method.

Stability domains

When λ has a negative real part, we should at least require the solution to be bounded in the limit $t \to \infty$. Applying the method iteratively with the same step size, we see that this is equivalent to the requirement $|R(z)| < 1$. This leads to the definition of the stability domain for the method:

$$S = \{z \in \mathbb{C}; |R(z)| \leq 1\}. \tag{5.18}$$

For real λ, this restricts the step size to $h < \lambda^{-1}$ in the forward Euler case. For $\lambda > 0$, this is actually not so bad, since we would already need a step size somewhat smaller than the time scale λ^{-1} to get a reasonable accuracy. However, when λ has a negative real part and after y has decayed, we must still respect this bound on h. The stability domains for the two explicit methods and the backwards Euler method are shown in figure 5.1. Looking at figure 5.1, we observe that the whole complex plane with negative real part is included in the stability domain of the backwards Euler method. This guarantees, that this method is not limited by stability for any decaying component. This is called A-stability. If we also require $|R(z)| \to 0$ for $z \to -\infty$, the method is called L-stable, and inspection of the stability function (5.17) reveals the backwards Euler method to be L-stable as well.

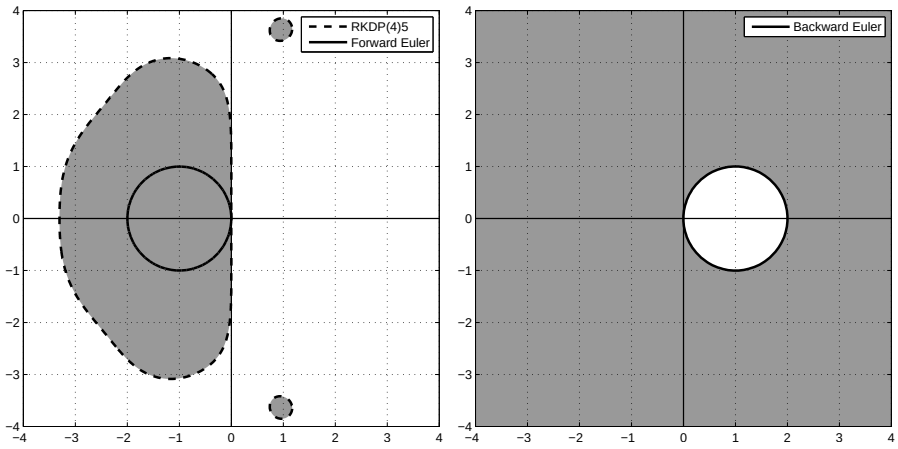


Figure 5.1: Stability domains in the complex plane for two explicit Runge-Kutta methods to the left and the backwards Euler method to the right.

5.7 RADAU5

Method and stability

We observed that the backwards Euler method is L-stable, but because it is an $\mathcal{O}\left(h^1\right)$ method, the step size must be small for the solution to be accurate. Just as we found a higher order explicit Runge-Kutta method to be superior to forward Euler, we may look for a higher order implicit Runge-Kutta method which still features L-stability. However, while adding stages to an explicit Runge-Kutta method is relatively cheap, adding stages to an implicit Runge-Kutta method is much more costly, since we must solve for all k_i, $i = 1, 2, \ldots, s$ simultaneously. This means that the non-linear algebraic system of equations is $[ns \times ns]$ where n is the number of equations and s is the number of stages, so we naturally prefer as few stages as possible. There is actually one method that meets all these requirements, and it is called RADAU5. It is a 5-th order, 3-stage, L-stable method with the Butcher tableau [93]

$$\begin{array}{c|ccc} \frac{4-\sqrt{6}}{10} & \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{4+\sqrt{6}}{10} & \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ 1 & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \\ \hline \boldsymbol{b} & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{array} \tag{5.19}$$

The stability domain for RADAU5 is given by the [2/3] Padé-approximation to the complex exponential function:

$$R^{\mathrm{R5}}(z) = [2/3]_{\exp}(z) = \frac{1 + \frac{1}{3}z}{1 - \frac{2}{3}z + \frac{1}{3}\frac{z^2}{2!}}. \tag{5.20}$$

The stability functions of the forward Euler method (5.16) and backward Euler method (5.17) are actually also Padé approximants. We have $R^{\mathrm{FE}}(z) = [1/0]_{\exp}(z)$ and $R^{\mathrm{BE}}(z) = [0/1]_{\exp}(z)$. The stability domain of RADAU5 is shown in figure 5.2.

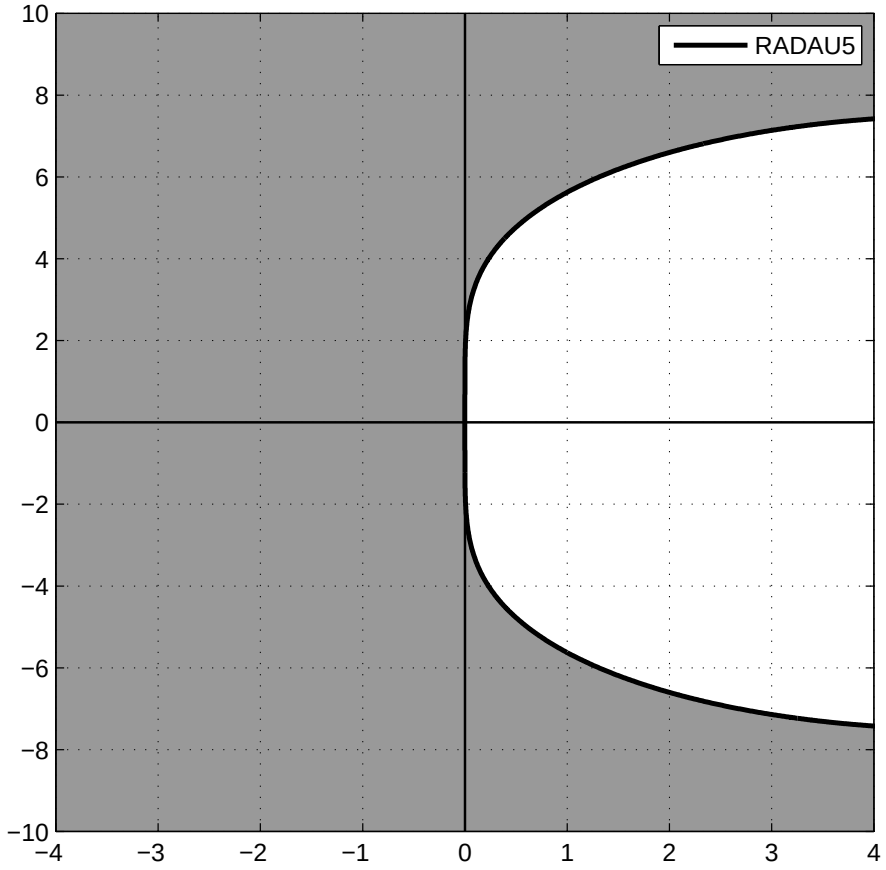


Figure 5.2: The stability domain for the L-stable `RADAU5` method.

Solving the non-linear equations

Since we have $a_{sj} = b_j$ for the `RADAU5` method, we find from definition (5.11)

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + z_s. \tag{5.21}$$

Let us define $\boldsymbol{G}$ by

$$\boldsymbol{G}(\boldsymbol{z}) \equiv \boldsymbol{z} - hA\boldsymbol{F}(\boldsymbol{z}), \tag{5.22}$$

where we have suppressed the dependence of $\boldsymbol{F}$ on the constant $\boldsymbol{c}$ as well as t_n and $\boldsymbol{y}_n$. We linearise $\boldsymbol{G}$ around $\boldsymbol{z}^{(k)}$, the k-th iterative of $\boldsymbol{y}_z$, according to Newtons algorithm. The Newton step is then

$$\boldsymbol{z}^{(k+1)} = \boldsymbol{z}^{(k)} + \Delta\boldsymbol{z}, \tag{5.23}$$

where $\Delta\boldsymbol{z}$ is given by the linear equation

$$\boldsymbol{J}_{\boldsymbol{G}}\left(\boldsymbol{z}^{(k)}\right)\Delta\boldsymbol{z} = -\boldsymbol{G}\left(\boldsymbol{z}^{(k)}\right) \Rightarrow \tag{5.24}$$

$$\left(\mathbb{1}_{[s\times s]} - hA\boldsymbol{J}_{\boldsymbol{F}}\right)\Delta\boldsymbol{z} = -\boldsymbol{z}^{(k)} + hA\boldsymbol{F}\left(\boldsymbol{z}^{(k)}\right). \tag{5.25}$$

Computing the Jacobian matrix of $\boldsymbol{F}$, $\boldsymbol{J}_{\boldsymbol{F}}$, requires $s \times n$ evaluations of f, and it must be computed n_{iter} times in each time step, where n_{iter} is the number of Newton iterations. However, we can make a good approximation for $\boldsymbol{J}_{\boldsymbol{F}}$ by

assuming $J_f(t_n + c_i, \boldsymbol{y}_n + z_i) \simeq J_f(t_n, \boldsymbol{y}_n)$. J_f is non-bold by intention, since it is not a matrix in the "stage-space" like $\boldsymbol{J_F}$ and A. In this approximation we have

$$\left(\mathbb{1}_{[s\times s]} - hA\mathbb{1}_{[s\times s]}J_f\right)\Delta\boldsymbol{z} = -\boldsymbol{z}^{(k)} + hA\boldsymbol{F}\left(\boldsymbol{z}^{(k)}\right), \tag{5.26}$$

$$\left(\mathbb{1}_{[s\times s]} - hAJ_f\right)\Delta\boldsymbol{z} = -\boldsymbol{z}^{(k)} + hA\boldsymbol{F}\left(\boldsymbol{z}^{(k)}\right). \tag{5.27}$$

Because the matrix $\left(\mathbb{1}_{[s\times s]} - hAJ_f\right)$ is now independent of the iterate $\boldsymbol{z}^{(k)}$, it can be LU-factorised once in each time-step and then used for solving the system for $\Delta\boldsymbol{z}$. This still means that we need to decompose a $[ns\times ns]$ matrix, but we can in fact do better than that as we shall see.

Block-diagonalisation of the linear system

The inverse Runge-Kutta matrix A^{-1} has a complex eigenvalue, and since the characteristic polynomial is a third order polynomial with real coefficients, we know that we have one real eigenvalue $\hat{\gamma}$ and two complex conjugate pairs $\hat{\alpha} \pm \mathrm{i}\hat{\beta}$. There exist a real matrix T with a real inverse T^{-1}, which brings A^{-1} to a simple form known as the real Jordan form. It is given by

$$T^{-1}A^{-1}T = \Lambda = \begin{bmatrix} \hat{\gamma} & & \\ & \hat{\alpha} & -\beta \\ & \beta & \alpha \end{bmatrix} \tag{5.28}$$

where we have

$$\hat{\gamma} = 3 - \sqrt[3]{3} + \sqrt[3]{3^2}, \quad \hat{\alpha} = 3 + \frac{1}{2}\left(\sqrt[3]{3} - \sqrt[3]{3^2}\right), \quad \hat{\beta} = \frac{\sqrt[6]{3}}{2}\left(3 + \sqrt[3]{3^2}\right), \tag{5.29}$$

and the inverse matrix A^{-1} is given by

$$A^{-1} = \begin{bmatrix} 2 + \sqrt{\frac{3}{2}} & -\frac{6}{5} + \frac{29}{5\sqrt{6}} & \frac{1}{15}\left(6 - 4\sqrt{6}\right) \\ -\frac{6}{5} - \frac{29}{5\sqrt{6}} & 2 - \sqrt{\frac{3}{2}} & \frac{1}{15}\left(6 + 4\sqrt{6}\right) \\ -1 + 8\sqrt{23} & -1 - 8\sqrt{23} & 5 \end{bmatrix}. \tag{5.30}$$

There is no simple analytic form of the transformation matrices except for the third row of T. I will write them out with 17 significant digits for completeness:

$$T = \begin{bmatrix} 0.094438762488975242 & 0.030029194105147425 & -0.14125529502095421 \\ 0.25021312296533331 & -0.38294211275726194 & 0.20412935229379993 \\ 1 & 0 & 1 \end{bmatrix}$$

$$T^{-1} = \begin{bmatrix} 4.1787185915519047 & 0.32768282076106239 & 0.52337644549944955 \\ 0.50287263494578688 & -2.5719269498556054 & 0.59603920482822493 \\ -4.1787185915519047 & -0.32768282076106239 & 0.47662355450055045 \end{bmatrix}.$$

We can now make a change of basis from $\boldsymbol{z}$ to $\boldsymbol{w}$ by $\boldsymbol{z} = T\boldsymbol{w}$. Applying this change of basis to the Newton iteration (5.27) yields

$$\begin{aligned}
\left(\mathbb{1}_{[s\times s]} - hAJ_f\right) T\Delta\boldsymbol{w} &= -T\boldsymbol{w}^{(k)} + hA\boldsymbol{F}\left(T\boldsymbol{w}^{(k)}\right) \Rightarrow \\
T^{-1}A^{-1}\left(\mathbb{1}_{[s\times s]} - hAJ_f\right) T\Delta\boldsymbol{w} &= -T^{-1}A^{-1}T\boldsymbol{w}^{(k)} + hT^{-1}A^{-1}A\boldsymbol{F}\left(T\boldsymbol{w}^{(k)}\right) \Rightarrow \\
\left(\Lambda - h\mathbb{1}_{[s\times s]}J\right)\boldsymbol{w} &= -\Lambda\boldsymbol{w}^{(k)} + hT^{-1}\boldsymbol{F}\left(T\boldsymbol{w}^{(k)}\right) \Rightarrow \\
\left(h^{-1}\Lambda - \mathbb{1}_{[s\times s]}J\right)\boldsymbol{w} &= -h^{-1}\Lambda\boldsymbol{w}^{(k)} + T^{-1}\boldsymbol{F}\left(T\boldsymbol{w}^{(k)}\right), \qquad (5.31)
\end{aligned}$$

and the next iteration is $\boldsymbol{w}^{(k+1)} = \boldsymbol{w}^{(k)} + \Delta\boldsymbol{w}$. Defining $\gamma = h^{-1}\hat{\gamma}$, $\alpha = h^{-1}\hat{\alpha}$ and $\beta = h^{-1}\hat{\beta}$, we can finally write our linear system as

$$\left(h^{-1}\Lambda - \mathbb{1}_{[s\times s]}J\right) = \begin{bmatrix} \gamma - J & & \\ & \alpha - J & -\beta \\ & \beta & \alpha - J \end{bmatrix} \begin{bmatrix} \Delta w_1 \\ \Delta w_2 \\ \Delta w_3 \end{bmatrix} = \begin{bmatrix} \mathrm{rhs}_1 \\ \mathrm{rhs}_2 \\ \mathrm{rhs}_3 \end{bmatrix}. \qquad (5.32)$$

We have managed to separate the system into two, one $[n \times n]$ system and one $[2n \times 2n]$ system. The $[2n \times 2n]$ system is equivalent to the $[n \times n]$ complex system

$$\left([\alpha + \mathrm{i}\beta] - J\right)\left(\Delta w_2 + \mathrm{i}\Delta w_3\right) = \left(\mathrm{rhs}_2 + \mathrm{irhs}_2\right), \qquad (5.33)$$

which is the final optimisation for the linear system.

5.8 Numerical differentiation formula

This section describes `ndf15`, the stiff ODE-solver for `CLASS`. It is a not a Runge-Kutta method but a variable order, adaptive step size, multi-step method.

Backwards differentiation formula

In section 5.3 we approximated the derivative by a forward difference and a backward difference and found the Euler methods. This is an extension of this idea. We define the backward difference operator ∇^k recursively by

$$\nabla^0 \boldsymbol{y}_n \equiv \boldsymbol{y}_n, \qquad (5.34)$$

$$\nabla^{j+1}\boldsymbol{y}_n \equiv \nabla^j \boldsymbol{y}_n - \nabla^j \boldsymbol{y}_{n-1}. \qquad (5.35)$$

The backwards differentiation formula (BDF) of order k is now

$$\sum_{j=1}^{k} \frac{1}{j}\nabla^j \boldsymbol{y}_{n+1} = hf_{n+1}. \qquad (5.36)$$

For $k = 1$ we find $\boldsymbol{y}_{n+1} - \boldsymbol{y}_n = hf_{n+1}$ which is just the backwards Euler method. The advantage of equation 5.36 is that it can be of high order (the method is order k) and it is relatively stable as we shall see later. Most importantly, the non-linear system that we need to solve in each step is only of size $[n \times n]$ - if we count 4 floating point operations per complex multiplication, this is roughly 5 times cheaper than our optimised `RADAU5` method!

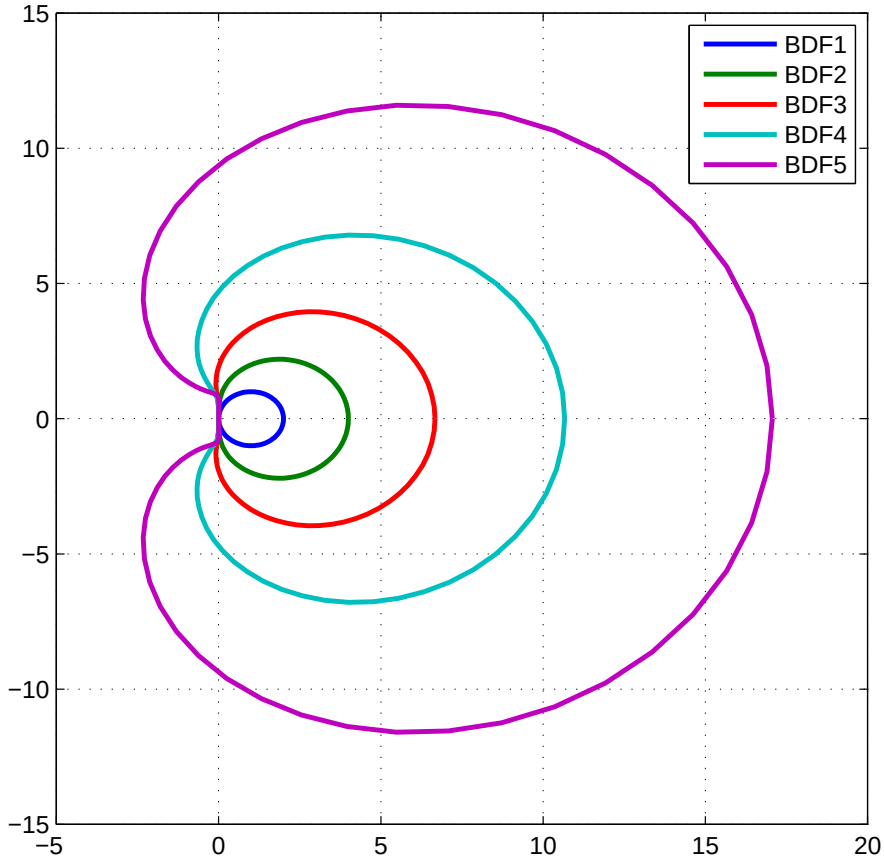


Figure 5.3: The stability domain for the BDF methods.

Stability

The stability of the BDF-method is quite good as indicated on figure 5.3. As we know, the $k = 1$ backwards Euler method is L-stable, and the same is true for $BDF2$. However, for $k > 2$, the BDF-methods loses A-stability.

Slightly improved formula

It is possible to add a free parameter κ to the method (5.36) without changing the order of the method:

$$\sum_{j=1}^{k} \frac{1}{j} \nabla^j \boldsymbol{y}_{n+1} + \kappa \left(\sum_{j=1}^{k} \frac{1}{j} \right) \nabla^{k+1} \boldsymbol{y}_{n+1} = h f_{n+1}. \tag{5.37}$$

The method (5.37) is called the numerical differentiation formula (NDF). The parameter κ can be used to either optimise the region of stability for $k \geq 3$, or to reduce the truncation error for all methods [122]. The latter approach is taken in `ndf15`.

Implementation

The implementation is an adaptive step size routine, but the step-size h is kept constant in the formula. This means that every time the step size is changed,

the values of the k last points must be interpolated. Integration starts with a NDF1 step, and since the code is only allowed to change order by 1, we are no longer restricted by a lack of memory in the sense of missing $\boldsymbol{y}_{n-j}$ points. The order selection is based on heuristics and works well. For the project in chapter 9, we would some times need to restrict the maximum order to 2 to get maximum stability.

Chapter 6

Sparse linear algebra

6.1 Data storage

An implicit ODE solver such as `RADAU5` or `ndf15` spends most of their time doing linear algebra. The Jacobian of $f(t, \boldsymbol{y})$ is often very sparse because many components do not couple directly. This is the case in linear Boltzmann codes, since a multi-pole only couples to its nearest neighbour and the perturbations that enter into the perturbed Einstein equations which is evident from equations (2.37). The same is true for the quantum kinetic equations of chapter 3 and 9, since each momentum bin evolve independently of the others except for the coupling to a few integrated quantities. For this reason we must take advantage of sparse matrix methods. We store a sparse $[n \times n]$ matrix `S` in compressed column form, such that the `nz` non-zero values of J is stored column after column in `S->Ax`. `S->Ax` has length `nz`, and the corresponding row-indices are stored in an integer array `S->Ai`. The integer array `S->Ap` of length `n+1` is constructed such that `S->Ap[0]=0` and that the number of entries in row `j` is given by `S->Ap[j+1]-S->Ap[j]`. Here is an example of how to loop through the matrix `S`:

```
for(col=0; col<n; col++){
    for(i=S->Ap[col]; i<S->Ap[col+1]; i++){
        row = S->Ai[i];
        S_dense[row][col] = S->Ax[i];
    }
}
```

6.2 LU-factorisation

Our sparse LU factorisation is a left-looking LU with partial pivoting which roughly reduces the problem to solving sparse, lower-triangular systems with sparse right hand sides [143]. For dense matrices and vectors this is solved trivially by back-substitution, but this is not so easy for a sparse method. We must first find out which elements in the solution vector $\boldsymbol{x}$ which will be non-zero for the method to be optimal. This step is necessary but also time consuming. However, if the Jacobian has the same structure, and its values has not changed by many orders of magnitude, we can reuse our partial-pivoting

scheme. Such a re-factorisation takes only a fraction of the time of a normal factorisation, and is a prime example on what can be gained by not relying on external black box libraries.

Part III

Scientific Research

Chapter 7

Sommerfeld Enhancement of DM Annihilation: Resonance Structure, Freeze-Out and CMB Spectral Bound

Steen Hannestad and Thomas Tram

Abstract In the last few years there has been some interest in WIMP Dark Matter models featuring a velocity dependent cross section through the Sommerfeld enhancement mechanism, which is a non-relativistic effect due to massive bosons in the dark sector. In the first part of this article, we find analytic expressions for the boost factor for three different model potentials, the Coulomb potential, the spherical well and the spherical cone well and compare with the numerical solution of the Yukawa potential. We find that the resonance pattern of all the potentials can be cast into the same universal form.

In the second part of the article we perform a detailed computation of the Dark Matter relic density for models having Sommerfeld enhancement by solving the Boltzmann equation numerically. We calculate the expected distortions of the CMB blackbody spectrum from WIMP annihilations and compare these to the bounds set by FIRAS. We conclude that only a small part of the parameter space can be ruled out by the FIRAS observations.

7.1 Introduction

The most probable candidate for Dark Matter is arguably the WIMP, a Weakly Interacting Massive Particle. This is due to what is often called the 'WIMP miracle'; new physics is required at the TeV scale when unitarity breaks down in the Standard Model, and some symmetry should protect the proton from decaying too fast through new diagrams involving these new degrees of freedom. The symmetry may ensure that at least one of the new particles is

stable on cosmological timescales, while the annihilation cross section is typically within an order of magnitude of what is required to explain the observed relic abundance.

In the last few years there have been some observations [13–16, 21–23, 25, 26] pointing to anomalous emission within the galaxy. It is possible, that these observations may be explained by WIMPs annihilating in the galactic halo [17–20, 24]. Common to all these suggestions is that the necessary annihilation rate requires an annihilation cross section a few orders of magnitude larger than the one consistent with a thermal relic density, $\langle \sigma v \rangle_{\mathrm{TH}} \simeq 3 \cdot 10^{-26} \mathrm{cm}^3/\mathrm{s}$.

This apparent discrepancy, as well as the necessary suppression of hadronic final states, has been suggested to stem from new GeV-mass force carriers in the dark sector [5, 31] which leads to Sommerfeld enhancement of the Dark Matter pair annihilation cross section. Although the anomalies mentioned above may turn out to be explained by normal astrophysical processes, there may still be force carriers in the dark sector which are light compared to the mass of the Dark Matter particle, so the mechanism of Sommerfeld enhancement is interesting and generic. Sommerfeld enhancement applied to WIMP annihilations in general, not related to the cosmic ray excess, was first studied in [32, 33] and later in [45] and [34].

7.2 Sommerfeld enhancement

Sommerfeld enhancement is a consequence of having light force carriers mediate an attractive interaction in the dark sector, which creates a Yukawa-potential for the WIMPs. In this work we consider only s-wave annihilation. We write this potential as

$$V(r) = -\frac{\lambda^2}{4\pi r} e^{-m_\phi r} = -\frac{\alpha}{r} e^{-m_\phi r}, \tag{7.1}$$

where λ is the coupling parameter, m_ϕ is the mass of the force carrier and r is the relative distance between the WIMPs. Since the potential is only a function of the relative distance, it only enters in the reduced one-particle Schrödinger equation for the relative motion. In spherical coordinates this is

$$-\frac{1}{2\mu}\nabla^2 \psi_k = \left(\frac{k^2}{2\mu} - V(r) \right) \psi_k, \tag{7.2}$$

where $\mu = m_\chi/2$ is the reduced mass. We are interested in the probability density of ψ_k when $r = 0$, since the boost factor for the s-wave case is given by $S_k = |\psi_k(0)|^2$, when ψ_k is normalized to the asymptotic form

$$\psi \to e^{ikz} + f(\theta)\frac{e^{ikr}}{r} \text{ as } r \to \infty. \tag{7.3}$$

This form of the boost factor can be derived [5] from a simple argument: Annihilations are assumed to proceed by a delta function interaction, so the rate of this process must be proportional to the norm squared of the reduced wave function at zero separation, $\sigma \propto |\psi(r=0)|^2$. Denoting the $V = 0$ wave function by ψ^0, we must have

$$S = \frac{\sigma}{\sigma^0} = \frac{|\psi(0)|^2}{|\psi^0(0)|^2}. \tag{7.4}$$

When the potential is rotationally symmetric, we know that the solution of equation (7.2) becomes invariant under rotations around the axis of the incoming particle, and hence we can expand ψ in products of Legendre polynomials and a radial wave function R_{kl}. Only R_{k0} is non-zero at the origin, so we define $\chi_k \equiv rR_{k0}$, which leads to the equation

$$\frac{1}{m_\chi}\frac{\mathrm{d}^2\chi}{\mathrm{d}r^2} = \left(-\frac{\alpha}{r}e^{-m_\phi r} - m_\chi\beta^2\right)\chi \Rightarrow \tag{7.5}$$

$$\frac{m_\phi}{m_\chi}\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} = \left(-\frac{\alpha}{x}e^{-x} - \frac{m_\chi}{m_\phi}\beta^2\right)\chi \Rightarrow \tag{7.6}$$

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} = -\left(\frac{\alpha}{fx}e^{-x} + \left(\frac{\beta}{f}\right)^2\right)\chi, \tag{7.7}$$

where we introduced the dimensionless distance $x \equiv m_\phi r$ and the ratio of the two masses $f \equiv m_\phi/m_\chi$. Equation (7.7), together with the boundary conditions

$$\chi(0) = 0 \tag{7.8a}$$

$$\chi \to \sin\left(\frac{\beta}{f}x + \delta\right), \text{ as } x \to \infty. \tag{7.8b}$$

defines the problem.

Numerical solution

In principle we should solve the boundary value problem (7.8) numerically, but since the Schrödinger equation is linear, we can exchange the condition (7.8b) by setting the derivative of χ to unity at $x = 0$ and solve what is now an initial value problem for $\widetilde{\chi}$. The solution to the original problem is just given by $\chi = \widetilde{\chi}/A$, where A is the asymptotic amplitude of $\widetilde{\chi}$. The enhancement factor is then given by

$$S_k = \left|\frac{R_{k0}^2(0)}{k}\right|^2 = \left|\frac{1}{k}\frac{\mathrm{d}\chi_k}{\mathrm{d}r}(0)\right|^2 = \left|\frac{f}{A\beta}\right|^2. \tag{7.9}$$

In figure 7.1 we have plotted the Sommerfeld boost factor for a velocity of 150km/s for the whole (α, f)-parameter space. The vertical lines in figure 7.1 are ragged by thin, straight lines of resonance which have, as we shall see later, a rather large impact on the freeze out process.

To calculate the amplitude A in (7.9), we must know when our waveform $\widetilde{\chi}$ has reached its asymptotic form. This must happen when the Yukawa potential becomes much less than the kinetic energy of the particle, so we find the position of equality:

$$\begin{aligned}\frac{\alpha}{x_{\text{range}}}e^{-x_{\text{range}}} &= \frac{\beta^2}{f} \Rightarrow \\ x_{\text{range}}e^{x_{\text{range}}} &= \frac{f\alpha}{\beta^2} \Rightarrow \\ x_{\text{range}} &= W\left(\frac{f\alpha}{\beta^2}\right),\end{aligned} \tag{7.10}$$

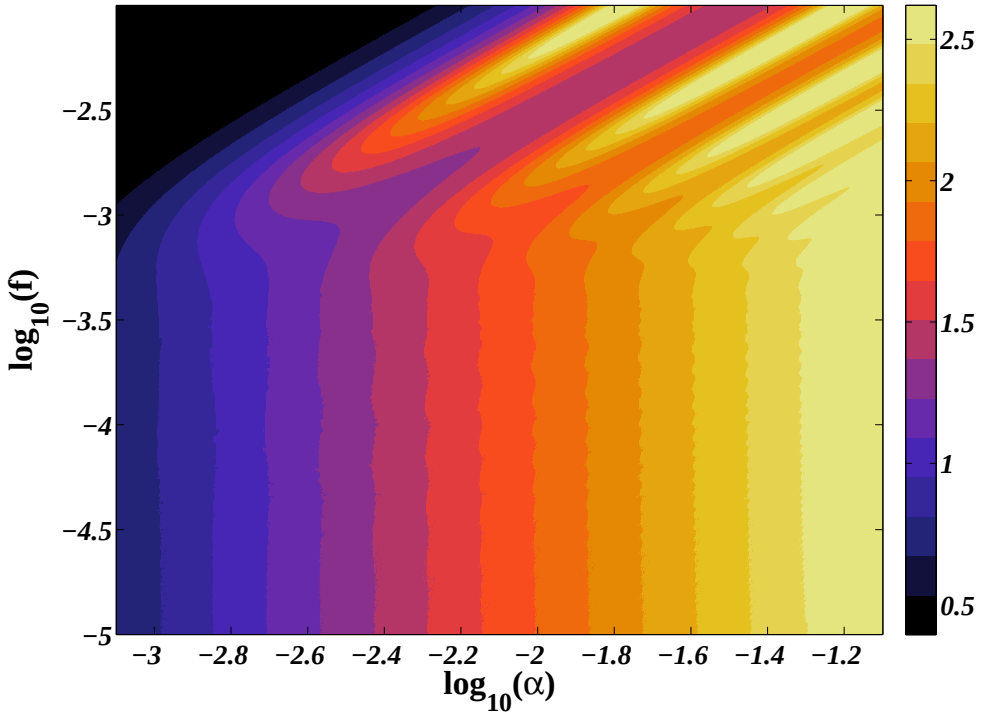


Figure 7.1: Sommerfeld boost factor in $\log_{10}$ for a relative velocity of 150km/s.

where W is the Lambert W-function. To calculate the amplitude we use the asymptotic waveform (7.8b):

$$\widetilde{\chi} = A \sin\left(\frac{\beta}{f}x + \delta\right) \tag{7.11a}$$

$$\frac{\mathrm{d}\widetilde{\chi}}{\mathrm{d}x} = \frac{\beta}{f} A \cos\left(\frac{\beta}{f}x + \delta\right), \tag{7.11b}$$

which leads to this expression for the amplitude A:

$$A = \sqrt{\widetilde{\chi}^2 + \left(\frac{f}{\beta}\frac{\mathrm{d}\widetilde{\chi}}{\mathrm{d}x}\right)^2}. \tag{7.12}$$

We evolve our waveform $\widetilde{\chi}$ to $1.5x_{\text{range}}$, and after this point we calculate A at each succeeding point until A has converged.

Coulomb potential

In the limit of a massless force carrier, the Yukawa potential becomes a Coulomb potential. In this case, the enhancement factor can be calculated in closed form,

$$S_{\epsilon_C} = \frac{\pi/\epsilon_C}{1 - e^{-\pi/\epsilon_C}}, \qquad \epsilon_C = \frac{\beta}{\alpha} \tag{7.13}$$

by solving the Schrödinger equation in terms of hyper geometric functions. The derivation of equation (7.13) can be found in section 7.5. We want to stress, that the applicability of this limit does not only depend on $f \ll 1$,

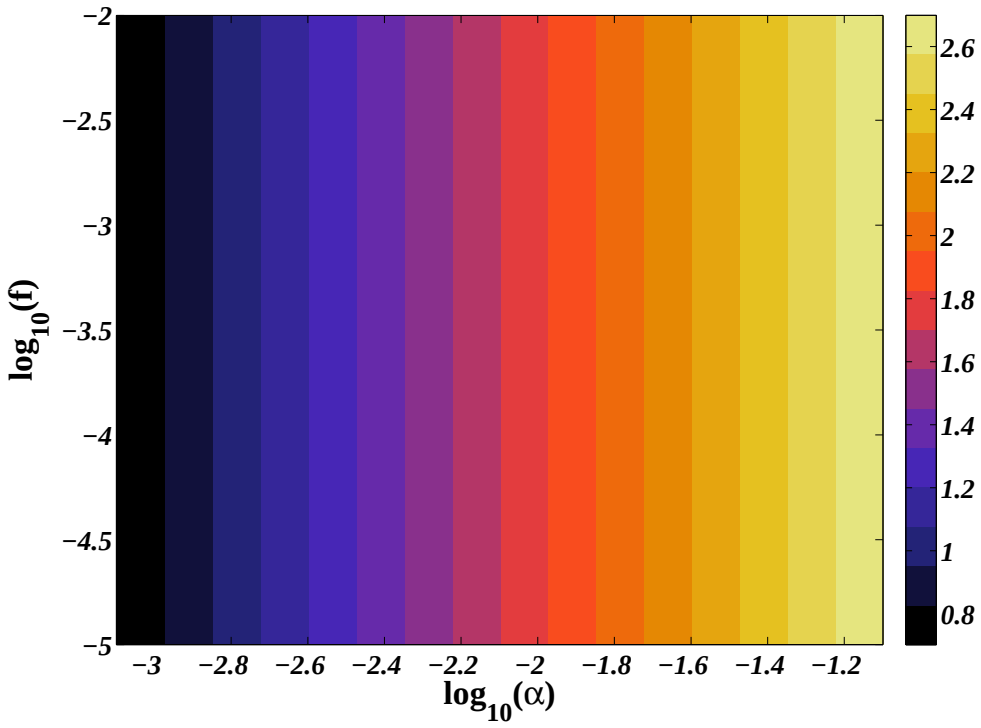


Figure 7.2: Boost factor in $\log_{10}$ for the Coulomb potential in $\log_{10}$ for a relative velocity of 150km/s.

but is also dependent on α and β. The correct question to ask is whether or not the scattering will take place in a regime in which the potential looks like a Coulomb potential. We want the range of the potential x_{range} from (7.10), to be less than something like $\ln(2)$. This would keep the exponential in the Yukawa potential at order 1 during the interaction, so we should be safe in assuming a Coulomb potential instead. (However, this does not take into account the non-perturbative effects of resonances.) The boost factor (7.13) is plotted in figure 7.2. If we compare with the Sommerfeld-case in figure 7.1, we can see that the lower left part of the Sommerfeld parameter space is well described by the boost from the Coulomb potential. This part of the parameter space is also the part where the range (7.10) of the potential is less than about 0.7 as expected. However, we do not find resonances in the Coulomb case as we do for the Sommerfeld enhancement, because the potential is not localized. To learn more about these resonances, we relate the Yukawa potential to two model potentials which can be solved analytically.

Spherical Well

The resonance structure of the Sommerfeld-enhancement does not appear in the massless limit, but we can understand it qualitatively by examining the spherical well and relating the well to the Yukawa potential[1]. We look at a

[1] In [33], the authors also related the Yukawa potential to the spherical well to understand the effect of resonances.

potential of the form

$$V(r) = \begin{cases} -V_0, & r \leq L/m_\phi \\ 0, & r > L/m_\phi. \end{cases} \tag{7.14}$$

We set V_0 to a value such that the potential integrated over volume agrees with the Yukawa potential (7.1). We find

$$V_0 = \frac{3\alpha m_\phi}{L^3}, \tag{7.15}$$

where L is the range of the potential in units of m_ϕ^{-1}, and should be order 1. The Schrödinger equation for this potential is

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} = -\left(\frac{V_0}{m_\phi f} + \left(\frac{\beta}{f}\right)^2\right)\chi \Rightarrow \tag{7.16}$$

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} = -\left(K^2 + \epsilon^2\right), \tag{7.17}$$

where we are using $x = m_\phi r$ and

$$K = \sqrt{\frac{3\alpha}{fL^3}} \quad \epsilon = \frac{\beta}{f}. \tag{7.18}$$

This equation is solved by sines and cosines, so if we define $p = \sqrt{K^2 + \epsilon^2}$, we can write the general solution as

$$\chi^{x<L}(x) = A\sin\left(px + \gamma\right) \tag{7.19a}$$

$$\chi^{x>L}(x) = B\sin\left(\epsilon(x - L) + \delta\right), \tag{7.19b}$$

and using boundary conditions (7.8), we can set $\gamma = 0$ and $B = 1$. We must now match the wave function and its derivative at the boundary at $x = L$. We get the equations

$$A\sin\left(pL\right) = \sin\left(\delta\right) \tag{7.20a}$$

$$Ap\cos\left(pL\right) = \epsilon\cos\left(\delta\right) \tag{7.20b}$$

and we solve for the amplitude A. We find:

$$A^2 = \frac{1}{1 + \left(\frac{K}{\epsilon}\right)^2\cos^2\left(pL\right)}, \tag{7.21}$$

which results in a boost factor

$$S = \frac{1 + \left(\frac{K}{\epsilon}\right)^2}{1 + \left(\frac{K}{\epsilon}\right)^2\cos^2\left(pL\right)}. \tag{7.22}$$

We have plotted this boost factor for $L = 1$ in figure 7.3 for the same (α, f)-parameter space as before. If we compare with figure 7.1, we can see the same type of resonances, although they become suppressed in the Yukawa case when the scattering becomes Coulomb-like. The structure of (7.22) tells us

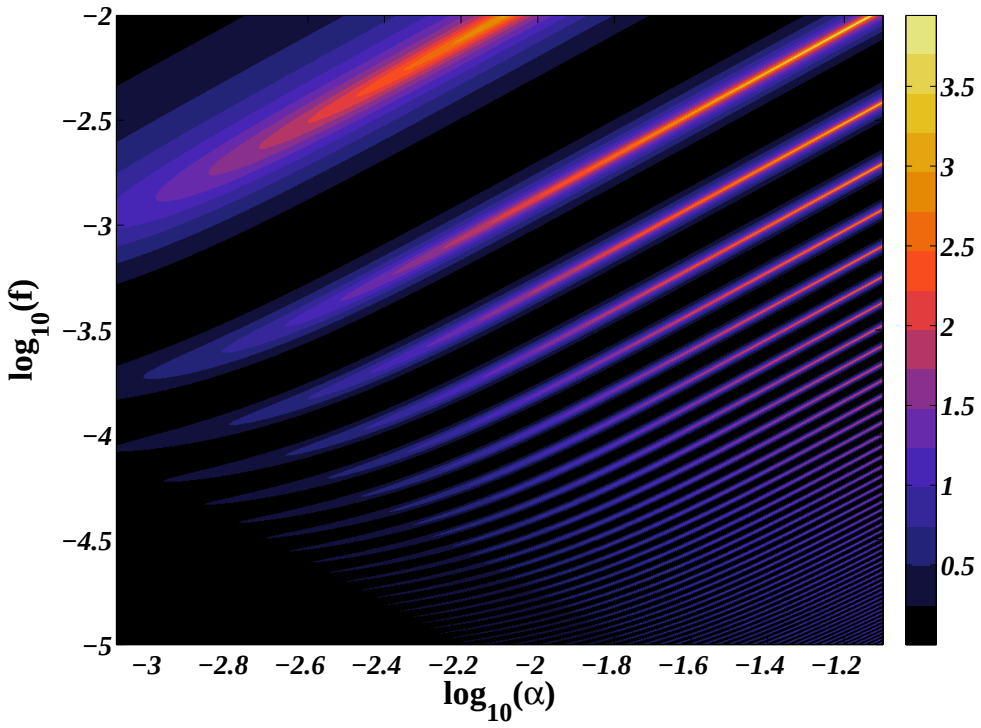


Figure 7.3: Boost factor in $\log_{10}$ for the spherical well with $v = 150$km/s. The depth of the well is related to the strength of the corresponding Yukawa potential.

something about the behavior of resonances, even for the Yukawa-case. First off, since the denominator is always larger than unity, the maximal boost must be of the order $\sim K^2/\epsilon^2$. Secondly, if the cosine is of order unity, the boost will be order unity as well. So the resonances occur when $pL/\pi \simeq n + 1/2$.

If we are interested in resonances giving S of order 100 or more, we can make the approximation $\epsilon/K \ll 1$, which allows us to use the zeroth order approximation $p \approx K$. Using the definition of K, we find

$$\frac{KL}{\pi} = \sqrt{\frac{3\alpha}{fL^3}}\frac{L}{\pi} = \sqrt{\frac{\alpha}{f}}\sqrt{\frac{3}{L\pi^2}}, \tag{7.23}$$

so the position of the resonances depends only on the ratio of α and f. This suggests that we introduce the rotated coordinates

$$u = \frac{\alpha}{f} \tag{7.24a}$$

$$v = \alpha f. \tag{7.24b}$$

Using (7.23), we find that the resonances happens at

$$\sqrt{u_n} = \sqrt{\frac{L\pi^2}{3}}\left(n + \frac{1}{2}\right) \Rightarrow$$

$$u_n = \frac{L\pi^2}{3}\left(n + \frac{1}{2}\right)^2, \quad n = 0, 1, \ldots \tag{7.25}$$

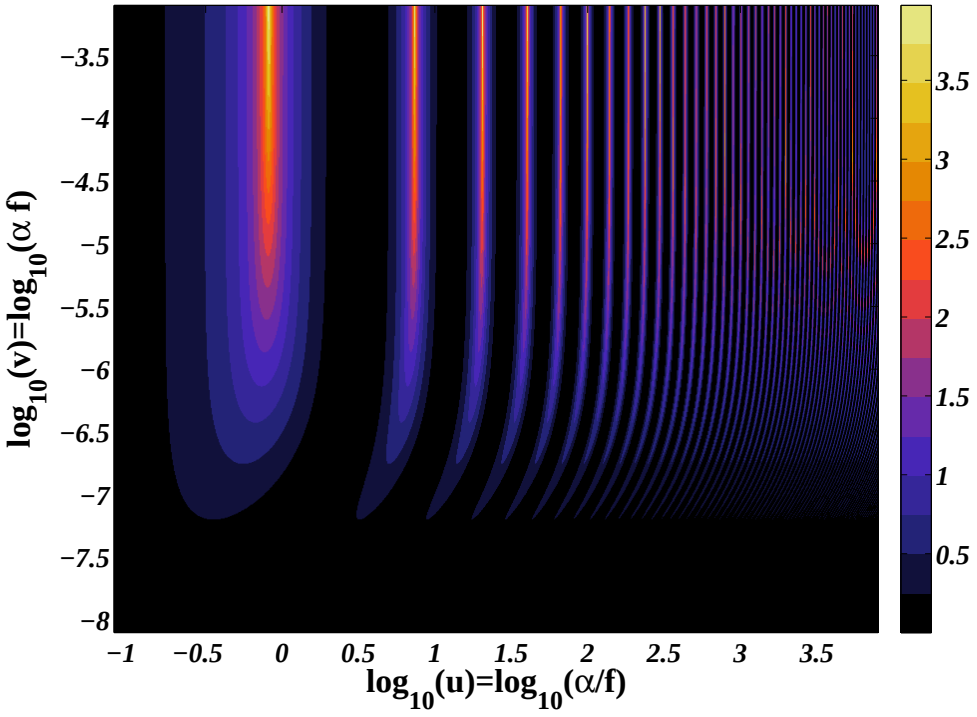


Figure 7.4: Boost factor for the spherical well with $v = 150$km/s in the new coordinates u and v.

Since the distance between resonances increases only by n^2, it is clear that they will happen closer and closer when viewed in a logarithmic plot, just as we saw in the Yukawa-case. The other parameter v controls the order of the boost factor, since we have

$$\frac{K^2}{\epsilon^2} = \frac{3}{L^3}\frac{f\alpha}{\beta^2} = \frac{3}{L^3}\frac{v}{\beta^2}. \tag{7.26}$$

We have shown the boost plot in the rotated coordinates (u, v) in figure 7.4. We suspect that this choice of coordinates would also work well in the Yukawa case, so we calculate the Sommerfeld boost factor in the (u, v) coordinates, and the result is shown in figure 7.5. As is clear from the plot, u alone determines the position of the peaks, just as for the spherical well. This is in agreement with [45], as our coordinate u is identical to their parameter ϵ_ϕ.

Spherical Slope Well

We now examine another model potential, the spherical slope well, to see how the position of resonances change compared to the spherical well. We write the potential

$$V(r) = \begin{cases} -V_0\left(1 - \frac{m_\phi}{L}r\right), & r \leq L/m_\phi \\ 0, & r > L/m_\phi, \end{cases} \tag{7.27}$$

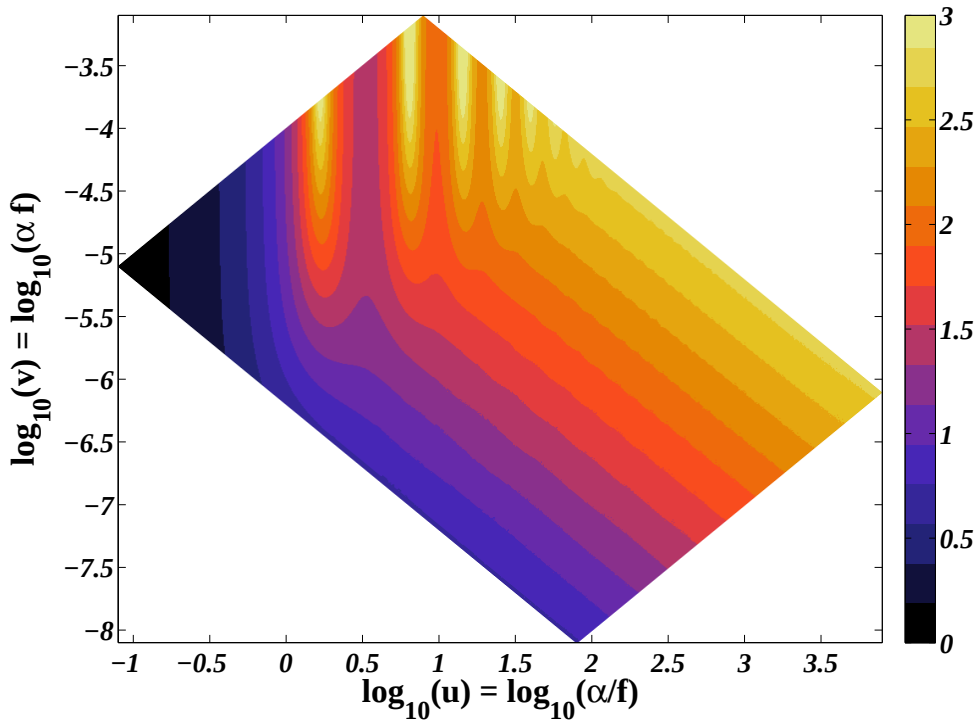


Figure 7.5: The Sommerfeld boost factor with $v = 150$km/s in the new coordinates u and v. We have cropped the plot at the (α, f) limits from figure 7.1

and as before, we set V_0 to a value such that the potential integrated over volume agrees with the Yukawa potential (7.1). In this case we find

$$V_0 = \frac{12\alpha m_\phi}{L^3}, \tag{7.28}$$

where L is the range of the potential in units of m_ϕ^{-1}, and should be order 1. The Schrödinger equation for this potential is

$$\begin{aligned}
\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} &= -\left(\frac{V_0}{m_\phi f}\frac{1}{L}(L-x) + \left(\frac{\beta}{f}\right)^2\right)\chi \Rightarrow \\
\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} &= -\left(K^2(L-x) + \epsilon^2\right) \Rightarrow \\
\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} &= -\left(-K^2 x + p^2\right),
\end{aligned}$$

where we are using $x = m_\phi r$, $K = \sqrt{\frac{12\alpha}{fL^4}}$ and $p = \sqrt{K^2 L + \epsilon^2}$. If we now do the substitution $\xi = K^{-4/3}\left(K^2 x - p^2\right)$, we get the Airy equation

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}\xi^2} = \xi\chi, \tag{7.29}$$

which has the general solution

$$\chi = c_1 \mathrm{Ai}(\xi) + c_2 \mathrm{Bi}(\xi). \tag{7.30}$$

We are interested in χ at two different values, $\xi_0 \equiv \xi(x=0)$ and $\xi_L \equiv \xi(x=L)$. These are given by

$$\xi_0 = K^{-4/3}\left(-p^2\right) = -K^{-4/3}p^2 \tag{7.31a}$$

$$\xi_L = K^{-4/3}\left(K^2 L - p^2\right) = -K^{-4/3}\epsilon^2. \tag{7.31b}$$

Since we have $\xi < 0$ in both cases, it makes sense to introduce the variable

$$\zeta = \frac{2}{3}\left(-\xi\right)^{3/2}, \tag{7.32}$$

and reexpress the solution χ and its derivative in terms of Bessel functions. We have:

$$\chi(\xi) = \sqrt{-\xi}\left[-AJ_{1/3}(\zeta) + BJ_{-1/3}(\zeta)\right] \tag{7.33a}$$

$$\frac{\mathrm{d}\chi}{\mathrm{d}\xi}(\xi) = -\xi\left[AJ_{-2/3}(\zeta) + BJ_{2/3}(\zeta)\right]. \tag{7.33b}$$

where the new coefficients A and B are related to the old ones by

$$A = -c_1/3 + c_2/\sqrt{3} \qquad B = c_1/3 + c_2/\sqrt{3}. \tag{7.34}$$

Since we require $\chi(\xi_0) = 0$, equation (7.33a) can be used to relate A and B, since we have

$$\begin{aligned} AJ_{1/3}(\zeta_0) &= BJ_{-1/3}(\zeta_0) \Rightarrow \\ B &= \frac{J_{1/3}(\zeta_0)}{J_{-1/3}(\zeta_0)}A \equiv tA. \end{aligned} \tag{7.35}$$

At this point we have one free parameter A left to determine from the solution $\chi^{x<L}$. The solution outside the well is just as before, (7.19b), and contains one free parameter, δ. Both parameters are fixed by matching the solution and its derivative at the boundary; we have two equations with two unknowns:

$$\sin(\delta) = \chi^{x<L}(x=L) \tag{7.36a}$$

$$\epsilon\cos(\delta) = \frac{\mathrm{d}\chi}{\mathrm{d}x}^{x<L}(x=L), \tag{7.36b}$$

and the interesting parameter A is most easily found by inserting the right hand sides in the well known trigonometric identity

$$\begin{aligned} \epsilon^2 &= \epsilon^2\sin^2(\delta) + (\epsilon\cos(\delta))^2 \\ &= \epsilon^2(-\xi_L)A^2\left[-J_{1/3}(\zeta_L) + tJ_{-1/3}(\zeta_L)\right]^2 + \\ &+ \xi_L^2 K^{4/3}A^2\left[J_{-2/3}(\zeta_L) + tJ_{2/3}(\zeta_L)\right]^2 \Rightarrow \\ A^2 &= \frac{K^{4/3}/\epsilon^2}{\left[-J_{1/3}(\zeta_L) + tJ_{-1/3}(\zeta_L)\right]^2 + \\ &\quad + \left[J_{-2/3}(\zeta_L) + tJ_{2/3}(\zeta_L)\right]^2}. \end{aligned} \tag{7.37}$$

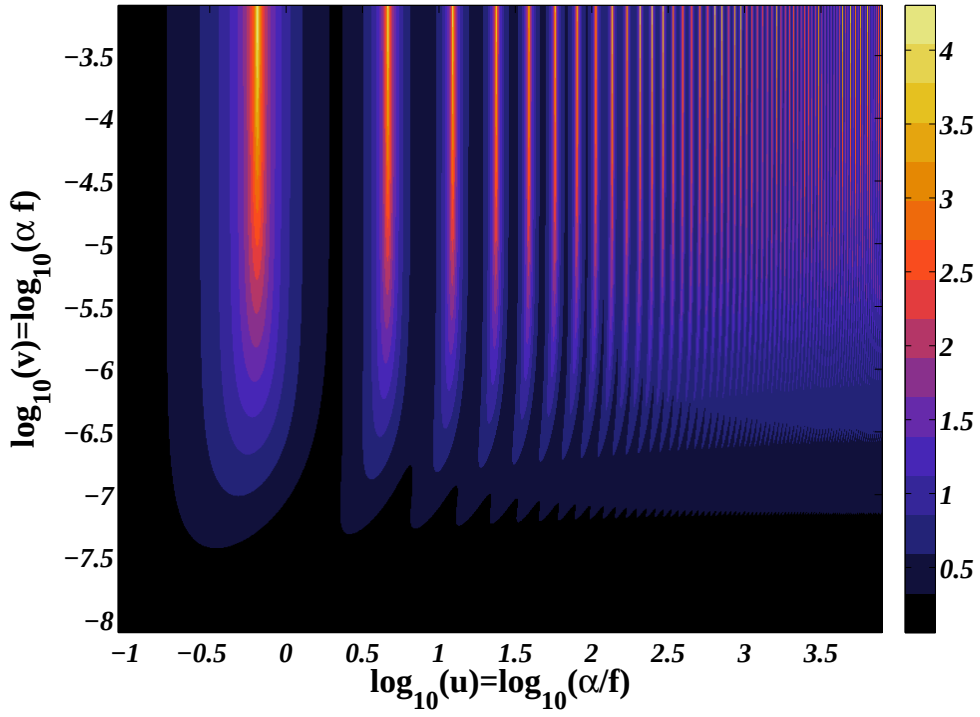


Figure 7.6: Boost factor in $\log_{10}$ for the spherical slope well with $v = 150\text{km/s}$. The depth of the well is related to the strength of the corresponding Yukawa potential.

Using equation (7.9), we can find the boost factor S:

$$\begin{aligned} S &= \frac{p^4}{\epsilon^2} K^{-4/3} A^2 \left[J_{-2/3}(\zeta_0) + t J_{2/3}(\zeta_0) \right]^2 \Rightarrow \\ S &= \frac{p^4}{\epsilon^4} \frac{\left[J_{-2/3}(\zeta_0) + t J_{2/3}(\zeta_0) \right]^2}{\left[-J_{1/3}(\zeta_L) + t J_{-1/3}(\zeta_L) \right]^2 + \\ &\quad + \left[J_{-2/3}(\zeta_L) + t J_{2/3}(\zeta_L) \right]^2}. \end{aligned} \tag{7.38}$$

This is a rather complicated expression, but the numerator can be simplified by using the series expansion for $J(z)$ and the following identity for the gamma function:

$$\begin{aligned} \Gamma(1/3)\Gamma(2/3) &= \Gamma(1/3)\Gamma(1 - 1/3) \\ &= \pi \csc(\pi/3) = 2\pi/\sqrt{3} \end{aligned} \tag{7.39a}$$

$$J_\nu(z) = (z/2)^\nu \sum_{k=0}^{\infty} \frac{(-)^k (z/2)^{2k}}{k! \Gamma(\nu + k + 1)} \tag{7.39b}$$

Having these, we can prove the neat identity

$$J_{-2/3}(z) J_{-1/3}(z) + J_{1/3}(z) J_{2/3}(z) = \frac{\sqrt{3}}{\pi z}. \tag{7.40}$$

Reinserting t, the numerator of equation (7.38) can now be written as

$$\begin{aligned}\text{num.} &= \Big[J_{-2/3}(\zeta_0) + tJ_{2/3}(\zeta_0)\Big]^2 \\ &= J_{-1/3}(z)^{-2}\times \\ &\times \Big[J_{-1/3}(\zeta_0)J_{-2/3}(\zeta_0) + J_{1/3}(\zeta_0)J_{2/3}(\zeta_0)\Big]^2 \\ &= \frac{3}{\pi^2\zeta_0^2 J_{-1/3}(\zeta_0)^2}. \end{aligned} \tag{7.41}$$

Inserting (7.41) into (7.38) and rearranging the denominator, we find:

$$\begin{aligned}S =& \frac{3^3K^4}{2^2\pi^2\epsilon^4p^2}\Big(J_{1/3}(\zeta_0)^2\left[J_{-1/3}(\zeta_L)^2 + J_{2/3}(\zeta_L)^2\right] + \\ &+ J_{-1/3}(\zeta_0)^2\left[J_{1/3}(\zeta_L)^2 + J_{-2/3}(\zeta_L)^2\right] + \\ &+ J_{-1/3}(\zeta_0)J_{1/3}(\zeta_0)J_{-1/3}(\zeta_L)\times \\ &\times \left[J_{-2/3}(\zeta_L) - J_{1/3}(\zeta_L)\right]\Big)^{-1}. \end{aligned} \tag{7.42}$$

We have plotted this boost factor in figure 7.6 in the (u, v)-coordinates we introduced for the spherical well. We see the same behavior as before, but the expression for the boost factor is still too complicated for us to deduce where the resonances will be. Let us take another look at the denominator of equation (7.42). We want to estimate the numerical value of the two arguments, ζ_0 and ζ_L. We got:

$$\zeta_0 = \frac{2}{3}\left(-\xi_0\right)^{3/2} = \frac{2p^3}{3K^2} \tag{7.43a}$$

$$\zeta_L = \frac{2}{3}\left(-\xi_L\right)^{3/2} = \frac{2\epsilon^3}{3K^2}. \tag{7.43b}$$

If we are interested in regions of possibly large boost factors, we may assume $K \gg \epsilon$ as we did before, and we conclude that $\zeta_0 \gg 1$ and $\zeta_L \ll 1$. It turns out, that to a reasonable approximation, we can set the Bessel functions equal to their asymptotic values. If we remember that

$$\begin{aligned} J_\nu(z) &\simeq \frac{(z/2)^\nu}{\Gamma(\nu+1)}, && z \ll 1 \\ J_\nu(z) &\simeq \sqrt{\frac{2}{\pi z}}\cos\left(z - \nu\pi/2 - \pi/4\right), && z \gg 1 \end{aligned}$$

we can identify the two largest terms in the denominator of equation (7.42). We are looking for the largest negative powers of ζ_L, because they will give the largest contribution to the denominator. We have one term behaving as $\sim \zeta_L^{-4/3}$ and one term from the crossproduct is behaving as $\sim \zeta_L^{-1}$, and the rest of the 6 terms has more positive powers. These two worst terms multiply $J_{-1/3}(\zeta_0)$, so we suspect to have a resonance pattern that follow the zeros of

$J_{-1/3}(\zeta_0)$, that is:

$$J_{-1/3}(\zeta_0) \sim \cos\left(\frac{2p^3}{3K^2} - \frac{1}{12}\pi\right) = 0 \Rightarrow$$
$$\frac{2p^3}{3K^2} - \frac{1}{12}\pi \simeq \frac{2K^3L^{3/2}}{3K^2} - \frac{1}{12}\pi = \pi\left(n + \frac{1}{2}\right) \Rightarrow$$
$$\left(\frac{2}{3}KL^{3/2}\right)^2 = \left(n + \frac{7}{12}\pi\right) \Rightarrow$$
$$u_n = \frac{3\pi^2}{16}L\left(n + \frac{7}{12}\right)^2. \tag{7.45}$$

Thus, we have derived an analytic equation for the position of resonances, just as we did for the spherical well, and this expression agrees exactly with the resonances in figure 7.6.

Hulthén potential

Comparing equation (7.25) and (7.45), we notice that they look very similar. Only the fraction in front and the 'phase' is different. It seems likely that we may fit the peaks of the resonances in the Sommerfeld case by an expression:

$$u_n = L\pi^2\left(n + b\right)^2, \tag{7.46}$$

and this is indeed the case. We find the values $L = 0.1592$ and $b = 1.006$. The value of b is very close to 1, hinting that a similar treatment is doable in the Yukawa case, yielding a boost factor which depends only on a sine to lowest order. A model potential called the Hulthén potential admits an analytic solution for the s-wave case[2], and the Sommerfeld boost coming from this potential was studied in [43] and later in [44].

The potential looks like

$$V_{\mathrm{H}} = -\frac{A\delta e^{-\delta r}}{1 - e^{-\delta r}}, \tag{7.47}$$

where A and δ are parameters. We stress that this is not a general version of the Yukawa potential, but it is a model potential just like the spherical well and the spherical slope well. However, unlike these two potentials, the Hulthén potential reproduces the $1/r$-behavior of the Yukawa potential in the limit $r \to 0$, and it decays exponentially instead of having a fixed range. By a procedure similar[3] to ours, A and δ is found to be $A = \alpha$ and $\delta = km_\phi = \pi^2/6$. In our notation, the $l = 0$ case of the Sommerfeld boost just before equation 44 in [43], can be written as

$$S_{\mathrm{H}} = \frac{\frac{\pi\alpha}{\beta}\sinh\left(2\pi\frac{\beta}{kf}\right)}{\cosh\left(2\pi\frac{\beta}{kf}\right) - \cos\left(2\pi\frac{\beta}{kf}\sqrt{\frac{k\alpha f}{\beta^2} - 1}\right)}. \tag{7.48}$$

Equation (7.48) can be derived easily under the assumption $k\alpha f \leq \beta^2$, (which

[2] We wish to thank the referee for bringing this to our attention.

[3] [43] argues using the Lippmann-Schwinger equation, that the first moments of the potentials should be set equal. In our approach we set the volumes equal instead, i.e. the second moment. We checked how using the first moment would affect our results, and the effect was only to change the factor in front of equation (7.25) and (7.45) as well as making them independent of L.

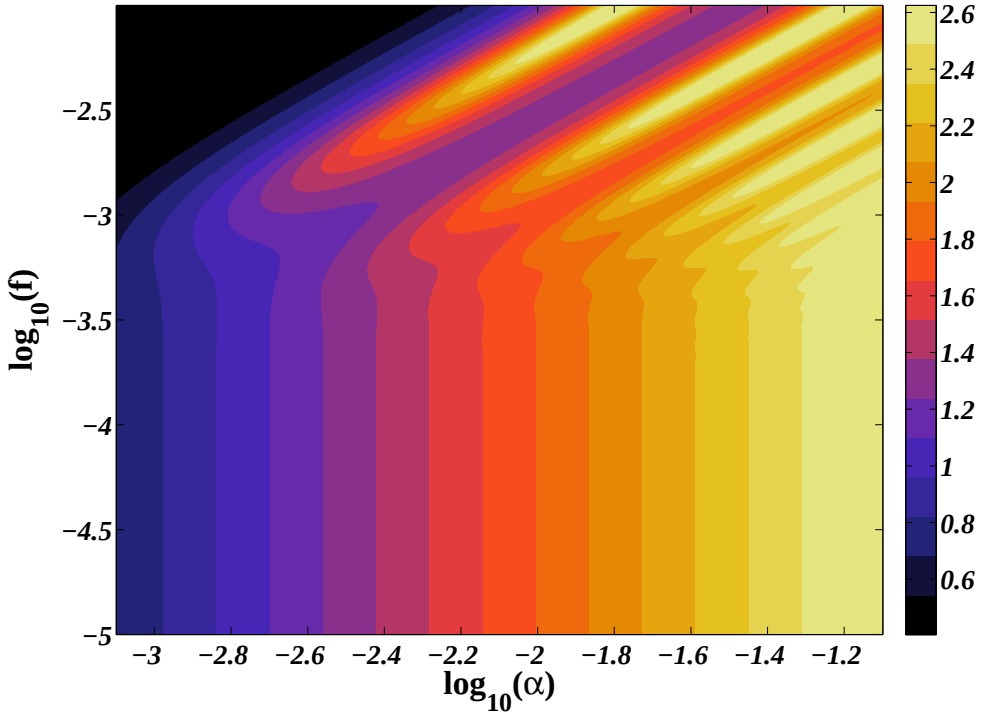


Figure 7.7: Sommerfeld boost factor in $\log_{10}$ for the Hulthén potential at a relative velocity of 150km/s.

is usually not the case), however, a more careful derivation reveals this equation to be true in any case. We have plotted the Hulthén boost factor in figure 7.7, and it looks very similar to the Yukawa case in figure 7.1. Inspection of equation (7.48) reveals the resonance pattern to be

$$\begin{aligned} u_n &= k\,(n+1)^2 \\ &= \frac{\pi^2}{6}\,(n+1)^2\,. \end{aligned} \tag{7.49}$$

However, for large n, the resonance pattern is shifted completely because of the slight difference between $\pi^2/6$ and the fit value. For a velocity of 150km/s we found equation (7.48) to be within 10% of the numerical solution at 82% of the plotted parameter space and within 30% at 98% of the parameter space. For a velocity of 10km/s, the Hulthén boost factor is within 10% in only 43% of the parameter space and within 30% in 74% of the parameter space. By using $k = 0.1592\pi^2$ and simultaneously correcting the phase by 0.006 by hand in equation (7.48), according to the fit parameters, we did somewhat better. The modified formula reproduced the resonance pattern perfectly, but the agreement of the boost magnitude at the first peak became worse.

7.3 Constraining Sommerfeld Enhancement with the CMB

There has been some work on constraining models incorporating Sommerfeld Enhancement using the CMB. WIMPs annihilating after recombination may contribute to reionisation, thereby changing the predicted optical depth which can be inferred from the CMB anisotropy spectrum. This was studied by [42]. Earlier, during the recombination phase, WIMP annihilations with Sommerfeld Enhancement also modify the standard scenario. An upper bound on this effect can again be inferred from the anisotropy spectrum, as was done in [35–41]. We will focus on WIMP annihilations happening even earlier, in the redshift range of approximately $1100 < z < 2.1 \cdot 10^6$. Annihilations occurring in this redshift range will not influence the anisotropy spectrum, but they will distort the Black Body spectrum.

Common to the above mentioned bounds is the fact that they use the expected cross section for a thermal relic, $\langle \sigma v \rangle_{\mathrm{TH}} \simeq 3 \cdot 10^{-26} \mathrm{cm}^3/\mathrm{s}$, and multiply this with the boost factor to obtain the effective annihilation cross section. As was pointed out in [7], this is not strictly correct, since the freeze out process is modified by the boost mechanism. So to consistently probe the parameter space, we must first find the relativistic annihilation cross section σ_0, which gives the correct relic abundance.

Relic density calculation

To catch the full nature of thermal freeze out in models with Sommerfeld enhancement, we must do a full calculation of the integrated Boltzmann equation because of the non-trivial velocity dependence. It should be noted, that we do expect something of the same order of magnitude as $\langle \sigma v \rangle_{\mathrm{TH}}$, since the Sommerfeld boost factor is very close to 1 at the time of freeze out. We start from the integrated Boltzmann equation for Dark Matter,

$$a^{-3} \frac{\mathrm{d}(n_\chi a^3)}{\mathrm{d}t} = \langle \sigma_{\mathrm{ann}} v \rangle \left\{ n_{\chi,\mathrm{eq}}^2 - n_\chi^2 \right\}, \tag{7.50}$$

where a is the scale factor, n_χ is the number density of Dark Matter, $n_{\chi,\mathrm{eq}}$ is the Dark Matter equilibrium number density and $\langle \sigma_{\mathrm{ann}} v \rangle$ is the thermally averaged annihilation cross section. We normalize the number density to the total entropy density $s \propto a^{-3}$ by introducing $Y \equiv \frac{n_\chi}{s}$:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = s \langle \sigma_{\mathrm{ann}} v \rangle \left\{ Y_{\mathrm{eq}}^2 - Y^2 \right\} \tag{7.51}$$

We also want to substitute the time parameter t by the dimensionless evolution parameter $x \equiv \frac{m_\chi}{T_\gamma}$. After some manipulations, we find the final form of the Boltzmann equation:

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = \sqrt{\frac{\pi}{45}} \frac{m_p m_\chi}{x^2} h_*^{\frac{1}{2}} \langle \sigma_{\mathrm{ann}} v \rangle \left\{ Y_{\mathrm{eq}}^2 - Y^2 \right\}. \tag{7.52}$$

The manipulations leading up to equation (7.52) can be found in section 7.5 of the appendix. In our case, $\sigma = \sigma_0 S(v, \ldots)$, where σ_0 is the s-wave annihilation cross section which is independent of velocity, and S is the Sommerfeld boost

factor, which depends on velocity and the model parameters. As was noted by the authors of [29, 30], σ_0 is not a completely free parameter but is related to α and m_χ by dimensional analysis, $\sigma_0 \sim \frac{\alpha^2}{m_\chi^2}$. However, the exact factor is highly model dependent, and for that reason we keep σ_0 as a parameter.

Inserting the Sommerfeld cross section in equation (7.52) leads to our final equation:

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = \frac{m_p m_\chi}{x^2} h_*^{\frac{1}{2}} \sigma_0 \langle S(v,\ldots)v\rangle \left\{Y_{\mathrm{eq}}^2 - Y^2\right\} \tag{7.53}$$

$$\equiv \lambda(x) \left\{Y_{\mathrm{eq}}^2 - Y^2\right\}. \tag{7.54}$$

The parameter σ_0 is then found by imposing the boundary condition that the WIMP must make up all of Dark Matter. In this article we are considering the case where the WIMP is not its own antiparticle, so we impose the boundary condition $\Omega_{\mathrm{DM}} = 2\Omega_{\chi,0}$ assuming that there is no asymmetry between χ and its antiparticle.

The equilibrium number density Y_{eq} can be calculated exactly under the assumption that Dark Matter follows a Maxwell-Boltzmann distribution. We got

$$Y_{\mathrm{eq}} = g_i \frac{45}{4\pi^4} \frac{x^2}{g_{*S}} K_2(x), \tag{7.55}$$

where K_2 is the modified Bessel function of the second kind. The derivation of equation (7.55) can be found in section 7.5 of the appendix. The values of g_{*S} in equation (7.55) and $h_*^{\frac{1}{2}}$ in equation (7.53) depends on the temperature, and it is found by interpolation in a precomputed table. The number of internal degrees of freedom, g_i, was set to 2. We also need the thermally averaged cross section. If we approximate the distribution function for the Dark Matter gas by the non-relativistic Maxwell-Boltzmann distribution

$$f(\beta) = \sqrt{\frac{2}{\pi}} x^{\frac{3}{2}} \beta^2 e^{-\frac{1}{2}x\beta^2}, \tag{7.56}$$

the thermal average of S becomes

$$\langle S(\beta,\ldots)v\rangle \simeq \sqrt{\frac{2}{\pi}} x^{\frac{3}{2}} \int_0^1 S(\beta,\ldots)\beta^2 e^{-\frac{1}{2}x\beta^2}. \tag{7.57}$$

This distribution is only normalized to 1 when the upper limit goes to infinity, but if we are in the non-relativistic limit this is a small correction. (In the calculation we set the upper limit dynamically to 4 times the position of the peak of the distribution.) Before freeze out the gas is not strictly non-relativistic and we should instead use the relativistic distribution:

$$f(\gamma)\mathrm{d}\gamma = \frac{x}{K_2(x)} \beta\gamma^2(\beta) e^{-x\gamma(\beta)} \Rightarrow \tag{7.58}$$

$$f(\beta)\mathrm{d}\beta = \frac{x}{K_2(x)} \gamma^5 \beta^2 e^{-\gamma x}, \tag{7.59}$$

where γ is the usual gamma factor and K_2 is the modified Bessel-function of the second kind. The thermal average then becomes

$$\langle Sv\rangle = \frac{x}{k_2(x)} \int_0^1 S(\beta,\ldots)\gamma^5\beta^2 e^{-x(\gamma-1)}, \tag{7.60}$$

where $k_2(x) = K_2(x)e^x$. When γ becomes close to 1 we must use the series expansion to calculate $1 - \gamma$ for numerical stability. The difference in the required cross section σ_0 from using (7.60) compared to (7.57) was negligible, however.

The numerical solution of equation (7.54) is not entirely trivial, especially not since the equation is stiff and we need to solve it repeatedly. In section 7.5 of the appendix, we have described the numerical scheme we use for this problem, which is the same used in the DarkSUSY software package [11]. We recommend this scheme to others interested in Freeze-Out calculations.

Distorting the Black Body spectrum

When energy is injected into the CMB photons, two types of processes are needed to restore a black body spectrum: Number changing processes and equilibrating processes. Double Compton scattering and bremsstrahlung belongs to the first category, while Compton scattering and inverse Compton scattering belongs to the second. Double Compton scattering freezes out at $z_{\mathrm{DC}} \simeq 2.1 \cdot 10^6$ and Compton scattering freezes out at $z_{\mathrm{C}} \simeq 5.4 \cdot 10^4$. Energy which is deposited in the photon gas after z_{DC} but before z_{C} will be redistributed to give an entropy maximising Planck spectrum with a chemical potential μ. Energy input after z_{C}, (but before recombination), can not equilibrate and will result in a Compton-y distortion [28] of the Planck spectrum.

The relevant quantity for deriving the size of both effects is thus the relative energy input to the CMB during these two epochs. Following [27], we write

$$\begin{aligned}\frac{\delta\rho_\gamma}{\rho_\gamma} &= \int_{t_1}^{t_2} \frac{\dot{\rho_{\mathrm{ann}}}}{\rho_\gamma}\mathrm{d}t \\ &= \int_{t_1}^{t_2} \frac{2Fm_\chi \langle\sigma_{\mathrm{ann}}v\rangle\, n_\chi^2}{\rho_{\gamma,0}(1+z)^4}\mathrm{d}t, \end{aligned} \tag{7.61}$$

where ρ_γ is the energy density of the CMB photons and z is the redshift. F denotes the fraction of energy which is transfered to the CMB photons. This is independent of redshift when $z \gtrsim 2500$, and according to table 1 of [35], this is between 30% and 90% depending on the annihilation channel. The factor of 2 in equation (7.61) stems from the fact that we, in consistence with our freeze out calculation, assume that χ is not its own antiparticle.

We introduce the following relations:

$$H^2(t) = \frac{4\pi^3}{45m_p^2} g_* T_{\gamma,0}^4 (1+z)^4 \Rightarrow \tag{7.62a}$$

$$\frac{\mathrm{d}t}{\mathrm{d}z} = -\sqrt{\frac{45}{\pi}}\frac{m_p}{2\pi} g_*^{-\frac{1}{2}} T_{\gamma,0}^{-2}(1+z)^{-3} \tag{7.62b}$$

$$\begin{aligned} n_\chi &= n_{\chi,0}(1+z)^3 = \frac{\rho_{\chi,0}}{m_\chi}(1+z)^3 \\ &= \Omega_\chi \frac{3H_0^2 m_p^2}{8\pi m_\chi}(1+z)^3 \end{aligned} \tag{7.62c}$$

$$\rho_\gamma = \rho_{\gamma,0}(1+z)^4 \tag{7.62d}$$

$$T_\gamma = T_{\gamma,0}(1+z) \tag{7.62e}$$

$$\langle\sigma_{\mathrm{ann}}v\rangle = \sigma_0 \langle S(\beta,\ldots)v\rangle \tag{7.62f}$$

Using equations (7.62) in equation (7.61) yields:

$$\begin{aligned}\frac{\delta\rho_\gamma}{\rho_\gamma} &= \frac{405}{64\pi^5}\sqrt{\frac{5}{\pi}}g_*^{-\frac{1}{2}}F\frac{\sigma_0}{m_\chi}\Omega_\chi^2\left[\frac{m_p^5H_0^4}{T_{\gamma,0}^6}\right]\times \\ &\times\int_{z(t_2)}^{z(t_1)} S_{\text{avg}}(z,\ldots)\frac{1}{1+z}\mathrm{d}z \\ &= \frac{405}{64\pi^5}\sqrt{\frac{5}{\pi}}g_*^{-\frac{1}{2}}f\left[\frac{100\text{GeV}}{m_\chi}\right]\left[\frac{\sigma_0}{10^{-26}\text{cm}^3/\text{s}}\right]\times \\ &\times(\Omega_\chi h)^2\,C_{-7}\int_{z(t_2)}^{z(t_1)} S_{\text{avg}}(z,\ldots)\frac{1}{1+z}\mathrm{d}z. \end{aligned} \tag{7.63}$$

Here $C_{-7}\simeq 2.8696\cdot 10^{-7}$ is a numerical constant. Since the Dark Matter is very cold at this time, the distribution function is strongly peaked around its mean value, so to a good approximation, we may write $S_{\text{avg}}(z)\simeq S(\beta_{\text{mean}}(z))$. We may also assume that the Sommerfeld enhancement has saturated at this stage, making the approximation $S(\beta_{\text{mean}}(z))\simeq S(\beta_{\text{mean}}(z(t_1)))$, allowing us to pull S outside the integral which can then be done analytically:

$$\begin{aligned}\frac{\delta\rho_\gamma}{\rho_\gamma} &\approx \frac{405}{64\pi^5}\sqrt{\frac{5}{\pi}}g_*^{-\frac{1}{2}}F\left[\frac{100\text{GeV}}{m_\chi}\right]\left[\frac{\sigma_0}{10^{-26}\text{cm}^3/s}\right]\times \\ &(\Omega_\chi h)^2\,C_{-7}S(\beta_{\text{mean}}(z_1))\ln\left|\frac{1+z(t_1)}{1+z(t_2)}\right|. \end{aligned} \tag{7.64}$$

In these calculations we have assumed a radiation dominated universe as described by the Friedmann equation (7.62a), which breaks down at $z\sim z_{\text{eq}}\simeq 3300$, well before recombination. But as can be seen from the analytical approximation, the dependence on $z(t_2)$ is logarithmic, so the error in doing this is insignificant. When plotting the μ and $|y|$-distortions, we have put $F=1$ for convenience, but since the dependence is linear in F, it can be reinstated by multiplying the distortions by F.

Results

For each point in the (α, f)-parameter space, we solved the Boltzmann equation from $x_1=1$ to $x_2=m_\chi/T_{\gamma,0}$ to find the value of σ_0. This depends weakly on the WIMP mass as well as the kinetic decoupling temperature but the overall dependence on the Sommerfeld parameters can be seen in figure 7.8. The kinetic decoupling temperature was taken as a parameter, and the effect on the freeze out process is shown on figure 7.9. We find that this effect can be as large as 30% in agreement with [7, 8]. We expect σ_0 to be nearly independent of the WIMP mass and this is confirmed by figure 7.10 which shows maximally one percent difference between a 200GeV WIMP and a 1000GeV WIMP for a kinetic decoupling value of $x_{\text{KD}}=2\cdot 10^3$ for both. However, in a real particle physics model we expect x_{KD} to have a slight mass dependence, so the actual effect of different masses may be somewhat bigger [9].

For each point in the parameter space, we solve the integral in equation (7.63) on both the interval which is relevant for Compton-$|y|$ distortions as well as the interval relevant for the CMB photons to develop a chemical potential μ. The results for a 200GeV particle are shown in figure 7.11 and 7.12.

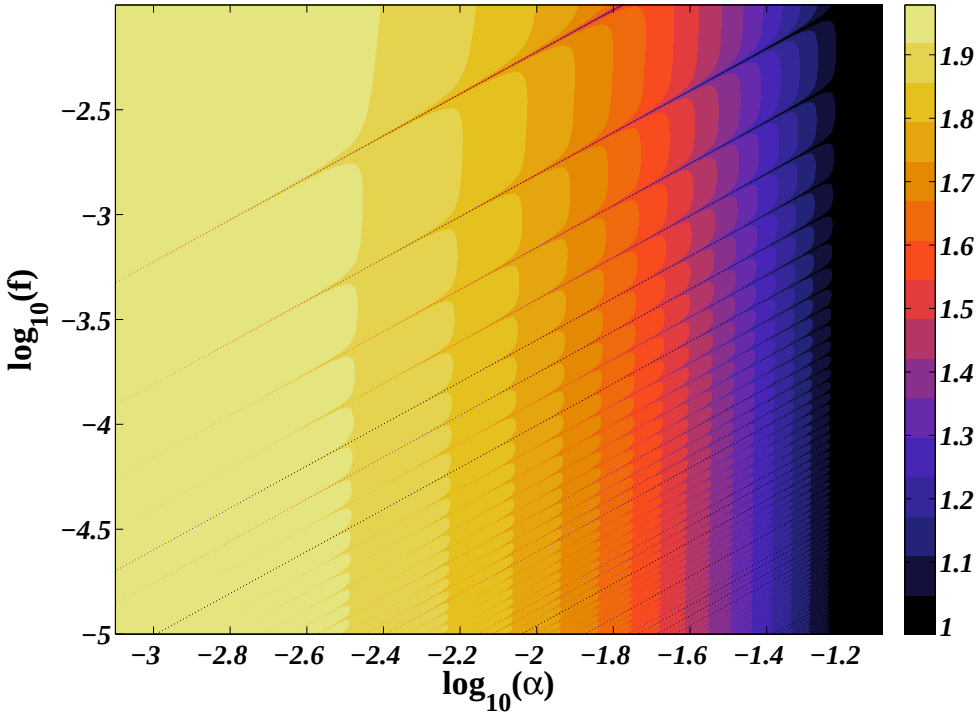


Figure 7.8: Required annihilation cross section σ_0 in units of $10^{-26}\mathrm{cm}^3/\mathrm{s}$ to explain the total Dark Matter abundance for a 200GeV WIMP and a kinetic decoupling temperature of 8MeV.

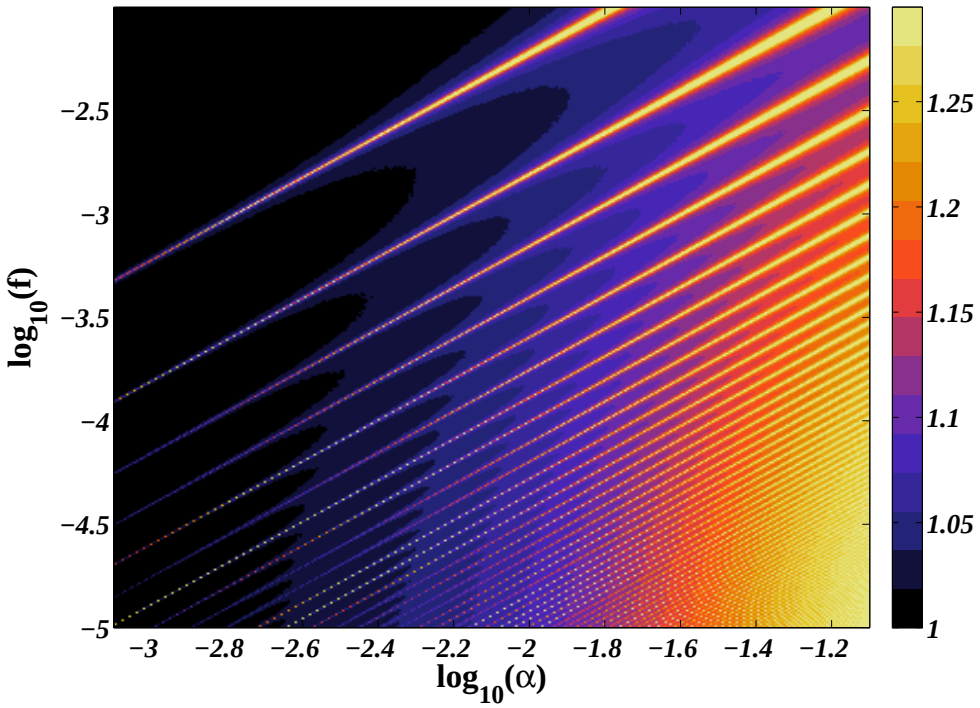


Figure 7.9: The ratio $\sigma_0^{\mathrm{KD=8MeV}}/\sigma_0^{\mathrm{KD=500MeV}}$ for a 200GeV WIMP.

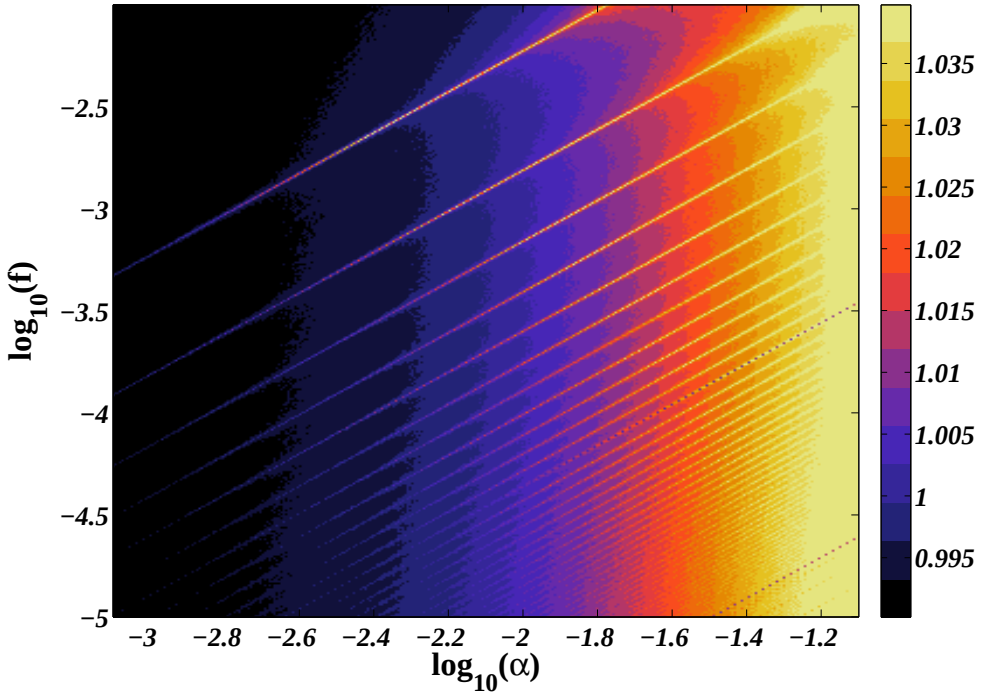


Figure 7.10: The ratio σ_0/σ_0' of cross sections for a 200GeV WIMP and a 1000GeV WIMP respectively, which decouples kinetically at $x_{\mathrm{KD}} = 2 \cdot 10^3$.

Considering the current bounds on μ- and y-distortions from FIRAS [6] of $|\mu| < 9 \cdot 10^{-5}$, $|y| < 1.5 \cdot 10^{-5}$, only a small portion of the parameter space can be ruled out. But there is a rather large part of lower right part of the parameter space which is close at saturating the current bound. The analysis for the μ-distortion was already carried out in [8], and our results agrees with theirs.

As is evident from figure 7.11 and 7.12, the two bounds are degenerate in the way that they both tends to rule out the resonances and the lower right part of the parameter space. With the current limits, the $|y|$-bound is always as strong or stronger than the μ-bound. One would suspect this to be the case, since the y-distortion happens at a later time than the μ-distortion, giving the WIMPs more time to cool. This is indeed true, but only for a small subset of the parameter space, at the resonances in the lower left part. For the rest of the parameter space, the Sommerfeld enhancement has already saturated at this point, and no further enhancement is possible.

We can consider what would be the allowed possibilities for the annihilation cross section in a halo having a fiducial velocity dispersion of 150km/s by removing all points that exceeds either bound. We have plotted this in figure 7.13 for our 200GeV example WIMP. It is suspected that a new FIRAS-like satellite, if built, could bring the bound on $|y|$ down to the order $\sim 10^{-7}$. In figure 7.14 we have shown what figure 7.13 would look like with a fiducial bound of $|y| < 10^{-7}$. It is also worth comparing this bound with the anisotropy

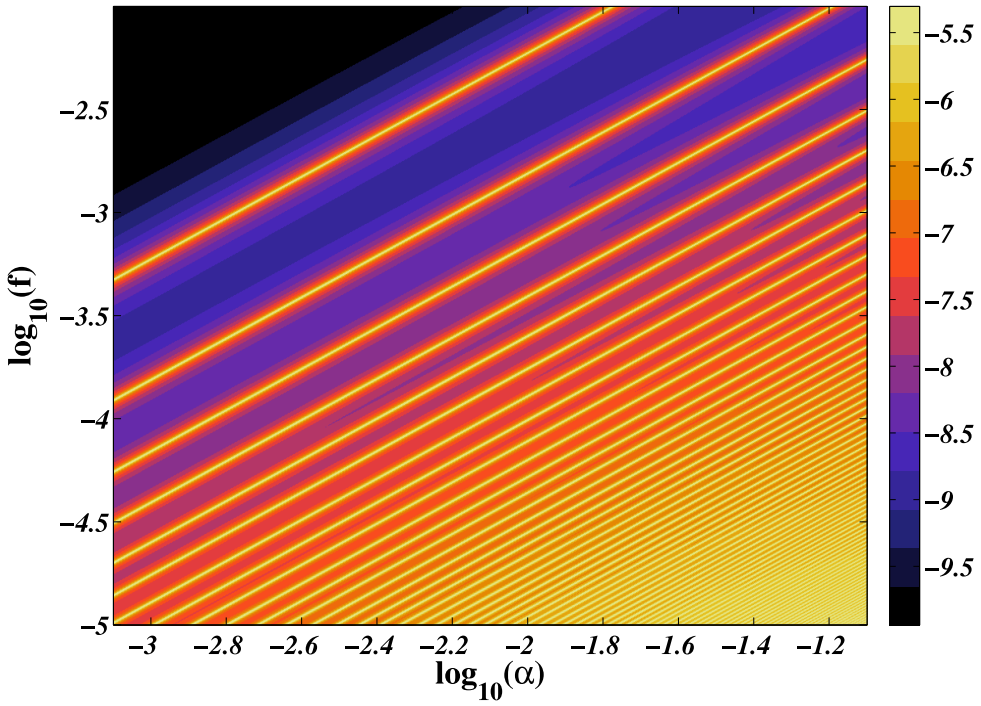


Figure 7.11: The magnitude of the Compton-$|y|$ parameter in $\log_{10}$ for a 200GeV WIMP and a kinetic decoupling temperature of 8MeV.

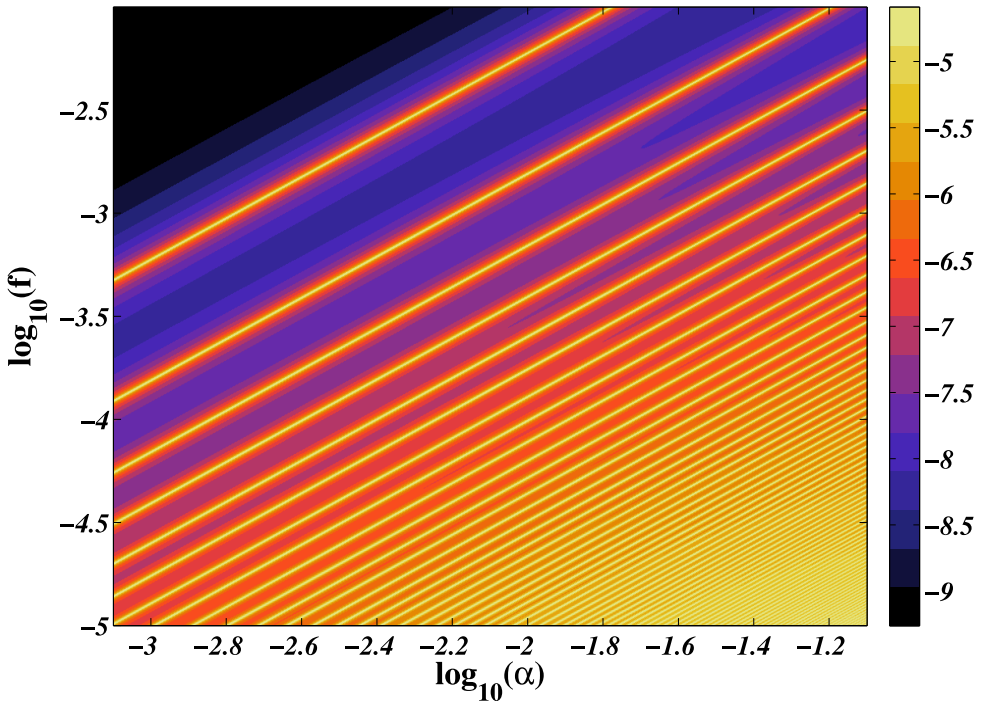


Figure 7.12: The magnitude of the chemical potential μ in $\log_{10}$ for a 200GeV WIMP and a kinetic decoupling temperature of 8MeV.

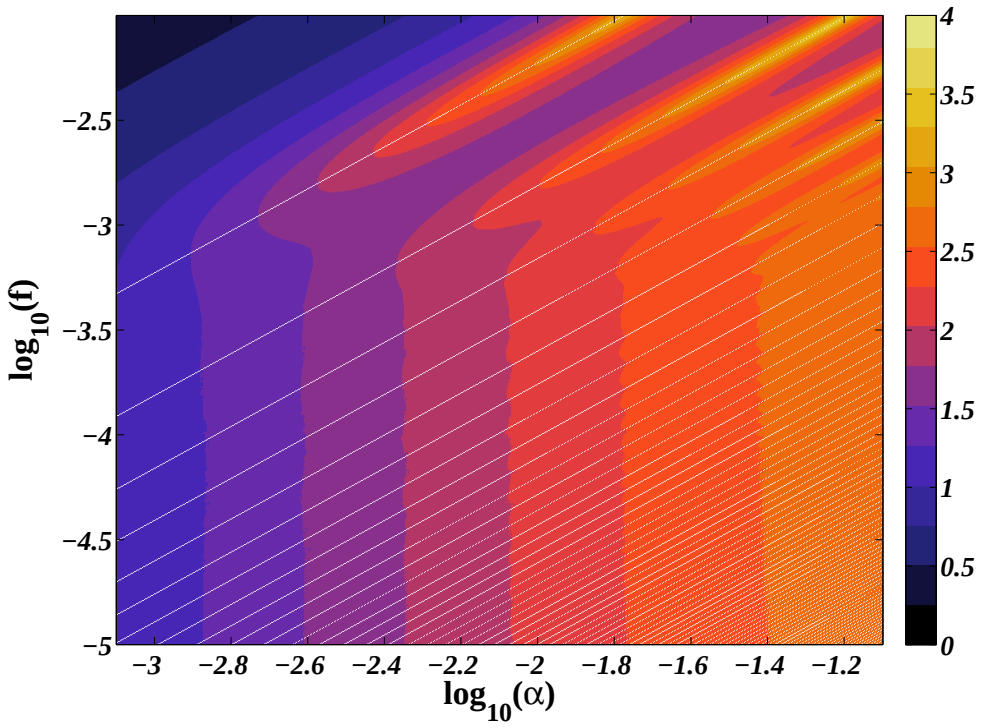


Figure 7.13: Logarithm of the annihilation cross section in units of $10^{-26}\mathrm{cm}^3/\mathrm{s}$ in a halo with velocity dispersion of approximately 150km/s. White points are ruled out by either μ- or $|y|$-bound.

bound. From [35], we have the bound

$$\langle\sigma_{\mathrm{ann}}v\rangle_{\mathrm{saturated}} < \frac{360\cdot 10^{-26}\mathrm{cm}^3/\mathrm{s}}{F_{\mathrm{rc}}}\frac{m_\chi}{1\mathrm{TeV}}, \tag{7.65}$$

where F_{rc} is the average fraction of energy being transfered to the CMB at the time of recombination. This has different values depending on the annihilation channel, as can be seen in table 1 of [35]. However, it is roughly of order 30% for most processes.

As we discussed earlier, plugging in the standard value $\langle\sigma_{\mathrm{ann}}v\rangle_{\mathrm{TH}}$ for a thermal relic is not strictly accurate. It is easily fixed however, by using our calculated values of σ_0, as well as the thermally averaged Sommerfeld factor at recombination. We get

$$\frac{\sigma_0}{10^{-26}\mathrm{cm}^3/\mathrm{s}}S_{\mathrm{avg}}(z_{\mathrm{rc}},\ldots) < \frac{360}{F_{\mathrm{rc}}}\frac{m_\chi}{1\mathrm{TeV}}. \tag{7.66}$$

In figure 7.15 we have again showed the effective boost factor, but this time with the CMB bound, equation (7.66). It is clear, that this bound is stronger than even the forecasted y-bound by roughly an order of magnitude.

7.4 Conclusion

We have analysed the Sommerfeld enhancement mechanism in detail and found coordinates in which the position of resonances depends only on one coordi-

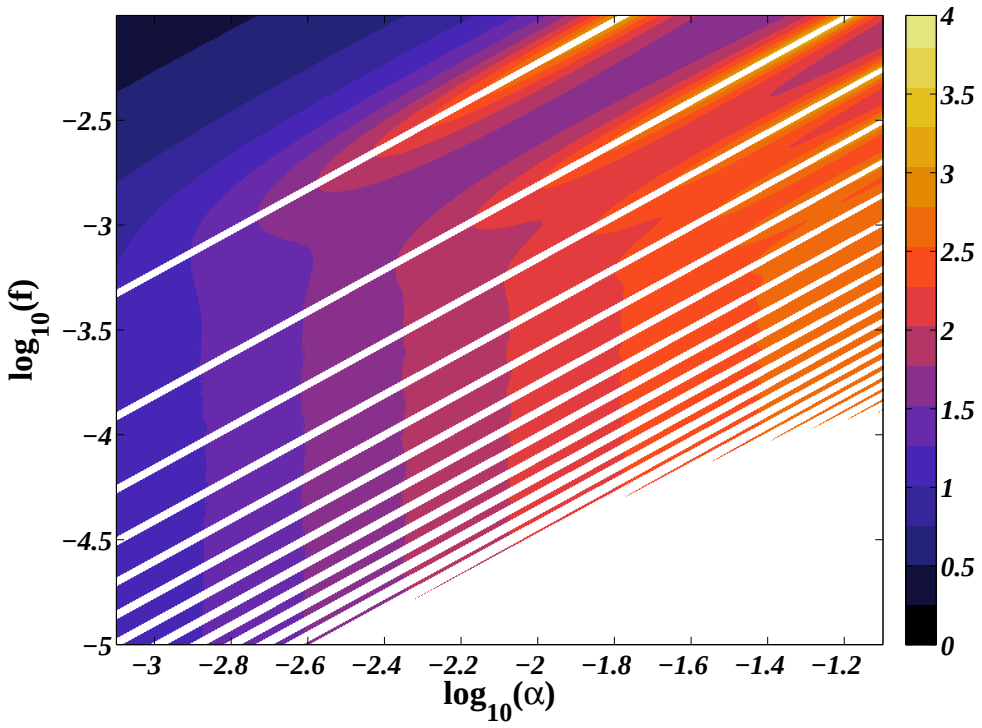


Figure 7.14: Logarithm of the annihilation cross section in units of $10^{-26}\mathrm{cm}^3/\mathrm{s}$ in a halo with velocity dispersion of approximately 150km/s. White points can be ruled out if the bound gets improved by a factor of 500.

nate, u, and we have derived analytical expressions for the position of resonances for two model potentials. A numeric treatment of the Sommerfeld case showed that the a similar simple relationship exists for this potential, and we found the following fit to agree excellently:

$$u_n = 0.1592\,(n + 1.006)^2 \qquad n = 0, 1, \ldots \tag{7.67}$$

This is a nice result, since knowing the position of resonances beforehand is helpful for doing numerical calculations.

In the second part we did the full freeze out calculation for the Sommerfeld parameter space by solving the Boltzmann equation all the way through chemical and kinetic freeze out, and we then calculated the effect of annihilations on the CMB blackbody spectrum. We found that only a small portion of the parameter space, directly on the resonances, can be ruled out by the current bound from FIRAS. We also noted that a future measurement of the $|y|-$ or $\mu-$distortion would rule out a huge portion of the parameter space. But as we showed, the anisotropy bound is already stronger than this and is likely to improve with Planck data. However, we think that it is worth mentioning, that the $|y|-$ and $\mu-$bounds are obtained at a different epoch and by different observations than the anisotropy bounds, and thus should be considered complementary to those.

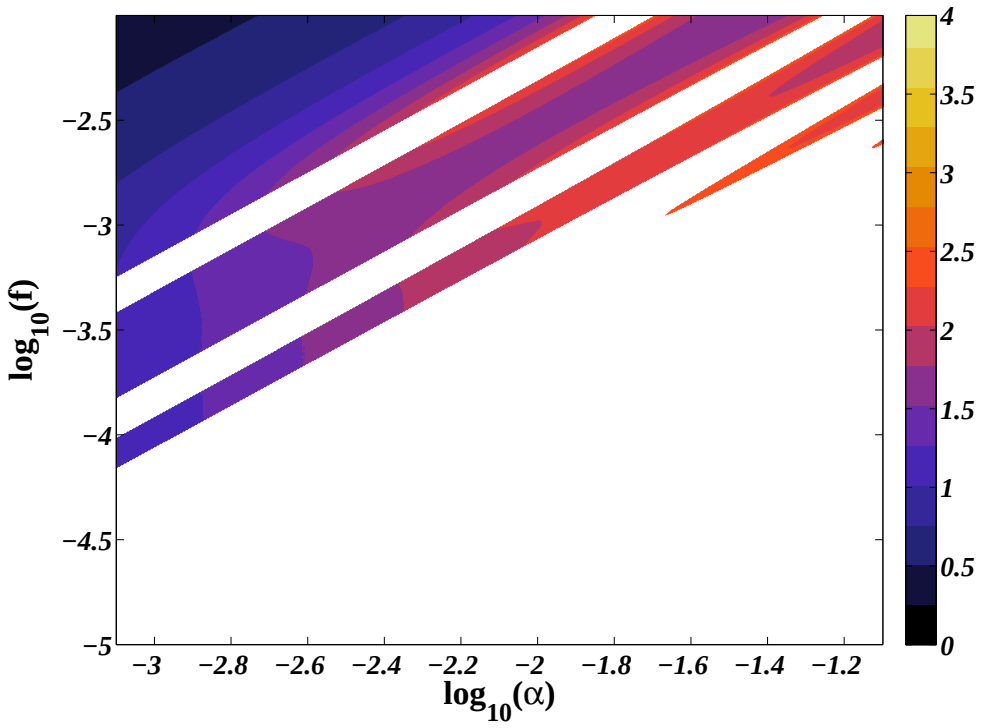


Figure 7.15: Logarithm of the annihilation cross section in units of $10^{-26}\mathrm{cm}^3/\mathrm{s}$ in a halo with velocity dispersion of approximately 150km/s. White points are ruled out by the CMB anisotropy bound.

7.5 Appendix

Sommerfeld enhancement for the Coulomb potential

We want to derive equation (7.13), the boost factor for the Coulomb potential. We start from the radial equation (7.5) with $m_\phi = 0$ and do the substitution $x = \alpha m_\chi r$:

$$\begin{aligned}\frac{1}{m_\chi}\frac{\mathrm{d}^2\chi}{\mathrm{d}r^2} &= \left(-\frac{\alpha}{r}e^{-m_\phi r} - m_\chi\beta^2\right)\chi \Rightarrow \\ \frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} &= -\left[\frac{1}{x} + \left(\frac{\beta}{\alpha}\right)^2\right]\chi \Rightarrow \\ &= \left[\frac{1}{x} + \epsilon_C^2\right]\chi. \end{aligned} \tag{7.68}$$

We analyze equation (7.68) the usual way by considering the asymptotic limits of the equation. We got:

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} \simeq -\frac{1}{x}\chi, \qquad x \to 0 \tag{7.69a}$$

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} \simeq -\epsilon_C\chi, \qquad x \to \infty. \tag{7.69b}$$

The analytic solution of (7.69b) is just an exponential, while the analytic solution of (7.69b) is more complicated. In general we have

$$\chi(x) = C_1\sqrt{(}x)J_1(2\sqrt{x}) + C_2\sqrt{x}Y_1(2\sqrt{x}),$$

where J_1 and Y_1 are the Bessel functions of the first and second order, respectively. However, $\sqrt{x}Y_1(2\sqrt{x})$ is not well behaved for $x \to 0$, so we are left with the C_1 term. Since we are looking at $x \ll 1$, we may use $J_1(z) \sim z/2$. Our Ansatz for the solution then becomes:

$$\chi = xe^{i\epsilon_C x}v(x), \tag{7.70}$$

where $v(x)$ interpolates between the asymptotic solutions. Taking the second derivative of the Ansatz (7.70) yields

$$\frac{\mathrm{d}^2\chi}{\mathrm{d}x^2} = e^{i\epsilon_C x}\left[(2i\epsilon_C - \epsilon_C^2)v + (2 + 2i\epsilon_C x)v' + xv''\right],$$

which can be inserted into equation (7.68). The prime on v denotes differentiation w.r.t x. The result is

$$0 = xv'' + (2 + 2i\epsilon_C x)v' + (2i\epsilon_C + 1)v. \tag{7.71}$$

We can bring this equation on a more recognizable form by making the substitution $z = -2i\epsilon_C x$. This leads to

$$0 = z\frac{\mathrm{d}^2v}{\mathrm{d}z^2} + (2 - z)\frac{\mathrm{d}v}{\mathrm{d}z} - \left(1 - \frac{i}{2\epsilon_C}\right)v, \tag{7.72}$$

which is the confluent hyper geometric equation. The solution which is regular in $x = z = 0$ is then:

$$v(z) = CF\left(1 - \frac{i}{2\epsilon_C}, 2, z\right). \tag{7.73}$$

We now need to apply the asymptotic boundary condition (7.8b) to our solution, so we need the asymptotic formula for $F(a, b, z)$ in the limit $x \to \infty$ or, equivalently, $z \to -i\infty$. To do this, define the following:

$$\begin{aligned} g(a, b, z) &\equiv \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{n!}z^{-n} \\ (c)_n &= c(c + 1)(c + 2)\cdots(c + n - 1), (c)_0 = 1 \\ F(a, b, z) &= \frac{\Gamma(b)}{\Gamma(b - a)}g(a, 1 + a - b, -z)(-z)^{-a} + \\ &\quad \frac{\Gamma(b)}{\Gamma(a)}g(b - a, 1 - a, z)e^z z^{a-b}. \end{aligned}$$

Since $g(a, b, z) \to 1$ as $|z| \to \infty$, $F(a, b, z)$ has the asymptotic form

$$F^\infty(a, b, z) = \frac{\Gamma(b)}{\Gamma(b - a)}(-z)^{-a} + \frac{\Gamma(b)}{\Gamma(a)}e^z z^{a-b}.$$

Now consider the gamma-function $\Gamma(b-a)$ for our values of $a = 1 - i/(2\epsilon_C)$ and $b = 2$:

$$\Gamma(b-a) = \Gamma\left(2 - \left(1 - \frac{i}{2\epsilon_C}\right)\right) = \Gamma(a^*) = \Gamma(a)^*,$$

and if we now define η by $\Gamma(a) = |\Gamma(a)| e^{i\eta}$, we find the following:

$$\frac{\Gamma(b)}{\Gamma(b-a)} = \frac{\Gamma(b)}{|\Gamma(a)|} e^{i\eta} \tag{7.75a}$$

$$\frac{\Gamma(b)}{\Gamma(a)} = \frac{\Gamma(b)}{|\Gamma(a)|} e^{-i\eta}. \tag{7.75b}$$

The asymptotic form F^∞ can now be written like:

$$F^\infty(a,b,z) = \frac{\Gamma(b)}{|\Gamma(a)|}\left(e^{i\eta}(-z)^{-a} + e^{-i\eta+z} z^{a-b}\right) \Rightarrow$$

$$F^\infty\left(1 - \frac{i}{2\epsilon_C}, 2, -2i\epsilon_C x\right) = \frac{e^{\frac{1}{2}z} e^{i\eta - \frac{\pi}{4\epsilon_C}}}{\Gamma\left(1 - \frac{i}{2\epsilon_C}\right)\epsilon_C x} \times$$

$$\times \sin\left(\epsilon_C x + \frac{1}{2\epsilon_C}\ln(2\epsilon_C x) + \eta\right).$$

Substituting the asymptotic solution back into the Ansatz (7.70) yields

$$\chi^\infty(x) = C_1 \frac{e^{i\eta - \frac{\pi}{4\epsilon_C}}}{\Gamma\left(1 - \frac{i}{2\epsilon_C}\right)\epsilon_C} \times \sin\left(\epsilon_C x + \frac{1}{2\epsilon_C}\ln(2\epsilon_C x) + \eta\right), \tag{7.76}$$

where C_1 is now set by the boundary condition (7.8b). We find

$$C_1 = \frac{\epsilon_C \Gamma\left(1 - \frac{i}{2\epsilon_C}\right)}{e^{i\eta - \frac{\pi}{4\epsilon_C}}}, \tag{7.77}$$

which we can finally insert into (7.73) and the Ansatz (7.70):

$$\chi(x) = x e^{i\epsilon_C x} \frac{\epsilon_C \Gamma\left(1 - \frac{i}{2\epsilon_C}\right)}{e^{i\eta - \frac{\pi}{4\epsilon_C}}} \times F\left(1 - \frac{i}{2\epsilon_C}, 2, -2i\epsilon_C x\right).$$

We are interested in the boost factor (7.9), so we calculate

$$S_{\epsilon_C} = \left|\frac{R_{k0}}{k}\right|^2 = \left|\frac{1}{\epsilon_C}\frac{\mathrm{d}\chi}{\mathrm{d}x}(x=0)\right|^2. \tag{7.78}$$

We take the derivative of our solution (7.78) and evaluate it in $x = 0$:

$$\frac{\mathrm{d}\chi}{\mathrm{d}x}(x=0) = \epsilon_C \Gamma\left(1 - \frac{i}{2\epsilon_C}\right) e^{-i\eta + \frac{\pi}{4\epsilon_C}} \Rightarrow$$

$$\begin{aligned} S_{\epsilon_C} &= \left|\Gamma\left(1 - \frac{i}{2\epsilon_C}\right) e^{-i\eta + \frac{\pi}{4\epsilon_C}}\right|^2 \\ &= \left|\Gamma\left(1 - \frac{i}{2\epsilon_C}\right)\right|^2 e^{\frac{\pi}{2\epsilon_C}} \\ &= \frac{-\frac{\pi}{2\epsilon_C}}{\sinh\left(-\frac{\pi}{2\epsilon_C}\right)} e^{\frac{\pi}{2\epsilon_C}} \\ &= \frac{\pi/\epsilon_C}{1 - e^{-\pi/\epsilon_C}}, \end{aligned} \tag{7.79}$$

which is the result obtained by Sommerfeld and quoted in [5].

How to solve the integrated Boltzmann equation

We want to derive the evolution equation (7.52) for the number density of Dark Matter. Starting from the integrated Boltzmann equation, we get:

$$a^{-3}\frac{\mathrm{d}(n_\chi a^3)}{\mathrm{d}t} = \langle\sigma_{\mathrm{ann}}v\rangle\left\{n_{\chi,\mathrm{eq}}^2 - n_\chi^2\right\}, \tag{7.80}$$

where a is the scale factor, n_χ is the number density of Dark Matter, $n_{\chi,\mathrm{eq}}$ is the Dark Matter equilibrium number density and $\langle\sigma_{\mathrm{ann}}v\rangle$ is the thermally averaged annihilation cross section. We normalize the number density to the total entropy density $s \propto a^{-3}$ by introducing $Y \equiv \frac{n_\chi}{s}$:

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = s\,\langle\sigma_{\mathrm{ann}}v\rangle\left\{Y_{\mathrm{eq}}^2 - Y^2\right\}. \tag{7.81}$$

We want to substitute the time parameter t by the dimensionless evolution parameter $x \equiv \frac{m_\chi}{T_\gamma}$. We need a few formulas to proceed. Using $s \propto a^{-3}$, we find:

$$\frac{\mathrm{d}sa^3}{\mathrm{d}t} = 0 \Rightarrow \dot{s} = -3sH. \tag{7.82}$$

The entropy density s is given by

$$s(T) = kg_{*S}(T)T^3, \tag{7.83}$$

where $k = \frac{2\pi^2}{45}$ is a constant and g_{*S} is the relativistic degrees of freedom for the entropy density. Taking the time derivative yields:

$$\begin{aligned}\dot{s} = \frac{\mathrm{d}s}{\mathrm{d}T}\dot{T} &= \dot{T}\left[k\frac{\mathrm{d}g_{*S}}{\mathrm{d}T}T^3 + 3kg_{*S}T^3\frac{1}{T}\right] \\ &= 3s\frac{\dot{T}}{T}\left[\frac{\mathrm{d}g_{*S}}{\mathrm{d}T}\frac{1}{3g_{*S}} + 1\right] \equiv 3s\frac{\dot{T}}{T}\frac{g_*^{\frac{1}{2}}h_*^{\frac{1}{2}}}{g_{*S}},\end{aligned} \tag{7.84}$$

where the last line defines the parameter $h_*^{\frac{1}{2}}$. Equating equation (7.82) and (7.84), yields

$$\frac{\dot{T}}{T} = -H\frac{g_{*S}}{g_*^{\frac{1}{2}}h_*^{\frac{1}{2}}} = -\sqrt{4\pi^3}45T^2g_{*S}(T)\frac{1}{m_p h_*^{\frac{1}{2}}}, \tag{7.85}$$

where the Friedmann equation in the radiation dominated universe has been used in the last equation. Using the chain rule on the left hand side of (7.51) with this equation gives

$$\begin{aligned}\frac{\mathrm{d}Y}{\mathrm{d}t} = \frac{\mathrm{d}Y}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{\mathrm{d}Y}{\mathrm{d}x}x\left(\frac{\dot{T}}{T}\right) \\ &= \frac{\mathrm{d}Y}{\mathrm{d}x}x\sqrt{\frac{4\pi^3}{45}}\left(\frac{m_\chi}{x}\right)^2 g_{*S}(T)\frac{1}{m_p h_*^{\frac{1}{2}}}.\end{aligned} \tag{7.86}$$

This gives our final form of the evolution equation:

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = \sqrt{\frac{\pi}{45}}\frac{m_p m_\chi}{x^2}h_*^{\frac{1}{2}}\,\langle\sigma_{\mathrm{ann}}v\rangle\left\{Y_{\mathrm{eq}}^2 - Y^2\right\}. \tag{7.87}$$

Calculating the average number density

The number density of Dark Matter in equilibrium, n_{eq}, can be found by integrating the distribution function $f(\mathbf{x}, \mathbf{p})$ over momentum space. We assume a Maxwell-Boltzmann distribution with zero chemical potential:

$$f(\mathbf{x}, \mathbf{p}) = g e^{\frac{-E(p)}{T}}. \tag{7.88}$$

Using $E(p) = \sqrt{p^2 + m^2}$, and introducing $x = m_\chi/T$ and $\alpha = p/T$ we get:

$$\begin{aligned} n_{\text{eq}} &= g \int \frac{\mathrm{d}^3 p}{(2\pi)^3} e^{-\frac{E}{T}} \\ &= \frac{4\pi g T^3}{(2\pi)^3} \int_1^\infty \mathrm{d}\alpha \alpha^2 e^{-\sqrt{\alpha^2 + x^2}} \\ &= \frac{4\pi g T^3}{(2\pi)^3} x^3 \int_1^\infty \mathrm{d}t t\sqrt{t^2 - 1} e^{-xt} \\ &= \frac{4\pi g T^3}{(2\pi)^3} x^2 K_2(x) = \frac{g m_\chi^3}{2\pi^2 x} K_2(x) \Rightarrow \\ Y_{\text{eq}} &= g \frac{45}{4\pi^4} \frac{x^2}{g_{*S}} K_2(x), \end{aligned} \tag{7.89}$$

where K_2 is the modified Bessel function of the second kind.

Solving the differential equation

Equation (7.54) is an example of a stiff differential equation, meaning that it involves vastly different scales. In this case it is roughly the timescale at which annihilations take place versus the timescale at which expansion happens. Explicit solvers like Runge-Kutta either fails or need a very low step size to maintain stability. We implement the implicit scheme, which is also used in the software package DarkSUSY [10, 11]. The derivative y' of a function can be approximated as follows:

$$y_i' \simeq \frac{y_{i+1} - y_i}{h} \qquad \text{Forward difference} \tag{7.90a}$$

$$y_{i+1}' \simeq \frac{y_{i+1} - y_i}{h} \qquad \text{Backward difference.} \tag{7.90b}$$

Equation (7.90b) inserted in the general ODE $y' = f(x, y)$ gives rise to the implicit backward euler scheme:

$$y_{i+1} = y_i + h f_{i+1}(x, y) + \mathcal{O}(h). \tag{7.91}$$

Adding equations (7.90) yields the implicit trapezoidal rule

$$y_{i+1} = y_i + \frac{h}{2}(f_{i+1} + f_i) + \mathcal{O}\left(h^2\right). \tag{7.92}$$

These two schemes are the $s = 0$ and $s = 1$ Adams-Moulton methods respectively. We can use the difference between the two methods as an error-estimate for controlling the step size. An implicit scheme like this usually results in a

system of non-linear algebraic equations which must then be solved numerically in each step. Fortunately, for the Boltzmann-equation, and using method (7.91) or (7.92), this is just a second order equation which can be solved analytically. We introduce the variables suggested in [11]:

$$Y_{eq} \equiv q \tag{7.93a}$$
$$u \equiv h\lambda_{i+1} \tag{7.93b}$$
$$\rho \equiv \frac{\lambda_i}{\lambda_{i+1}} \tag{7.93c}$$
$$c \equiv 2Y_i + u\left[(q_{i+1}^2 + \rho q_i^2) - \rho Y_i^2\right] \tag{7.93d}$$
$$c' \equiv 4(Y_i + uq_{i+1}^2). \tag{7.93e}$$

The Euler method (7.91) for the Boltzmann equation (7.54) is

$$\begin{aligned} 4Y_{i+1} &= 4y_i + 4h\lambda_{i+1}(q_{i+1}^2 - Y_{i+1}^2) \\ &= c' - 4uy_{i+1}^2 \Rightarrow \\ Y_{i+1} &= -\frac{1 \mp \sqrt{1+uc'}}{2u} \Rightarrow \\ Y_{i+1} &= \frac{1}{2}\frac{c'}{1+\sqrt{1+uc'}}, \end{aligned} \tag{7.94}$$

where we have chosen the solution which gives a positive Y_{i+1}. It works the same way for the trapezoidal rule (7.92):

$$\begin{aligned} 2Y_{i+1} &= 2Y_i + h\left[\lambda_i(q_i^2 - Y_i^2) + \lambda_{i+1}(q_{i+1}^2 - Y_{i+1}^2)\right] \\ &= 2Y_i + u\left[\rho q_i^2 + q_{i+1}^2 - \rho Y_i^2\right] - uY_{i+1}^2 \\ &= c - uY_{i+1}^2 \Rightarrow \\ Y_{i+1} &= \frac{-1 \pm \sqrt{1+ucu}}{u} \Rightarrow \\ Y_{i+1} &= \frac{c}{1+\sqrt{1+uc}}. \end{aligned} \tag{7.95}$$

If we let Y'_{i+1} denote the Euler estimate of the next step, we estimate the relative error err as

$$\text{err} = \left|\frac{Y'_{i+1} - Y_{i+1}}{Y_{i+1}}\right|, \tag{7.96}$$

and since the trapezoidal rule is second order in h, we modify the step size according to

$$\text{hnext} = \min\left(hS\sqrt{\frac{\text{eps}}{\text{err}}}, 5h\right), \tag{7.97}$$

where S is a safety factor set at 0.9 and eps is the wanted accuracy. We have also demanded that h can only grow with a factor of 5 in each step.

Chapter 8

CLASS III: Non-cold Relics

Julien Lesgourgues and Thomas Tram

Abstract We present a new flexible, fast and accurate way to implement massive neutrinos, warm dark matter and any other non-cold dark matter relics in Boltzmann codes. For whatever analytical or numerical form of the phase-space distribution function, the optimal sampling in momentum space compatible with a given level of accuracy is automatically found by comparing quadrature methods. The perturbation integration is made even faster by switching to an approximate viscous fluid description inside the Hubble radius, which differs from previous approximations discussed in the literature. When adding one massive neutrino to the minimal cosmological model, `CLASS` becomes just 1.5 times slower, instead of about 5 times in other codes (for fixed accuracy requirements). We illustrate the flexibility of our approach by considering a few examples of standard or non-standard neutrinos, as well as warm dark matter models.

8.1 Introduction

The inclusion of massive, non-cold relics in a Boltzmann code is complicated by the fact that it is necessary to evolve the perturbation of the distribution function on a momentum grid. A grid size of N points together with L terms in the expansion of the perturbation leads to $N \cdot L$ added equations to the system. In public Boltzmann codes like `CMBFAST` [59], `CAMB` [58] and `CMBEASY` [57], distributions are sampled evenly with fixed step size and maximum momentum, adapted to the case of a Fermi-Dirac shaped distribution function $f(p)$. Moreover, the analytic expression for $f(p)$ is hard-coded in many places in those codes, and implicitly assumed e.g. in the mass to density relation, so that exploring other models like neutrinos with chemical potentials and flavour oscillations, neutrinos with non-thermal corrections, extra sterile neutrinos or any kind of warm dark matter candidate requires non-trivial changes to these codes.

We present here the way in which generic Non-Cold Dark Matter (NCDM) relics are implemented in the new Boltzmann code `CLASS`[1] (Cosmic Linear

[1] available at `http://class-code.net`. This paper is based on version `v1.1` of the code.

Anisotropy Solving System), already presented in a series of companion papers [54–56]. In order to ensure a complete flexibility, `CLASS` assumes an arbitrary number of NCDM species, each with an arbitrary distribution function $f_i(p)$. For each species, this function can be passed by the user under some (arbitrarily complicated) analytic form in a unique place in the code, or in a file in the case of non-trivial scenarios that requires a numerical simulation of the freeze-out process. All other steps (finding a mass-density relation, optimising the momentum sampling and computing the derivative of $f_i(p)$) are done automatically in order to ensure maximum flexibility.

In Sec. 8.2, we present an automatic quadrature method comparison scheme which allows `CLASS` to find an optimal momentum sampling, given $f_i(p)$ and some accuracy requirement, and in Sec. 8.3, we devise a new approximation scheme allowing us to drastically reduce the computational time for wavelengths inside the Hubble radius. Finally, in Sec. 8.4 and 8.5, we illustrate these methods with several examples based on standard and non-standard massive neutrinos, and different types of warm dark matter candidates.

8.2 Optimal momentum sampling

The formalism describing the evolution of any NCDM species is given by the massive neutrino equations of Ma & Bertschinger [61]. We will follow the notations from this paper closely, with the exceptions

$$q \equiv \frac{q_{\mathtt{MB}}}{T_{\mathtt{ncdm},0}}\,, \qquad \epsilon \equiv \frac{\epsilon_{\mathtt{MB}}}{T_{\mathtt{ncdm},0}} = \left(q^2 + a^2 \frac{m^2}{T^2_{\mathtt{ncdm},0}}\right)^{\frac{1}{2}}\,, \tag{8.1}$$

where $T_{\mathtt{ncdm},0}$ is the temperature of the non-cold relic today, in the case of a thermal relic. If the relic is non-thermal, $T_{\mathtt{ncdm},0}$ is just a scale of the typical physical momentum of the particles today. Note that the perturbation equations Eq. (8.4) are still the same as in [61], since they depend only on the ratio q/ϵ which is not affected by this rescaling.

Perturbations on a grid

We are not interested in the individual momentum components of the perturbation, Ψ_l, but only in the perturbed energy density, pressure, energy flux and shear stress of each NCDM species, which are integrals over Ψ_l [61]:

$$\delta\rho_{\mathtt{ncdm}} = 4\pi \left(\frac{T_{\mathtt{ncdm},0}}{a}\right)^4 \int_0^\infty f_0(q)\mathrm{d}q q^2 \epsilon \Psi_0, \tag{8.2a}$$

$$\delta p_{\mathtt{ncdm}} = \frac{4\pi}{3} \left(\frac{T_{\mathtt{ncdm},0}}{a}\right)^4 \int_0^\infty f_0(q)\mathrm{d}q \frac{q^4}{\epsilon} \Psi_0, \tag{8.2b}$$

$$(\bar{\rho}_{\mathtt{ncdm}} + \bar{p}_{\mathtt{ncdm}})\,\theta_{\mathtt{ncdm}} = 4\pi k \left(\frac{T_{\mathtt{ncdm},0}}{a}\right)^4 \int_0^\infty f_0(q)\mathrm{d}q q^3 \Psi_1, \tag{8.2c}$$

$$(\bar{\rho}_{\mathtt{ncdm}} + \bar{p}_{\mathtt{ncdm}})\,\sigma_{\mathtt{ncdm}} = \frac{8\pi}{3} \left(\frac{T_{\mathtt{ncdm},0}}{a}\right)^4 \int_0^\infty f_0(q)\mathrm{d}q \frac{q^4}{\epsilon} \Psi_2. \tag{8.2d}$$

In the rest of the article, we will omit all `ncdm` subscripts, and dots will denote derivatives with respect to conformal time, τ.

Note that Ψ_0 and Ψ_1 are gauge-dependent quantities, while higher momenta are not. The gauge transformation can be derived from the corresponding gauge transformation of the integrated quantities. The relation between Ψ_1 in the conformal Newtonian gauge and in the synchronous one reads:

$$\Psi_{1,\mathrm{Con.}} = \Psi_{1,\mathrm{Syn.}} + \alpha k \left[\frac{\epsilon}{q} + \frac{1}{3}\frac{q}{\epsilon}\right] \ , \tag{8.3}$$

with $\alpha \equiv (\dot{h} + 6\dot{\eta})/(2k^2)$, where h and η are the usual scalar metric perturbations in the synchronous gauge. In the rest of this paper, we will work exclusively in the synchronous gauge. The evolution of the Ψ_l's are governed by the Boltzmann equation as described in [61], and leads to the following system of equations:

$$\dot{\Psi}_0 = -\frac{qk}{\epsilon}\Psi_1 + \frac{\dot{h}}{6}\frac{\mathrm{d}\ln f_0}{\mathrm{d}\ln q}, \tag{8.4a}$$

$$\dot{\Psi}_1 = \frac{qk}{3\epsilon}\left(\Psi_0 - 2\Psi_2\right), \tag{8.4b}$$

$$\dot{\Psi}_2 = \frac{qk}{5\epsilon}\left(2\Psi_1 - 3\Psi_3\right) - \left(\frac{\dot{h}}{15} + \frac{2\dot{\eta}}{5}\right)\frac{\mathrm{d}\ln f_0}{\mathrm{d}\ln q}, \tag{8.4c}$$

$$\dot{\Psi}_{l\geq 3} = \frac{qk}{(2l+1)\,\epsilon}\left(l\Psi_{l-1} - (l+1)\,\Psi_{l+1}\right). \tag{8.4d}$$

We can write the homogeneous part of this set of equations as

$$\dot{\mathbf{\Psi}} = \frac{qk}{\epsilon}A\mathbf{\Psi} \equiv \alpha\left(\tau\right)A\mathbf{\Psi}, \tag{8.5}$$

where A is given by

$$A = \begin{bmatrix} & -1 & & & & \\ \frac{1}{3} & & -\frac{2}{3} & & & \\ & \ddots & & \ddots & & \\ & & \frac{l}{2l+1} & & -\frac{l+1}{2l+1} & \\ & & & \ddots & & \ddots \\ & & & & \ddots & \end{bmatrix} \tag{8.6}$$

The solution can be written in terms of the matrix exponential,

$$\mathbf{\Psi}\left(\tau\right) = e^{\int_{\tau_i}^{\tau}\mathrm{d}\tau'\alpha(\tau')A}\mathbf{\Psi}\left(\tau_i\right) \tag{8.7}$$

$$= Ue^{\int_{\tau_i}^{\tau}\mathrm{d}\tau'\alpha(\tau')D}U^{-1}\mathbf{\Psi}\left(\tau_i\right), \tag{8.8}$$

where A has been diagonalised such that $A = UDU^{-1}$ and D is a diagonal matrix of eigenvalues of A. The largest eigenvalue of A (using the complex norm) goes toward $\pm i$ for $l_{\max} \to \infty$, so the largest frequency oscillation in the system is

$$\omega_{\max} \simeq k\int_{\tau_i}^{\tau}\mathrm{d}\tau'\left(1 + \frac{M^2}{q^2}a\left(\tau'\right)^2\right)^{-\frac{1}{2}}. \tag{8.9}$$

Quadrature strategy

There is no coupling between the momentum bins, so our only concern is to perform the indefinite integrals numerically with sufficient accuracy while using the fewest possible points. We are interested in the integrals in Eq. (8.2), which are all on the form

$$\mathcal{I} = \int_0^\infty \mathrm{d}q f_0(q)\, g(q)\,, \tag{8.10}$$

where $f_0(q)$ is the phase space distribution and $g(q)$ is some function of q. We will assume that $g(q)$ is reasonably well described by a polynomial in q, which we checked explicitly for the functions in Eq. (8.2). Under this assumption, we can determine the accuracy of any quadrature rule on $\mathcal{I}$ by performing the integral

$$\mathcal{J} = \int_0^\infty \mathrm{d}q f_0(q)\, t(q)\,, \tag{8.11}$$

where $t(q)$ is a test function. Given a set of different quadrature rules for performing the integral $\mathcal{I}$, the idea is to choose the rule which can compute $\mathcal{J}$ to the required accuracy `tol_ncdm` using the fewest possible points.

We define a quadrature rule on $\mathcal{I}$ to be a set of weights W_i and a set of nodes q_i, such that

$$\mathcal{I} \simeq \sum_{i=1}^{n} W_i g(q_i)\,. \tag{8.12}$$

Note that the distribution function itself has been absorbed into the weights. The optimal quadrature rule will depend on both the distribution $f_0(q)$ and the accuracy requirement `tol_ncdm`, but the specific method used for obtaining the rule is decoupled from the rest of the code; the output is just two lists of n points, $\{q_i\}$ and $\{W_i\}$. `CLASS` tries up to three different methods for obtaining the most optimal quadrature rule, each with its own strength and weaknesses. These are Gauss-Laguerre quadrature, adaptive Gauss-Kronrod quadrature and a combined scheme. We will now discuss each of them.

Gauss-Laguerre quadrature

Most of the time, the distribution function will be close to a Fermi-Dirac distribution, and the integrand is exponentially decaying with q. The Gauss-Laguerre quadrature formula is well suited for exponentially decaying integrands on the interval $(0;\infty)$, so this is an obvious choice. The rule is

$$\int_0^\infty \mathrm{d}q e^{-q} h(q) \simeq \sum_{i=1}^{n} w_i h(q_i)\,, \tag{8.13}$$

where the nodes q_i are the roots of L_n, the Laguerre polynomial of degree n and the weights can be calculated from the formula

$$w_i = \frac{q_i}{(n+1)^2 \left[L_{n+1}(q_i)\right]^2}. \tag{8.14}$$

If we put $h(q) = e^q f_0(q) g(q)$ we obtain the rule

$$W_i = w_i e^{q_i} f_0\left(q_i\right). \tag{8.15}$$

This rule will be very effective when the ratio f_0/e^{-q} is well described by a polynomial, but it will converge very slowly if this is not the case.

Adaptive sampling

When an integrand has structure on scales smaller than the integration interval, an adaptive integration scheme is often the best choice, since it will subdivide the interval until it resolves the structure and reach the required accuracy. We will use the 15 point Gauss-Kronrod quadrature formula as a basis for our adaptive integrator; 7 of the 15 points can be used to obtain a Gauss quadrature estimate of the integral, and the error estimate on the 15 point formula is then $\mathrm{err}_{\mathrm{est.}} = 200|G7 - K15|^{1.5}$.

The Gauss-Kronrod formula is defined on the open interval $(-1, 1)$, but it can be rescaled to work on an arbitrary open interval (a, b). We transform the indefinite integral into a definite integral by the substitution $x = (q+1)^{-1}$:

$$\int_0^\infty \mathrm{d}q f\left(q\right) = -\int_0^1 \mathrm{d}x \frac{\mathrm{d}q}{\mathrm{d}x} f\left(q\left(x\right)\right) = \int_0^1 \mathrm{d}x x^{-2} f\left(q\left(x\right)\right). \tag{8.16}$$

This integral can then be solved by the adaptive integrator. If the tolerance requirement is not met using the first 15 points, the interval is divided in two and the quadrature method is called recursively on each subinterval.

This method is very efficient when the integrand is smooth. For practical purposes, this will be the case unless the phase-space distribution is read from a file with sparse sampling: in this case, the code must interpolate or extrapolate the file values in order to cover the whole momentum range, and the next method may be more efficient.

Integration over tabulated distributions

If some distribution function is not known analytically, but only on a finitely sampled grid on $(q_{\mathrm{min}}, q_{\mathrm{max}})$, we have to interpolate the distribution function within the interval, and we have to extrapolate the behaviour outside the interval. Inside the interval we use a spline interpolation, while we assume $f(q < q_{\mathrm{min}}) \equiv f(q_{\mathrm{min}})$ close to zero. For the tail, we assume the form $f(q) = \alpha e^{-\beta q}$. Requiring the function and its first derivative to be continuous at the point $q = q_{\mathrm{max}}$ leads to the following equations for α and β:

$$\alpha = f(q_{\mathrm{max}}) e^{\beta q_{\mathrm{max}}}, \tag{8.17}$$

$$\beta = -f(q_{\mathrm{max}})^{-1} \left.\frac{\mathrm{d}f}{\mathrm{d}q}\right|_{q=q_{\mathrm{max}}}. \tag{8.18}$$

In the combined scheme we use the 4 point Gauss-Legendre method on the interval $(0, q_{\mathrm{min}})$, adaptive Gauss-Kronrod quadrature on $(q_{\mathrm{min}}, q_{\mathrm{max}})$ and the 6 point Gauss-Laguerre rule on the tail $(q_{\mathrm{max}}, \infty)$[2]. This scheme works well when the integrand is interpolated from tabulated points.

[2]This version of the rule is obtained by a simple substitution.

Implementation in CLASS

When `CLASS` initialises the background structure, it will find optimal momentum samplings for each of the species. More specifically, we start by computing the integral of the distribution function multiplied by the test function at high accuracy, which gives a reference value which can be used for comparison. It also creates a binary tree of refinements, from which we can extract integrals at various levels, where level 1 is the best estimate. We choose the highest possible level which results in an error which is less than the input tolerance, and we extract the nodes and weights from that level.

The code will now search for the lowest number of nodes required for computing the integral with the desired accuracy using Gauss-Laguerre quadrature. The most efficient method, the method using the lowest number of points, is then chosen. For a distribution not departing too much from a Fermi-Dirac one, this will usually be Gauss-Laguerre quadrature.

The scheme suggested here has the benefit, that there is just one tolerance parameter directly related to how well the integral is approximated, *independently* of the distribution function. However, for this to be exactly true, we require the test function to be a sufficiently realistic representation of $q^n\Psi_l$ for $n = 2, 3, 4$ and $l = 0, 1, 2$ for the perturbations. We have checked this using different test functions, but in the end we found the polynomial $f(q) = a_2q^2 + a_3q^3 + a_4q^4$ to be adequate. The coefficients were chosen such that

$$a_n \int_0^\infty \mathrm{d}q \frac{q^n}{e^q + 1} = 1. \tag{8.19}$$

When the phase-space distribution function is passed in the form of a file with tabulated (q_j, f_j) values, the code compares the three previous methods (still with a common tolerance parameter) and keeps the best one, which is usually the third one in the case of a poor sampling of the function, or one of the other two in the opposite case.

Note that higher accuracy is needed for integrating background quantities (density, pressure, etc.) than perturbed quantities (the Ψ_l's). On the other hand, the code spends a negligible time in the computation of the former, while reducing the number of sampling points for perturbations is crucial for reducing the total computing time. Hence, `CLASS` calls the quadrature optimisation algorithm twice for each NCDM species, with two different accuracy parameters. The background tolerance is set to a smaller value leading to a finer sampling.

We conclude this section by noticing that this whole process sounds very sophisticated, but requires a negligible computing time in `CLASS`. What really matters is to reduce the number of discrete momenta in the perturbation equations, and this is indeed accomplished thanks to the previous steps (as we shall see in Sec. 8.5).

8.3 Sub-Hubble Approximation

Fluid approximation

Various kinds of approximations for massive neutrino perturbations have been discussed in the past [47, 60, 62]. The approximation discussed here is different and consists in an extension of the Ultra-relativistic Fluid Approximation presented in [55], applying only to the regime in which a given mode has entered the Hubble radius. The idea is that after Hubble crossing, there is an effective decoupling between high multipoles (for which power transfers from smaller l's to higher l's, according to the free-streaming limit) and low multipoles (just sourced by metric perturbation). Hence, when $k\tau$ exceeds some threshold, we can reduce the maximum number of multipoles from some high $l_{\rm max}$ down to $l_{\rm max} = 2$. We showed in [55] that this Ultra-relativistic Fluid Approximation (UFA) allows simultaneously to save computing time (by reducing the number of equations) and to increase precision (by avoiding artificial reflection of power at some large cut-off value $l_{\rm max}$).

In the case of massive neutrinos, we expect the same arguments to hold in the relativistic regime, while in the non-relativistic limit all multipoles with $l > 1$ decay and the species behave more and more like a pressureless fluid. Hence, some kind of fluid approximation is expected to give good results in all cases.

We write the continuity equation and the Euler equation in the usual way. In the synchronous gauge we have

$$\dot{\delta} = -(1+w)\left(\theta + \frac{\dot{h}}{2}\right) - 3\frac{\dot{a}}{a}\left(c^2_{\rm Syn.} - w\right)\delta, \tag{8.20a}$$

$$\dot{\theta} = -\frac{\dot{a}}{a}\left(1 - 3c_g^2\right)\theta + \frac{c^2_{\rm Syn.}}{1+w}k^2\delta - k^2\sigma. \tag{8.20b}$$

Here, c_g^2 is the adiabatic sound speed, and $c^2_{\rm Syn.} \equiv \frac{\delta p}{\delta \rho}$ is the effective sound speed squared in the synchronous gauge. The latter can be related to the physical sound speed defined in the gauge comoving with the fluid, that we denote $c_{\rm eff}$. The above equations can then be written as:

$$\dot{\delta} = -(1+w)\left(\theta + \frac{\dot{h}}{2}\right) - 3\frac{\dot{a}}{a}\left(c^2_{\rm eff} - w\right)\delta + 9\left(\frac{\dot{a}}{a}\right)^2(1+w)\left(c^2_{\rm eff} - c_g^2\right)\frac{\theta}{k^2}, \tag{8.21a}$$

$$\dot{\theta} = -\frac{\dot{a}}{a}\left(1 - 3c^2_{\rm eff}\right)\theta + \frac{c^2_{\rm eff}}{1+w}k^2\delta - k^2\sigma. \tag{8.21b}$$

Later on, we will close the system by an evolution equation for the shear σ, but first we will discuss how to calculate the adiabatic sound speed and how to approximate the effective sound speed $c^2_{\rm eff}$.

Sound speeds

The adiabatic sound speed can be expressed as

$$c_g^2 = \frac{\dot{p}}{\dot{\rho}} = w\frac{\dot{p}}{p}\left(\frac{\dot{\rho}}{\rho}\right)^{-1} = -w\frac{\dot{p}}{p}\left(\frac{\dot{a}}{a}\right)^{-1}\frac{1}{3(1+w)}$$

$$= \frac{w}{3\,(1+w)}\left(5 - \frac{\mathfrak{p}}{p}\right), \tag{8.22}$$

where the quantity $\mathfrak{p}$ (called the pseudo-pressure inside `CLASS`) is a higher moment pressure defined by

$$\mathfrak{p} \equiv \frac{4\pi}{3} a^{-4} \int_0^\infty f_0(q) \mathrm{d}q \frac{q^6}{\epsilon^3}. \tag{8.23}$$

With this formulation, we can compute the adiabatic sound speed in a stable and accurate way, without needing to evaluate the time-derivative of the background pressure $\dot{p}$. When the `ncdm` species is no longer relativistic, its pressure perturbation δp defined in Eq. (8.2b) is an independent quantity. Since we do not have an evolution equation for δp, we approximate $c^2_{\rm eff}$ by c^2_g. This approximation is sometimes as much as a factor 2 wrong as shown on Fig. 8.1.

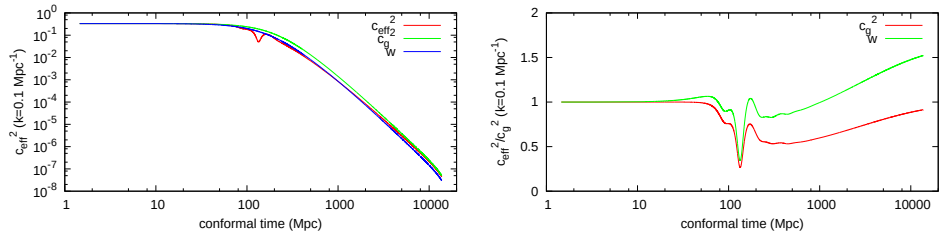


Figure 8.1: Effective sound speed squared $c^2_{\rm eff}$ for a mass of $m = 2.0$ eV. *Left panel:* The effective sound speed plotted together with the adiabatic sound speed squared c^2_g and the equation of state parameter w. In the relativistic and the non-relativistic limit we have $c^2_{\rm eff} = c^2_g$ as expected, but the behaviour of $c^2_{\rm eff}$ in between the two limits are non-trivial. *Right panel:* The ratios $c^2_{\rm eff}/c^2_g$ and $c^2_{\rm eff}/w$. One can see that c^2_g is a better approximation to $c^2_{\rm eff}$ than w, but neither catches the full evolution.

Evolution equation for the shear

Given an ansatz for Ψ_3, we can derive a formally correct evolution equation for the shear. We follow Ma and Bertschinger, and close the system using their suggested recurrence relation for massive neutrinos[3]. The truncation law presented in Ma and Bertschinger is valid for $l_{\rm max} > 3$: in this case, all quantities are gauge-invariant. When writing the same ansatz for $l_{\rm max} = 3$, we have to face the issue of the gauge dependence of Ψ_1. Asssuming that the truncation law holds for gauge-invariant quantities, one obtains in the synchronous gauge:

$$\Psi_3 \approx \frac{5\epsilon}{qk\tau}\Psi_2 - \left(\Psi_1 + \alpha k \left[\frac{\epsilon}{q} + \frac{1}{3}\frac{q}{\epsilon}\right]\right). \tag{8.24}$$

[3]The recurrence relation in the massless limit is better motivated theoretically, since $\Psi_l \propto j_l(k\tau)$ when metric perturbations vanish or satisfy a simple constraint (namely, $\dot{\phi}+\dot{\psi}=0$ in the Newtonian gauge). In the massive case, the formal solution involves more complicated oscillating functions with arguments going from $\sim k\tau$ in the massless limit to $\sim (k\tau)^{-1}$ in the massive limit, as can be checked from eq. (8.9).

Throughout this subsection and the next one, one can recover Newtonian gauge equations by simply taking $\alpha = 0$. We now differentiate equation (8.2d):

$$\dot{\sigma} + \frac{\dot{a}}{a}\left(1 - 3c_g^2\right)\sigma = \frac{1}{\rho + p}\frac{8\pi}{3}a^{-4}\int_0^\infty f_0(q)\mathrm{d}qq^4\frac{\partial}{\partial\tau}\left(\frac{\Psi_2}{\epsilon}\right) \ . \tag{8.25}$$

We can compute the right-hand side using Eq. (8.4c) and replace Ψ_3 with its approximate expression from (8.24). After carrying out integrals over momentum, one gets:

$$\dot{\sigma} = -3\left(\tau^{-1} + \frac{\dot{a}}{a}\left[\frac{2}{3} - c_g^2 - \frac{1}{3}\frac{\Sigma}{\sigma}\right]\right)\sigma + \frac{2}{3}\left[\Theta + \alpha k^2\frac{w}{1+w}\left(3 + \frac{\mathfrak{p}}{p}\right)\right], \tag{8.26}$$

where we have borrowed the notation

$$(\rho + p)\,\Theta = 4\pi k a^{-4}\int_0^\infty f_0(q)\mathrm{d}qq^3\frac{q^2}{\epsilon^2}\Psi_1, \tag{8.27}$$

$$(\rho + p)\,\Sigma = \frac{8\pi}{3}a^{-4}\int_0^\infty f_0(q)\mathrm{d}q\frac{q^4}{\epsilon}\frac{q^2}{\epsilon^2}\Psi_2, \tag{8.28}$$

from [62]. From the definition it is clear that $\Theta \to \theta$ and $\Sigma \to \sigma$ in the relativistic limit, and that Θ and Σ become suppressed in the non-relativistic regime compared to θ and σ. Our differential equation for σ differs from its Newtonian gauge counterpart in [62], because we have used the recurrence relation to truncate the hierarchy, while Shoji and Komatsu have used $\Psi_3 = 0$. The evolution equation for the shear can be further simplified by using Eq. (8.22), leading to:

$$\dot{\sigma} = -3\left(\tau^{-1} + \frac{\dot{a}}{a}\left[\frac{2}{3} - c_g^2 - \frac{1}{3}\frac{\Sigma}{\sigma}\right]\right)\sigma + \frac{2}{3}\left[\Theta + \alpha k^2\left(8\frac{w}{1+w} - 3c_g^2\right)\right] \ . \tag{8.29}$$

Estimating higher order momenta

One way to close the system governing the fluid approximation is to replace Θ and Σ by the usual quantities θ and σ multiplied by functions depending only on background quantities (in the same way that we already approximated δp by $c_g^2\delta\rho$). More explicitly, our aim is to write an approximation of the type $\Sigma = 3w_\sigma\sigma$, where w_σ could be any function of time going from one third in the relativistic limit to zero in the non-relativistic one. Since θ and Θ are not gauge-independent, we should search for a similar approximation holding on their gauge-independent counterpart. In the synchronous gauge, such an approximation would read

$$\left[\Theta + \alpha k^2\left(8\frac{w}{1+w} - 3c_g^2\right)\right] = 3w_\theta\left[\theta + \alpha k^2\right] \ . \tag{8.30}$$

However, we will stick to the notations of [60], who introduced a viscosity speed related to our w_θ through

$$c_{\mathrm{vis}}^2 = \frac{3}{4}w_\theta(1+w) \ . \tag{8.31}$$

With such assumptions, the approximate equation for the shear would read

$$\dot{\sigma} = -3\left(\frac{1}{\tau} + \frac{\dot{a}}{a}\left[\frac{2}{3} - c_g^2 - w_\sigma\right]\right)\sigma + \frac{4}{3}\frac{c_{\mathrm{vis}}^2}{1+w}\left[2\theta + 2\alpha k^2\right] \ . \tag{8.32}$$

Since the suppression factor q^2/ϵ^2 which appears in Eq. (8.27, 8.28) compared to Eq. (8.2c, 8.2d) is also found in the pressure integral compared to the energy density integral, we may guess that the relative behaviour is similar, i.e. related by w. This leads to a guess $w_\sigma = w$ and $w_\theta = w$ which implies $c^2_{\rm vis} = \frac{3}{4}w\,(1+w)$. However, the same logic would imply $c^2_{\rm eff} = w$, which we have shown in Fig. 8.1 is not exactly true.

Let us investigate a bit how to approximate higher momenta quantities like Θ and Σ. If we want to approximate Θ for instance, we may assume some functional form of $\Psi_1(q)$ described by a single (time dependent) parameter. We can make the ansatz $\Psi_1(q/\epsilon) \approx a_{1n}(t)\left(\frac{q}{\epsilon}\right)^n$, and then use θ to determine the parameter $a_{1n}(t)$. We then find

$$\Theta \approx \theta \frac{\int_0^\infty f_0(q)\mathrm{d}qq^3\frac{q^2}{\epsilon^2}\left(\frac{q}{\epsilon}\right)^n}{\int_0^\infty f_0(q)\mathrm{d}qq^3\left(\frac{q}{\epsilon}\right)^n}. \tag{8.33}$$

The guess $c^2_{\rm vis} = \frac{3}{4}w\,(1+w)$ can be seen to be a special case of this approach having $n = -1$. The problem is that the value of n best approximating the behavior of Ψ_1 and other momenta is not the same in the relativistic and non-relativistic limit. In fact our testing shows that this guess sources σ too much during the relativistic to non-relativistic transition compared to the exact solution. Instead we got much better results by using $c^2_{\rm vis} = 3wc_g^2$, which avoids this excessive sourcing during the transition, while still reducing to 1/3 in the relativistic limit.

For the ratio Σ/σ, the assumption of a q-independent Ψ_2 (i.e. $n = 0$) yields $w_\sigma = \mathfrak{p}/(3p)$, which provides satisfactory results and is adopted in the schemes described below.

We speculate that by pushing these kinds of considerations further, one could find better approximations for $c^2_{\rm eff}$, $c^2_{\rm vis}$ and w_σ. It is also possible that another independent equation could be found, and that it would allow a better determination of $c_{\rm eff}$.

Implementation in CLASS

For comparison, we have implemented 3 different Non-Cold Dark Matter Fluid Approximations (NCDMFA) in `CLASS` which differ only in their respective equation for the shear. In correspondence with the Ultra-relativistic Fluid Approximation discussed in [55], we have named the approximations `MB`, `Hu` and `CLASS`: in the relativistic limit, they reduce to their relativistic counterpart in [55]. In all three approximations we are using Eq. (8.21a) and (8.21b) as the first two equations with $c_{\rm eff} = c_g$. The respective equations for the shear read

$$\dot\sigma_{\texttt{MB}} = -3\left(\frac{1}{\tau} + \frac{\dot a}{a}\left[\frac{2}{3} - c_g^2 - \frac{1}{3}\frac{\mathfrak{p}}{p}\right]\right)\sigma + \frac{4}{3}\frac{c^2_{\rm vis}}{1+w}\left[2\theta + \dot h + 6\dot\eta\right], \quad c^2_{\rm vis} = 3wc_g^2, \tag{8.34a}$$

$$\dot\sigma_{\texttt{Hu}} = -3\frac{\dot a}{a}\frac{c_g^2}{w}\sigma + \frac{4}{3}\frac{c^2_{\rm vis}}{1+w}\left[2\theta + \dot h + 6\dot\eta\right], \qquad c^2_{\rm vis} = w, \tag{8.34b}$$

$$\dot{\sigma}_{\texttt{CLASS}} = -3\left(\frac{1}{\tau} + \frac{\dot{a}}{a}\left[\frac{2}{3} - c_g^2 - \frac{1}{3}\frac{\mathfrak{p}}{p}\right]\right)\sigma + \frac{4}{3}\frac{c_{\rm vis}^2}{1+w}\left[2\theta + \dot{h}\right], \qquad c_{\rm vis}^2 = 3wc_g^2. \tag{8.34c}$$

The second shear equation, named `Hu`, corresponds exactly to the prescription of Ref. [60] for approximating massive neutrinos. The first shear equation, `MB`, comes directly from (8.32) with the values of w_σ and $c_{\rm vis}$ motivated in the previous subsection. Finally, in [55], we found that removing the $\dot{\eta}$ term leads to slightly better results for the matter power spectrum, and can be justified using an analytic approximation to the exact equations. By analogy, we also define in the massive neutrino case a `CLASS` approximation identical to the `MB` one except for the omission of this term.

In Fig. 8.2 we have tested these three fluid approximations in a model with no massless neutrinos and 3 degenerate massive neutrinos. The three approximations work very well as long as the neutrinos are light and become non-relativistic after photon decoupling. Like in the massless case, the `CLASS` approximation is slightly better for predicting the matter power spectrum on small scales, and we set it to be the default method in the code. When the mass increases, the fluid approximation alters the CMB spectra on small angular scales ($l \geq 2500$), but the error remains tiny (only 0.02% for $l = 2750$ for three species with $m = 1$eV). The effect on the matter power spectrum is stronger: with three 1 eV neutrinos, the $P(k)$ is wrong by 1 to 3% for $k \in [0.05; 1]h\text{Mpc}^{-1}$. Hence, we recommend to use the fluid approximation for any value of the mass when computing CMB anisotropies, and only below a total mass of one or two eV's when computing the matter power spectrum. However, cosmological bounds on neutrino masses strongly disfavour larger values of the total mass. This means that in most projects, `CLASS` users can safely use the fluid approximation for fitting both CMB and large scale structure data.

8.4 Standard massive neutrinos

We first illustrate our approach with the simple case of standard massive neutrinos with a Fermi-Dirac distribution. In this case, for each neutrino, the user should provide two numbers in the input file: the mass m, and the relative temperature `T_ncdm` $\equiv T_\nu/T_\gamma$ (the ratio of neutrino to the photon temperature). The `CLASS` input file `explanatory.ini` recommends to use the value `T_ncdm`=0.71599, which is "fudged" in order to provide a mass-to-density ratio $m/\omega_\nu = 93.14$ eV in the non-relativistic limit. This number gives a very good approximation to the actual relic density of active neutrinos, resulting from an accurate study of neutrino decoupling [53]. However, when comparing the `CLASS` results with those from `CAMB`, we take `T_ncdm`=0.7133 in order to recover the mass-to-density ratio assumed in that code. Finally, if no temperature is entered, the code will default to the instantaneous decoupling value of $(4/11)^{1/3}$.

Agreement with CAMB

In Fig. 8.3, we compare the CMB and matter power spectrum from `CAMB` and `CLASS` (without the NCDM fluid approximation) for two massless and one

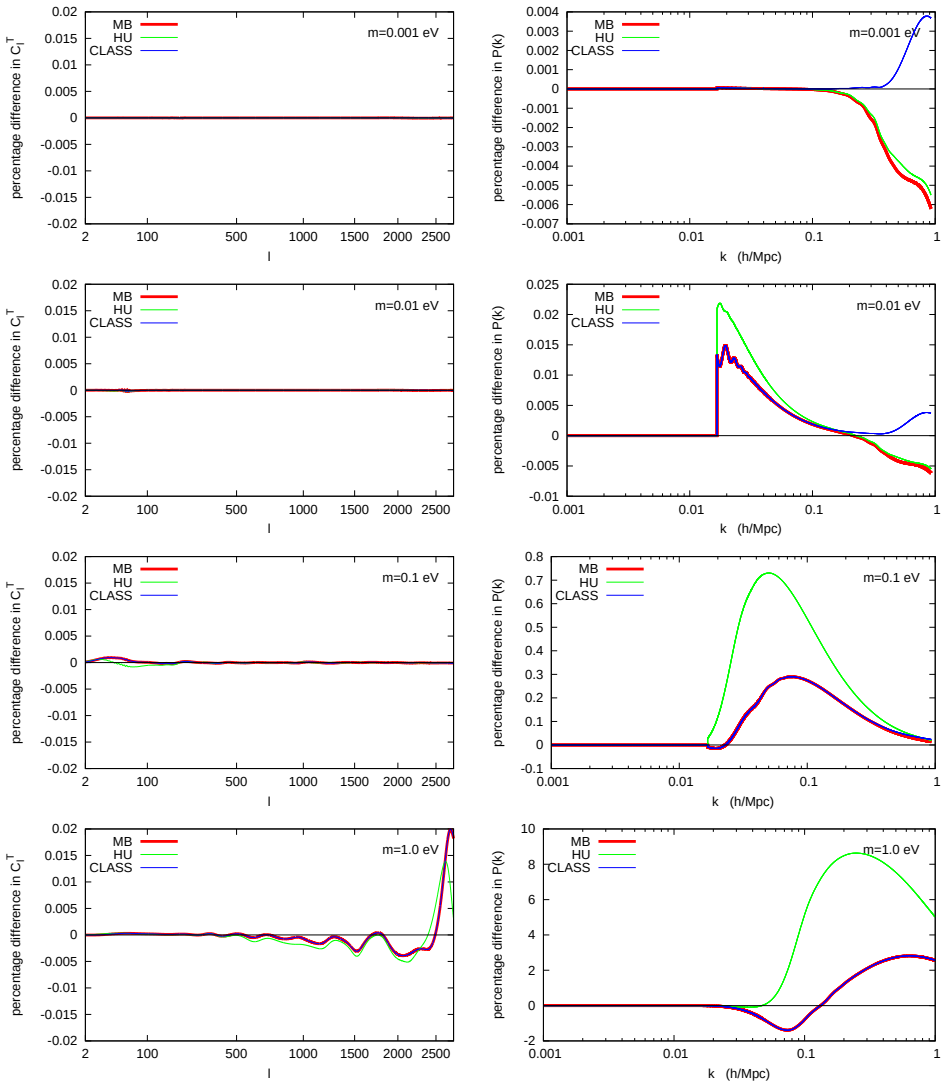


Figure 8.2: On the left, we have shown the percentage difference in the C_l^T for three degenerate neutrino species with mass $m = 0.001$eV, $m = 0.01$eV, $m = 0.1$eV and $m = 1$eV respectively, in runs with/without the fluid approximation. The fluid approximation works very well as long as the neutrinos are relativistic, so this is what we expect. On the right we have shown the matter power spectrum for the same masses. Here the agreement is not so good as the mass becomes higher.

massive neutrino with $\Omega_\nu = 0.02$ (corresponding to a mass $m \simeq 0.923$eV). We used high accuracy settings for `CAMB`, described in [56] under the name *[CAMB:07]*. For `CLASS`, we used the input file `cl_ref.pre`, which corresponds to the setting *[CLASS:01]* in [56] for parameters not related to NCDM; for the latter, `cl_ref.pre` contains the settings described in the first column of Table 8.1. For such settings and in absence of massive neutrinos, the two temperature spectra would agree at the 0.01% level in the range $l \in [20; 3000]$; at the 0.02% level for polarization in the same range; and at the 0.01% level

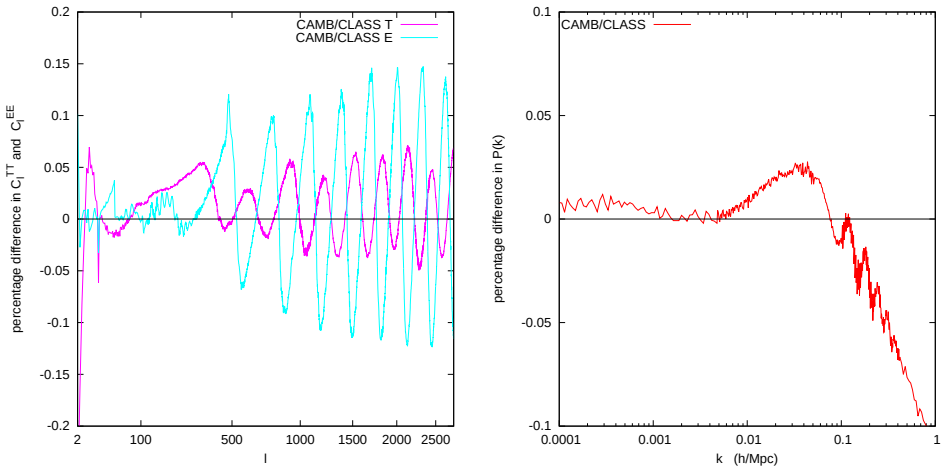


Figure 8.3: Relative difference between `CAMB` and `CLASS` spectra in a model with $\Omega_\nu = 0.02$, two massless neutrinos, and reference accuracy settings. The two codes agree rather well.

for the matter power spectrum for $k < 1h\text{Mpc}^{-1}$. With a neutrino mass close to 1 eV, we see in Fig. 8.3 that the discrepancy is approximately six times larger than in the massless case. However, it remains very small: even with massive neutrinos the two codes agree to better than 0.1% for the CMB and matter power spectra. This is by far sufficient for practical applications.

In a perfect implementation of massless and massive neutrinos in Boltzmann codes, we expect that in the relativistic limit $m \ll T_\nu^0$ (where T_ν^0 is the neutrino temperature today) the spectra would tend towards those obtained with three massless species (provided that we are careful enough to keep the same number of relativistic degrees of freedom $N_{\rm eff}$). We performed this exercise for both codes, and the results are presented in Fig. 8.4. It appears that with a small enough mass, `CLASS` can get arbitrarily close to the fully relativistic case: with a mass of 10^{-8}eV, the difference is at most of 0.03% in the C_l's and 0.05% in the $P(k)$. This test is another way to validate the accuracy of our implementation.

Accuracy settings

We now come to the question of defining degraded accuracy settings for computing the spectra in a fast way, while keeping the accuracy of the results under control. For such an exercise, we need to define a measure a precision. Like in [56], we will use an effective χ^2 which mimics the sensitivity of a CMB experiment like Planck to temperature and E-polarisation anisotropies. Taking the runs with accuracy settings `cl_ref.pre` as a reference, we decrease the precision for each parameter while keeping the $\Delta\chi^2$ roughly below a given limit, chosen to be either 0.1 or 1. This exercise was already performed in [56] for all parameters not related to NCDM, leading to the definition of two precision files `chi2pl0.1.pre` and `chi2pl1.pre` which are available on the `CLASS` web site. Here, we only need to set the NCDM precision parameters in these

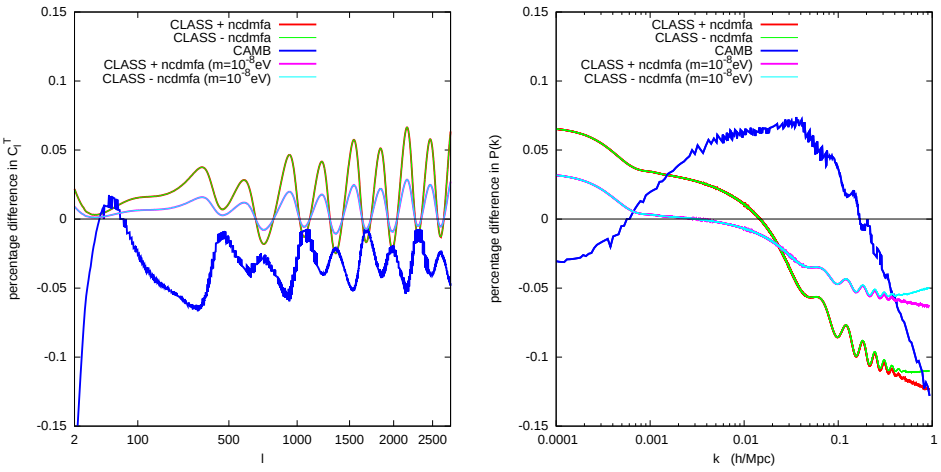


Figure 8.4: This is a test of how well `CAMB` and `CLASS` recovers the massless limit. We compute a model with $\Omega_\nu = 1.2 \cdot 10^{-4}$ and 3 massive neutrinos with degenerate mass. This setting corresponds to a neutrino mass of $m_i = 2.8 \cdot 10^{-4}$eV, which is not exactly massless, but it is the best we can do since the mass parameter can not be set directly in `CAMB`. Setting the mass parameter in `CLASS` to $m_i = 10^{-8}$eV reveals that we are in part seeing the effect of the neutrino going slightly non-relativistic at late times.

two files to correct values. Our results are listed in Table 8.1, in the second and third columns. They take advantage of the fluid approximation, and use an extremely small number of momenta (8 or 5 only). We checked that these settings provide the correct order of magnitude for $\Delta\chi^2$ within a wide range of neutrino masses, at least up to 2 eV. This is shown in Table 8.2 for the two cases `chi2pl0.1.pre` and `chi2pl1.pre`, as well as for the case `chi2pl1.pre` with the fluid approximation removed. Around $m = 2$ eV, the error induced by the fluid approximation starts increasing significantly: when exploring this region, the user should either turn off the approximation, or increase the value of the $k\tau$ trigger. Given current limits on active neutrino masses, the interesting mass range to explore is below 2 eV, and in most projects, the `CLASS` users can safely employ the default settings of `chi2pl0.1.pre` and `chi2pl1.pre` including the fluid approximation.

These settings are optimised for fitting the CMB spectra only. For the matter power spectra, the files `chi2pl0.1.pre` and `chi2pl1.pre` produce an error of the order of a few per cents in the range $k \in [0.05; 1]h\text{Mpc}^{-1}$ (for any neutrino mass and with/without the fluid approximation). In order to get accurate matter power spectra, it is better to employ the settings `cl_permille.pre`, `cl_2permille.pre`, `cl_3permille.pre`, which lead to a precision of 1, 2 or 3 per mille for C_l^{TT} in the range $2 < l < 3000$, even in the presence of neutrino masses. In these files, we fixed the fluid approximation trigger to a rather larger value in order to get a precision of one permille for the matter power spectrum for $k < 0.2h\text{Mpc}^{-1}$ and $m < 2$ eV, or a bit worse for mildly non-linear scales $k \in [0.2; 1]h\text{Mpc}^{-1}$. The power spectrum accuracy with such settings is indicated in Table 8.3 for various values of the mass.

	`cl_ref.pre`	`chi2pl0.1.pre`	`chi2pl1.pre`
`tol_ncdm_bg`	10^{-10}	10^{-5}	10^{-5}
`tol_ncdm`	10^{-10}	10^{-4}	10^{-3}
`l_max_ncdm`	51	16	12
fluid approximation	none	`ncdmfa_class`	`ncdmfa_class`
$k\tau$ trigger	–	30	16
number of q (back.)	28	11	11
number of q (pert.)	28	8	5
number of neutrino equations	1428	136→3	65→3

Table 8.1: Accuracy parameters related to NCDM in the three precision files `cl_ref.pre`, `chi2pl0.1.pre` and `chi2pl1.pre`. When the fluid approximation is used, the method described in section 8.3 is employed, and the switching time is set by the above values of $k\tau$. Below these parameters, we indicate the corresponding number of momenta sampled in background quantities and in perturbation quantities, as well as the number of neutrino perturbation equations integrated over time, equal to (`l_max_ncdm` + 1) times the number of sampled momenta when the fluid approximation is not used, and to three afterwards.

mass (eV)	`chi2pl0.1.pre`	`chi2pl1.pre`	same without approx.
10^{-3}	0.087	0.94	0.90
10^{-2}	0.087	0.93	0.92
0.1	0.092	0.90	0.92
1	0.083	0.96	0.82
2	0.157	1.10	0.93

Table 8.2: For a CMB instrument with the sensitivity of Planck, χ^2 difference between the spectra obtained with reference accuracy settings and with degraded accuracy settings, for various values of the neutrino mass (all models have two massless and one massive neutrinos). This shows that our accuracy settings `chi2pl0.1.pre` and `chi2pl1.pre` always lead to an accuracy of roughly $\Delta\chi^2 \sim 0.1$ or $\Delta\chi^2 \sim 1$ respectively. The last column correspond to the settings of `chi2pl1.pre`, but without the fluid approximation.

mass (eV)	$k < 0.2h\mathrm{Mpc}^{-1}$	$k \in [0.2; 1]h\mathrm{Mpc}^{-1}$
10^{-3}	0.04%	0.12%
10^{-2}	0.04%	0.12%
0.1	0.05%	0.12%
1	0.06%	0.8%
2	0.2%	1.5%

Table 8.3: Maximum error induced by any of the `cl_permille.pre`, `cl_2permille.pre` or `cl_3permille.pre` precision settings on the linear matter power spectrum $P(k)$, for approximately linear scales $k < 0.2h\mathrm{Mpc}^{-1}$ (first column) or mildly non-linear scales $k \in [0.2; 1]h\mathrm{Mpc}^{-1}$ (second column), and for various values of the neutrino mass (all models have two massless and one massive neutrinos). The fluid approximation introduces an error which remains below the per mille level until $k = 0.2h\mathrm{Mpc}^{-1}$ for $m < 2$ eV, and exceeds this level for larger masses.

Performance

The quadrature method reveals to be extremely useful since even with five values of the momenta, we get accurate results leading to 0.2%-0.3% accuracy on the $C_l's$, 0.1% accuracy on the $P(k)$ and $\Delta\chi^2 \sim 1$. Traditional Boltzmann codes employ 14 momenta in order to achieve a comparable precision. In the presence of massive neutrinos, the total execution time of a Boltzmann code is dominated by the integration of the perturbation equations, which depends on the total number of perturbed variables, itself dominated by the number of massive neutrino equations. By reducing the number of momenta from 14 to 5, the quadrature method speeds up the code by more than a factor two. We find that the use of the fluid approximation leads to an additional 25% speed up for standard accuracy settings (like those in the file `chi2pl1.pre`). In total, for a single massive neutrino, our method speeds up the code by a factor 3. This means that instead of being 4.5 times slower in presence of one massive neutrino, `CLASS` only becomes 1.5 times slower. We checked these numbers with various masses and accuracy settings.

Realistic mass schemes

We have proved in this section that `CLASS` can be employed in any project requiring high-precision computations of cosmological observables in presence of massive neutrinos. It is of course perfectly suited for realistic situations with different neutrino species and masses.

To illustrate this, we display in Figure 8.5 the ratio of pairs of matter power spectra for models with three massive neutrinos satisfying constraints from atmospheric/solar oscillation experiments [46] ($\Delta m^2_{21} = 7.6 \times 10^{-5}\text{eV}^2$, $\Delta m^2_{32} = \pm 2.4 \times 10^{-3}\text{eV}^2$). Each pair of models corresponds to one normal hierarchy and one inverted hierarchy scenario, with the same total mass M_ν, equal to 0.100 eV, 0.115 eV or 0.130 eV. The first total mass is very close to the minimum allowed value for the inverted hierarchy, $M_\nu \simeq 0.0994$ eV. For each pair or models with a given M_ν:

- on intermediate scales, the bump reflects the difference in the three free-streaming scales involved in the two models.

- in the large k limit, the two spectra are offset by 0.03% to 0.22%: it is known that in this limit, the suppression in the power spectrum induced by neutrino free-streaming depends mainly on the total mass (through the famous $-8f_\nu$ approximate formula), but also slightly on the mass splitting (in [52], a more accurate formula gives the suppression as a function of both the total mass and number of degenerate massive neutrinos). When M_ν increases, the two models are less different from each other (they go towards a common limit, namely the degenerate mass scenario), and the discrepancy is less pronounced.

- in the small k limit, the two spectra are nearly identical. The tiny difference, which increases when M_ν decreases, is due to the fact that in the inverted hierarchy model, there is a very light neutrino just finishing to complete its non-relativistic transition today. It therefore has a non-

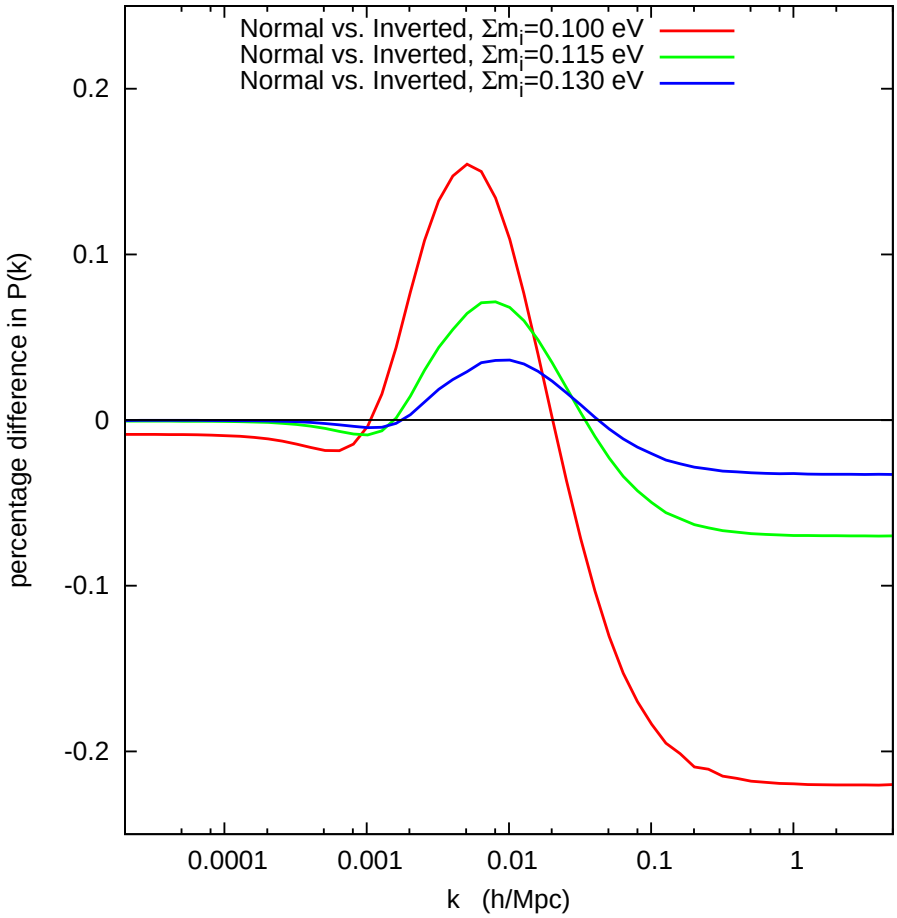


Figure 8.5: Ratio of matter power spectra for pairs of models with three massive neutrinos, obeying either to the normal or inverted hierarchy scenario, but with a common total mass for each pair: $M_\nu = 0.100$ eV, 0.115 eV or 0.130 eV. The various effects observed here are discussed in the text.

negligible pressure, which slightly affects metric perturbations on large wavelengths.

Observing the difference between these two models would be extremely challenging, although 21 cm surveys could reach enough sensitivity [51].

8.5 Beyond standard massive neutrinos

In this section we will illustrate the power and flexibility of the non-cold Dark Matter implementation in `CLASS`, by implementing different models which have already been studied elsewhere in the literature.

Massive neutrinos with large non-thermal corrections

It is plausible that some new physics can introduce non-thermal corrections to an otherwise thermal Fermi-Dirac distribution function. One might think of

using CMB and large scale structure data to put bounds on such non-thermal corrections, as was described e.g. in [63]. `CLASS` is ideally suited for playing with such models. As a test case, we take the following distribution from [63]:

$$f(q) = \frac{2}{(2\pi)^3}\left[\frac{1}{e^q+1} + \frac{A\pi^2}{q^2\sqrt{2\pi}\sigma}\exp\left(-\frac{(q-q_c)^2}{2\sigma^2}\right)\right], \tag{8.35}$$

which is the Fermi-Dirac distribution with an added Gaussian peak in the number density. This distribution could presumably be the result of some particle suddenly decaying into neutrinos at a late time.

In practise, we only need to change the expression for $f(q)$ in `CLASS`, which appears in a unique line (in the function `background_ncdm_distribution()`). All the rest, like density-to-mass relation and computation of the logarithmic derivative, is done automatically by the code. In particular, we do not need to change the accuracy parameters `tol_ncdm` and `tol_ncdm_bg`: the momentum sampling algorithm automatically increases the number of momenta by a significant amount, in order to keep the same precision. If this was not the case, the effect of the peak would be underestimated because of under sampling, and the parameter extraction would then likely be biased.

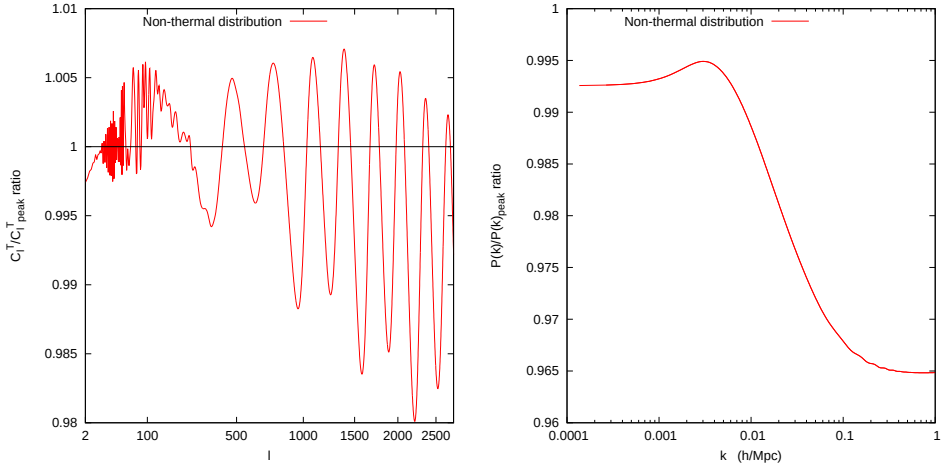


Figure 8.6: C_l's and $P(k)$'s for a model of 3 degenerate neutrinos with the non-thermal distribution (8.35) using parameters $m = 1.0$ eV, $A = 0.018$, $\sigma = 1.0$ and $q_c = 10.5$. This corresponds to $N_{\rm eff} = 3.98486$. We have compared this model to a model with degenerate thermal neutrinos with the same mass and $N_{\rm eff}$. The signal is due to a combination of background and perturbation effects: although the mass and the relativistic density are the same, the non-relativistic density and the average momentum differ significantly in the two models.

In Fig. 8.6, we show the CMB and matter power spectra for this model, relative to a standard model with three thermally distributed neutrinos. The two models are chosen to share exactly the same masses and the same initial number of relativistic degrees of freedom $N_{\rm eff}$. Nevertheless, they do not have the same non-relativistic neutrino density and average neutrino momentum; in particular, non-thermal neutrinos in the decay peak become non-relativistic

slightly later. This induces a combination of background and perturbation effects affecting CMB and matter power spectra in a significant way.

Warm dark matter with thermal-like distribution

There is an infinity of possible warm dark matter models, since the phase-space distribution of warm dark matter depend on the details of its production mechanism. The most widely studied model is that of non-resonantly produced warm dark matter with a rescaled Fermi-Dirac distribution, having the same temperature as that of active neutrinos. This model is implemented in the default `CLASS` version: when the user enters a temperature, a mass and a density $\Omega_{\texttt{ncdm}}$ (or $\omega_{\texttt{ncdm}}$) for the same species, the code knows that the degeneracy parameter in front of the Fermi-Dirac distribution must be rescaled in order to match these three constraints simultaneously. The code will also ensure that the perturbations begin to be integrated when the non-cold species is still relativistic, in order to properly follow the transition to the non-relativistic regime.

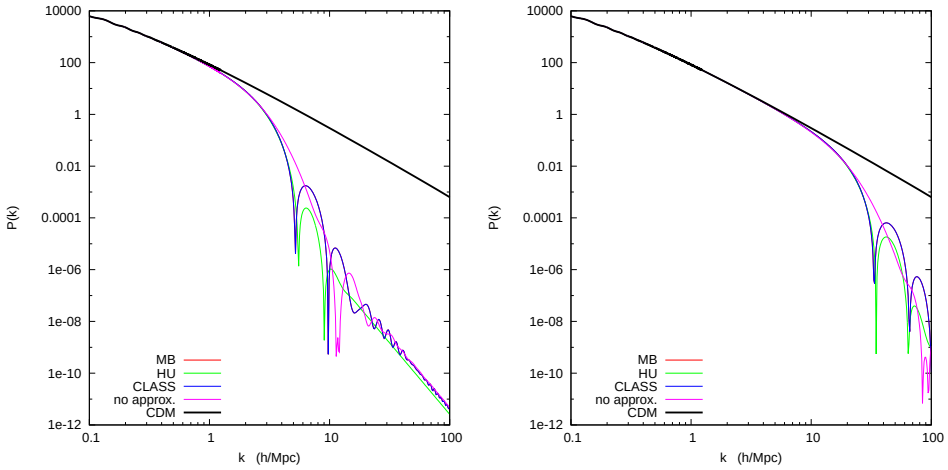


Figure 8.7: $P(k)$'s for a Warm Dark Matter model with $m = 1$ keV (left) and $m = 10$ keV (right). The fluid approximation can be seen to be a very good approximation in this case, though it does not catch the acoustic oscillations precisely.

We illustrate this by running a ΛWDM model with a mass of $m = 1$keV or $m = 10$keV and a density $\Omega_{\texttt{ncdm}} = 0.25$, with or without the fluid approximation. We compare the results with those of ΛCDM with $\Omega_{\texttt{cdm}} = 0.25$, in order to show the well-known suppression effect of WDM in the small-scale limit of the matter power spectrum. It appears that the fluid approximation works very well in those cases, unless one wants to resolve the details of the WDM acoustic oscillations on very small scales, first predicted in [48].

Warm dark matter with non-trivial production mechanism

Non-resonantly produced warm dark matter candidates are severely constrained by Lyman-α bounds, but such bounds do not apply to other warm particles

which could have been produced through more complicated mechanisms (e.g. resonant production), leading to a non-trivial, model-dependent phase-space distribution function [50]. It is not always easy to find a good analytic approximation for such a distribution; this is anyway not an issue for `CLASS`, since the code can read tabulated values of $f(p)$ from an input file.

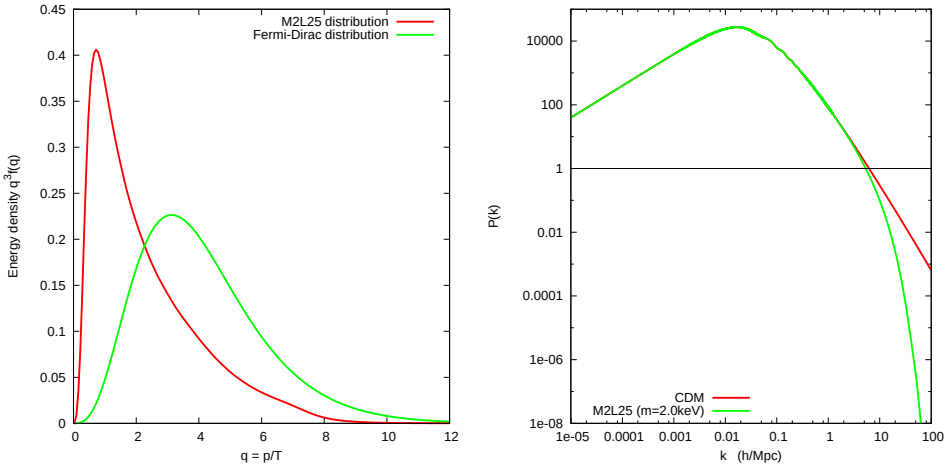


Figure 8.8: $P(k)$'s for a Warm Dark Matter model with a non-trivial production mechanism for a mass of $m = 2$ keV compared to the same model with Cold Dark Matter. Note that the normalisation of the distribution function is arbitrary; when both `m_ncdm` and one of {`Omega_ncdm`,`omega_ncdm`} is present for some species, `CLASS` will normalise the distribution consistently.

We illustrate this case by taking a particular model for resonantly produced sterile neutrinos, which distribution was computed numerically by [49] (simulating the details of sterile neutrino production and freeze-out), and stored in a file with discrete q_i, f_i values. Again, we only need to specify the name of this file in the `CLASS` input file, to enter a value for the mass and for the density $\Omega_{\texttt{ncdm}}$, and the rest is done automatically by the code (finding the mass-density relation and the correct normalization factor for $f(q)$, defining the new momentum steps, deriving $[d\ln f]/[d\ln q]$ with a good enough accuracy). The assumed $f(q)$ and the resulting matter power spectrum when the mass is set to $m = 2$keV is shown in Fig. 8.8. By eye, this spectrum seems identical to a thermal-like WDM one, but the cut-off is in fact much smoother due to an excess of low-momentum particles in this model (which behave like a small cold dark matter fraction).

8.6 Conclusions

A large fraction of the activity in cosmology consists in deriving bounds on particle physics in general, and on the neutrino and dark matter sector in particular. Fitting cosmological data with non-standard neutrinos or other non-cold relics require non-trivial changes in existing public Boltzmann codes. Moreover, running parameter extraction codes including massive neutrinos or more exotic non-cold relics is computationally expensive due to a significant

increase in the number of differential equations to be solved numerically for each set of cosmological parameters.

The newly released Cosmic Linear Anisotropy Solving System aims at rendering this task easy and fast. The code provides a very friendly and flexible input file in which users can specify a lot of non-standard properties for the NCDM sector: masses, temperatures, chemical potentials, degeneracy parameters, etc. Moreover, the Fermi-Dirac distribution function is not hard-coded in `CLASS`; it is just a default choice appearing in one line of the code, which can be very easily modified. Even when a non-thermal distribution $f(q)$ does not have a simple analytic expression, the code can be told to read it directly from a file. After reading this function, `CLASS` performs a series of steps in a fully automatic way: finding the mass-density relation, defining an optimal sampling in momentum space with a sophisticated but fast algorithm, and accurately computing the derivative of $f(q)$, needed in the perturbation equations.

In this paper, we presented the main two improvements related to the NCDM sector in `CLASS`: an adaptive quadrature sampling algorithm, which is useful both for the purpose of flexibility (the sampling is always adapted to any new distribution function) and speed (the code sticks to a minimum number of momenta, and hence, of perturbation equations); and a fluid approximation switched on inside the Hubble radius. We showed that the latter approximation works very well for realistic active neutrinos (with a total mass smaller than $1-2$eV), and for warm dark matter candidates becoming non-relativistic during radiation domination. In between these two limits, there is a range in which the accuracy of the fluid approximation is not well established, and in which the user may need to keep the approximation off, at the expense of increasing the execution time. However, the range between a few eV and few keV is usually not relevant in most realistic scenarios.

The adaptive quadrature sampling algorithm and the fluid approximation both contribute to a reduction in the total execution time of the code by a factor of three for ordinary neutrinos. This means that when one massive neutrino species is added to the ΛCDM model, `CLASS` becomes 1.5 times slower instead of 4.5 times slower like other codes. Since the code is already quite fast in the the massless case, we conclude that the global speed up is significant and appreciable when fitting cosmological data.

Acknowledgments

We wish to thank Oleg Ruchayskiy for providing the phase-space distribution of the Warm Dark Matter model, Fig. 8.8, which gave us an opportunity to test the code in a non-standard scenario. We also wish to thank Emanuele Castorina for many useful discussions on the fluid approximation for neutrinos during his stay at CERN.

Chapter 9

Thermalisation of light sterile neutrinos in the early universe

Steen Hannestad, Irene Tamborra and Thomas Tram

Abstract Recent cosmological data favour additional relativistic degrees of freedom beyond the three active neutrinos and photons, often referred to as "dark" radiation. Light sterile neutrinos is one of the prime candidates for such additional radiation. However, constraints on sterile neutrinos based on the current cosmological data have been derived using simplified assumptions about thermalisation of ν_s at the Big Bang Nucleosynthesis (BBN) epoch. These assumptions are not necessarily justified and here we solve the full quantum kinetic equations in the (1 active + 1 sterile) scenario and derive the number of thermalised species just before BBN begins ($T \simeq 1$ MeV) for null ($L = 0$) and large ($L = 10^{-2}$) initial lepton asymmetry and for a range of possible mass-mixing parameters. We find that the full thermalisation assumption during the BBN epoch is justified for initial small lepton asymmetry only. Partial or null thermalisation occurs when the initial lepton asymmetry is large.

9.1 Introduction

Sterile neutrinos are hypothetical $SU(2) \times U(1)$ singlets. They are supposed to mix with one or more of the active states without interacting with any other particle. Low-mass sterile neutrinos have been invoked to explain the excess $\bar{\nu}_e$ events in the LSND experiment [75–77] as well as the MiniBooNE excess events in both neutrino and antineutrino channels. Interpreted in terms of flavour oscillations, the MiniBooNE data require CP violation and thus no less than two sterile families [78–80] or additional ingredients such as non-standard interactions [82]. Recently a new analysis of reactor $\bar{\nu}_e$ spectra and their distance and energy variation [90–92] suggested indication for the possible existence of eV-mass sterile neutrinos. However the IceCube collaboration excluded part of the parameter space [81].

The most recent analysis of cosmological data suggest a trend towards the existence of "dark radiation," radiation in excess with respect to the three neutrino families and photons [83, 98, 99]. The cosmic radiation content is usually expressed in terms of the effective number of thermally excited neutrino species, $N_{\rm eff}$. Its standard value, $N_{\rm eff} = 3.046$, slightly exceeds 3 because of e^+e^- annihilation providing residual neutrino heating [53]. The Wilkinson Microwave Anisotropy Probe (WMAP) collaboration found $N_{\rm eff} = 4.34^{+0.86}_{-0.88}$ based on their 7-year data release and additional LSS data [100] at 1σ. Including the Sloan Digital Sky Survey (SDSS) data release 7 (DR7) halo power spectrum, [98] found $N_{\rm eff} = 4.78^{+1.86}_{-1.79}$ at 2σ. Measurements of the CMB anisotropy on smaller scales by the ACT [114] and SPT [113] collaborations also find tentative evidence for a value of $N_{\rm eff}$ higher than predicted by the standard model (see also [109–112, 115] for recent discussions of $N_{\rm eff}$).

Also, cosmological constraints coming from big bang nucleosynthesis (BBN) suggest that the relatively high ^{4}He abundance can be interpreted in terms of additional radiation during the BBN epoch [87, 88]. Low-mass sterile neutrinos have been considered among possible candidates for the extra-radiation content [84, 85, 89]. The cosmic microwave background anisotropies and big-bang nucleosynthesis in combination seem to favor an excess of radiation compatible with one family of sub-eV sterile neutrinos [83–86]. On the other hand, eV-mass sterile neutrinos are cosmologically viable only if additional ingredients are included since otherwise sterile neutrinos would contribute too much hot dark matter [89] (see also [107]).

However, cosmological constraints during the BBN epoch have usually been derived under the assumption that the extra sterile neutrino families were fully thermalised [89]. However, the validity of this assumption is not a priori clear and some preliminary studies [96, 97] already pointed toward this direction. It was shown in [95] that for plausible values of the mass and mixing parameters, and initial lepton asymmetries not excluded by current observations there are cases where little or no thermalisation occurs. For the charged fermions of the standard model the particle anti-particle asymmetry is known to be of order 10^{-10}. For neutrinos, however, no such bound exists, and the asymmetry can be many orders of magnitude larger without violating observational constraints. In the standard model with no sterile states the upper bound on the neutrino chemical potential is of order $\mu/T \lesssim \text{few} \times 10^{-2}$ [108, 117–121] and while no exact bound has been derived in models with sterile neutrinos, we expect that the upper bound is of the same order of magnitude.

The purpose of this paper is to quantitatively derive the amount of thermalisation as a function of neutrino parameters (mass, mixing, and initial lepton asymmetry). We solve the full quantum kinetic equations in the 1 active+1 sterile approximation, calculate the effective number of thermalised species just before BBN starts (at $T \simeq 1$ MeV) and define under which conditions the thermalisation hypothesis holds. The assumption of (1+1) families to evaluate the thermalisation degree is justified for small lepton asymmetries since the resonances in the active sector are decoupled from the conversions occurring in the active-sterile sector due to the larger mass difference. However, for large asymmetries active-sterile conversion is delayed and can occur simultaneously with active-active conversion. While this does not qualitatively change the overall picture there are some issues which we will return to in Section 3.

In our study, we calculate the number of thermalised extra families for the allowed mass-mixing parameter space for different initial lepton asymmetries. In Section 2 we introduce the adopted formalism and the quantum kinetic equations. In Section 3, we present our results for initial null and large ($L = 10^{-2}$) lepton asymmetry. Conclusions and perspectives are presented in Section 4.

9.2 Equations of motion

In this section, we introduce the quantum kinetic equations (QKEs) governing the evolution of neutrinos in the early universe [64–69]. We adopt a mapping of the Bloch vectors in terms of new vectors related to the active and sterile species grouping large and small dynamical variables.

Quantum Kinetic Equations

We consider oscillations of one active flavour ν_a (with $a = e$ or μ, τ) with a sterile neutrino state ν_s. Denoting with θ_s the mixing angle in vacuum and with ν_1 and ν_2 the two mass eigenstates, separated by the mass difference δm_s^2, we have:

$$\nu_a = \cos\theta_s \nu_1 - \sin\theta_s \nu_2 \ , \tag{9.1}$$

$$\nu_s = \sin\theta_s \nu_1 + \cos\theta_s \nu_2 \ . \tag{9.2}$$

In what follows we will refer $\delta m_s^2 > 0$ as the normal hierarchy scenario (NH) and $\delta m_s^2 < 0$ as the inverted hierarchy scenario (IH). Structure formation data strongly disfavour models with a total thermalised neutrino mass (the sum of all fully thermalised mass states) in excess of 0.5-1 eV. Given that all the active states are fully thermalised this disfavours the inverted hierarchy for sterile masses above 0.2-0.3 eV. However, for masses below this the inverted hierarchy is not disfavoured and for completeness we study the same mass and mixing parameter space for both NH and IH.

In order to describe the evolution of sterile neutrinos in the early universe, we use the density matrix formalism and we express the density matrix associated with each momentum p in terms of the Bloch vector components $(P_0, \mathbf{P}) = (P_0, P_x, P_y, P_z)$ [64, 65, 67],

$$\rho = \frac{1}{2} f_0 (P_0 + \boldsymbol{P} \cdot \boldsymbol{\sigma}) \ , \qquad \overline{\rho} = \frac{1}{2} f_0 (\overline{P}_0 + \overline{\boldsymbol{P}} \cdot \boldsymbol{\sigma}) \ , \tag{9.3}$$

where σ are the Pauli matrices and $f_0 = 1/(1+e^{p/T})$ is the Fermi-Dirac distribution function with no chemical potential. The neutrino kinetic equations in terms of the components of the Bloch vectors for each momentum mode are:

$$\dot{\mathbf{P}} = \mathbf{V} \times \mathbf{P} - D(P_x \mathbf{x} + P_y \mathbf{y}) + \dot{P}_0 \mathbf{z} \ , \tag{9.4}$$

$$\dot{P}_0 = \Gamma \left[\frac{f_{\rm eq}}{f_0} - \frac{1}{2}(P_0 + P_z) \right] \tag{9.5}$$

where the dot denotes the time derivative ($\mathrm{d}_t = \partial_t - Hp\partial_p$, with H the Hubble parameter) and $f_{\rm eq} = 1/(1 + e^{(p-\mu)/T})$.

Defining the comoving momentum $x = p/T$, the vector $\mathbf{V}$ has the following components

$$V_x = \frac{\delta m_s^2}{2xT} \sin 2\theta_s \,, \tag{9.6}$$
$$V_y = 0 \,, \tag{9.7}$$
$$V_z = V_0 + V_1 + V_L. \tag{9.8}$$

and

$$V_0 = -\frac{\delta m_s^2}{2xT} \cos 2\theta_s, \tag{9.9}$$
$$V_1^{(a)} = -\frac{7\pi^2}{45\sqrt{2}} \frac{G_F}{M_Z^2} x T^5 \left[n_{\nu_a} + n_{\bar{\nu}_a} \right] g_a \tag{9.10}$$
$$V_L = \frac{2\sqrt{2}\zeta(3)}{\pi^2} G_F T^3 L^{(a)}. \tag{9.11}$$

Here, $g_{\mu,\tau} = 1$ for $\nu_{\mu,\tau}$–ν_s mixing, $g_e = 1 + 4\sec^2\theta_W/(n_{\nu_e} + n_{\bar{\nu}_e})$ for ν_e–ν_s mixing and θ_W is the Weinberg angle. The dimensionless number densities $n_{\nu_a,(\bar{\nu}_a)}$ are the equilibrium active neutrino (antineutrino) densities normalised to unity in thermal equilibrium. The effective neutrino asymmetries $L^{(a)}$ are defined by

$$L^{(e)} = \left(\frac{1}{2} + 2\sin^2\theta_W\right) L_e + \left(\frac{1}{2} - 2\sin^2\theta_W\right) L_p - \frac{1}{2} L_n + 2L_{\nu_e} + L_{\nu_\mu} + L_{\nu_\tau}, \tag{9.12}$$
$$L^{(\mu)} = L^{(e)} - L_e - L_{\nu_e} + L_{\nu_\mu} \,, \tag{9.13}$$
$$L^{(\tau)} = L^{(e)} - L_e - L_{\nu_e} + L_{\nu_\tau} \,, \tag{9.14}$$

where $L_f \equiv (n_f - n_{\bar{f}}) N_f / N_\gamma$ with N_f (N_γ) the integrated active (photon) number density in thermal equilibrium. The potential V_L, defined as in Eq. (9.11), is the leading order contribution to V_z. The V_1 term is the finite temperature correction and for example in the case of ν_e–ν_s mixing it includes coherent interactions of ν_e with the medium through which it propagates. The condition for a matter induced resonance to occur is $V_z = 0$, and because V_z depends on $L^{(a)}$ any non-zero lepton asymmetry can have dramatic consequences for oscillation driven active-sterile neutrino conversion. In Appendix A we discuss the location of resonances in detail for all possible values of mass, mixing, and lepton asymmetry.

A detailed derivation of the quantum kinetic equations is presented in [66, 73]. Here we choose to adopt minimal assumptions on the collision terms. In particular, the term D is the damping term, quantifying the loss of quantum coherence due to ν_a collisions with the background medium. For example, considering ν_e, the elastic contribution should come from the elastic scattering of ν_e with e^- and e^+ and with the other active flavours ν_a and $\bar{\nu}_a$. The inelastic contribution comes from the scattering of ν_e with $\bar{\nu}_e$ (producing e^- and e^+ or ν_a and $\bar{\nu}_a$). In terms of the Bloch vectors such terms have the effect of suppressing the off-diagonal elements of the density matrix ($P_{x,y}$). The effective potentials contributing to this term have been previously calculated [72–74] and if thermal equilibrium is aasumed and the electron mass neglected, it is

approximately half the corresponding scattering rate Γ [65, 66, 70]

$$D = \frac{1}{2}\Gamma \ . \tag{9.15}$$

The evolution of P_0 is determined by processes that deplete or enhance the abundance of ν_a with the same momentum and its rate of change receives no contribution from coherent ν_a-ν_s oscillations. The repopulation term $\Gamma(f_{\rm eq}/f_0 - 1/2(P_0 + P_z))$ is an approximation for the correct elastic collision integral [70] with

$$\Gamma = C_a G_F^2 x T^5 \tag{9.16}$$

where $C_e \simeq 1.27$ and $C_{\mu,\tau} \simeq 0.92$ [64]. Note that the term including the effective collision rate, Γ, is an approximation to the full momentum dependent scattering kernel which repopulates neutrinos from the background plasma. The full expression has been derived in [66]. In [70] it was proven that the general form of D (and Γ) exactly reduces to Eqs. (9.15,9.16) for weakly interacting species in thermal equilibrium with zero chemical potential, and that it is the zero order approximation for particles with non-null chemical potential. The respective equations of motion for anti-neutrinos can be found by substituting $L^{(a)} = -L^{(a)}$ and $\mu = -\mu$ in the above equations. In our treatment we have not included the rate equations for the electrons and positrons since we are assuming that all the species electromagnetically interacting are kept in equilibrium.

Mapping with the active and the sterile variables

We can distinguish among large and small linear combinations of the dynamical variables in the particle and antiparticle sector to simplify the numerical treatment. For each momentum mode, we define for each component i (with $i = 0, x, y, z$) of the Bloch vector

$$P_i^{\pm} = P_i \pm \overline{P}_i \ . \tag{9.17}$$

We also separate active (a) and sterile (s) sectors

$$P_a^{\pm} = P_0^{\pm} + P_z^{\pm} = 2\frac{\rho_{aa}^{\pm}}{f_0} \ , \tag{9.18}$$

$$P_s^{\pm} = P_0^{\pm} - P_z^{\pm} = 2\frac{\rho_{ss}^{\pm}}{f_0} \ . \tag{9.19}$$

Therefore, in terms of the new vectors Eqs. (9.4, 9.5) become

$$\dot{P}_a^{\pm} = V_x P_y^{\pm} + \Gamma\left[2f_{eq}^{\pm}/f_0 - P_a^{\pm}\right] \ , \tag{9.20}$$

$$\dot{P}_s^{\pm} = -V_x P_y^{\pm} \ , \tag{9.21}$$

$$\dot{P}_x^{\pm} = -(V_0 + V_1)P_y^{\pm} - V_L P_y^{\mp} - D P_x^{\pm} \ , \tag{9.22}$$

$$\dot{P}_y^{\pm} = (V_0 + V_1)P_x^{\pm} + V_L P_x^{\mp} - \frac{1}{2}V_x(P_a^{\pm} - P_s^{\pm}) - D P_y^{\pm} \ , \tag{9.23}$$

where we have defined $f_{\rm eq}^{\pm} = f_{\rm eq}(p,\mu) \pm f_{\rm eq}(p,-\mu)$.

The lepton number can be directly calculated from the integral over the difference between the neutrino and the antineutrino distribution functions, i.e. P_a^-:

$$L^{(a)} = \frac{2}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 \rho_{aa}^- = \frac{1}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 f_0 P_a^- \ . \tag{9.24}$$

However, since the repopulation term is approximated by Eq. (9.16) which does not explicitly conserve lepton number we independently evolve $L^{(a)}$ as in [71] using an evolution equation where the repopulation term does not enter. Taking the time derivative of Eq. (9.24) and ignoring the repopulation part of $\dot{P}_a^-$, the evolution equation for $L^{(a)}$ is

$$\dot{L}^{(a)} = \frac{1}{8\zeta(3)} \int_0^\infty \mathrm{d}x x^2 f_0 V_x P_y^- \ . \tag{9.25}$$

Note that, in kinetic equilibrium, μ, or rather the degeneracy parameter $\xi \equiv \mu/T$, is related to the lepton number $L^{(a)}$ through the integral over f_{eq}^- [121]

$$L_{\mathrm{eq}}^{(a)} = \frac{1}{4\zeta(3)} \int_0^\infty \mathrm{d}x \ x^2 \left[\frac{1}{1+e^{x-\xi}} - \frac{1}{1+e^{x+\xi}} \right] = \frac{1}{12\zeta(3)} \left(\pi^2 \xi + \xi^3 \right) . \tag{9.26}$$

This is a third order equation, and using Chebyshev's cubic root, one can extract the corresponding and expression for ξ valid for any $L^{(a)}$ using trigonometric functions:

$$\xi = \frac{-2\pi}{\sqrt{3}} \sinh \left(\frac{1}{3} \mathrm{arcsinh} \left[-\frac{18\sqrt{3}\zeta(3)}{\pi^3} L^{(a)} \right] \right) . \tag{9.27}$$

In order to numerically solve the QKEs, we define the momentum grid in comoving coordinates ($x = p/T$). Therefore the grid becomes stationary and the partial differential equations become ordinary differential equations coupled through integrated quantities only. Using the temperature T as the evolution parameter, time derivatives, d_t, are replaced by $\to -HT\partial_T$ in the above equations, provided that the time derivative of the effective number of degrees of freedom can be ignored.

9.3 Results: thermalised sterile species

The fraction of sterile thermalised species is defined as

$$\delta N_{\mathrm{eff,s}} = \frac{\int \mathrm{d}x x^3 f_0 P_s^+}{4 \int \mathrm{d}x x^3 f_0} \ . \tag{9.28}$$

However, the total amount of radiation is given by the sum of active and sterile energy densities

$$\delta N_{\mathrm{eff}} = \frac{\int \mathrm{d}x x^3 f_0 \left(P_s^+ + P_a^+ - 4 \right)}{4 \int \mathrm{d}x x^3 f_0} \ . \tag{9.29}$$

Note that when the active state is in thermal equilibrium ($P_a^+ = 4$), $\delta N_{\mathrm{eff,s}} = \delta N_{\mathrm{eff}}$. When $L^{(a)}$ is large, the sterile sector may be populated so late that

the active sector does not have time to repopulate before it decouples. In this section, we discuss the fraction of thermalised species for initial $L^{(a)} = 0$ and $L^{(a)} = 10^{-2}$ and for a range of $(\delta m_s^2, \sin^2 2\theta_s)$.

In terms of late-time cosmological constraints on light neutrinos both $\delta N_{\rm eff,s}$ and $\delta N_{\rm eff}$ can be relevant quantities. Models with a modified light neutrino sector are most often parametrised in terms of the neutrino mass, m_ν, and $N_{\rm eff}$ in such a way that $N_{\rm eff}$ neutrino species all share the same common mass m_ν (i.e. it is assumed that the mass spectrum is degenerate). However, in models with a single sterile state one instead has either $\delta N_{\rm eff,s}$ steriles with mass m_s and $3.046 + \delta N_{\rm eff} - \delta N_{\rm eff,s}$ massless active states (NH) or $3.046 + \delta N_{\rm eff} - \delta N_{\rm eff,s}$ massive active state with degenerate mass and $\delta N_{\rm eff,s}$ massless sterile states (IH). These two cases are different when it comes to structure formation and should in principle be treated separately (see e.g. [116] for a discussion about this point). Since the goal of this paper is to calculate $\delta N_{\rm eff}$, not to provide quantitative constraints on specific models, we simply use $\delta N_{\rm eff}$ from this point on.

Numerical solution of the quantum kinetic equations

Solving the quantum kinetic equations numerically is non-trivial task. The number of differential equations are roughly $8N$ where N is the number of momentum bins, and since the resonances can be very narrow we need a few hundred points to obtain good precision. There are many vastly separated time-scales involved, so the problem is stiff, and once L changes, the system becomes extremely non-linear. We used two different solvers, one based on the numerical differentiation formulae of order $1-5$ (`ndf15`) due to Shampine [122], and one based on the fifth order implicit Runge-Kutta method `RADAU5` due to Hairer and Wanner [93]. If the maximum order of the first method is reduced to two, both solvers are L-stable, and thus excellent for stiff problems. Because of the large number of equations and the sparsity of the Jacobian we must use sparse matrix methods for the linear algebra operations needed in both solvers. For this purpose, we are employing a small sparse matrix package based on [123].

To sample the momentum-space in an optimal way we are mapping the x-interval $[x_{\rm min}; x_{\rm max}]$ to a u-interval $[0; 1]$ by

$$u(x) = \frac{x - x_{\rm min}}{x_{\rm max} - x_{\rm min}} \times \frac{x_{\rm max} + x_{\rm ext}}{x + x_{\rm ext}}, \tag{9.30}$$

where $x_{\rm ext}$ is the extremal point of some moment of the Fermi-Dirac distribution. We chose the values $x_{\rm min} = 10^{-4}$, $x_{\rm ext} = 3.1$ and $x_{\rm max} = 100$, and then sampled u uniformly. This is the same mapping employed in [94], but they go one step further and introduce an adaptive grid that follows the resonances. This is not necessary for this project since our mixing angles are comparably larger, and we are not looking at chaotic amplification of an initially small value of L.

We evolved the system from an initial temperature of 60 MeV to a final

temperature of 1 MeV for the following grid of masses and mixing angles:

$$10^{-3}\ \text{eV}^2 \leq \delta m_s^2 \leq 10\ \text{eV}^2 \text{ and } 10^{-4} \leq \sin^2 2\theta_s \leq 10^{-1} \text{ for } \quad L^{(a)} = 0\ , \tag{9.31a}$$

$$10^{-1}\ \text{eV}^2 \leq \delta m_s^2 \leq 10\ \text{eV}^2 \text{ and } 10^{-3.3} \leq \sin^2 2\theta_s \leq 10^{-1} \text{ for } \quad L^{(a)} = 10^{-2}\ . \tag{9.31b}$$

We ran the complete grids for different number of momentum bins, different accuracy parameters and both differential equation solvers with no noticeable difference.

Sterile neutrino production for zero lepton asymmetry

The simplest case, and the one most often studied in the literature, is the one where the lepton asymmetry is zero. For $L^{(a)} = 0$, the evolution of P_i^+ is decoupled from P_i^- [see Eqs. (9.20,9.23)] and the asymmetry remains zero for the whole evolution (as can be seen from Eq. (9.25)).

From Eqs. (9.35,9.39) in Appendix A it can be seen that there is either no resonance (NH) or that the resonances are identical for neutrinos and anti-neutrinos (IH). As it is well known, in IH the resonance propagates to higher values of x as the universe expands and eventually covers the entire momentum distribution of neutrinos.

In Fig. 9.1 we show the fraction of thermalised neutrinos, δN_{eff}, for the range of mixing parameters given in Eq. (9.31a) with initial asymmetry $L^{(\mu)} = 0$. The top panel shows the normal hierarchy, $\delta m_s^2 > 0$, and the bottom panel the inverted hierarchy, $\delta m_s^2 < 0$. The smaller parameter space described by (9.31b) is denoted with a dashed rectangle to facilitate comparison with the results presented in Sec 9.3.

We mark with a green hexagon the best fit point of the $3+1$ global analysis presented in [106], obtained from a joint analysis of Solar, reactor, and short-baseline neutrino oscillation data $(\delta m_s^2, \sin^2 2\theta_s) = (0.9\text{eV}^2, 0.089)$. For that point $\delta N_{\text{eff}} = 1$ in both hierarchies, i.e. complete thermalization occurs. In addition we show the parameter range preferred by CMB and large scale structure (LSS) data. The $1-2-3\sigma$ contours have been obtained interpolating the likelihood function obtained in [84] for each fixed δm_s^2 and N_{eff}. In both cases the lower left corners of parameter space where little thermalization occurs are disfavoured because of the CMB+LSS preference for extra energy density.

It is also of interest to see how the thermalization proceeds as a function of temperature. In Fig. 9.2 we show the evolution of δN_{eff} as a function of temperature for the NH scenario for a variety of different δm_s^2 and $\sin^2 2\theta_s$. For the non-resonant NH, the thermalization rate of sterile neutrinos is approximately $\Gamma_s \sim \frac{1}{2} \sin^2 2\theta_s \Gamma$. The maximum thermalisation rate occurs at a temperature of approximately $T_{\text{max}} \sim 10\,(\delta m_s^2)^{1/6}$ MeV and the final δN_{eff} depends only on $\sin^2 2\theta_m$ at that temperature (see [64] for a detailed discussion). In the top panel of Fig. 9.2 this behaviour can be seen. For very large vacuum mixing $\Gamma_s/H > 1$ already before T_{max} such that complete thermalisation has occurred already before T_{max} reached. For smaller mixing Γ_s/H never exceeds 1 and even though thermalisation proceeds fastest around T_{max} it is never fast enough to equilibrate the sterile states.

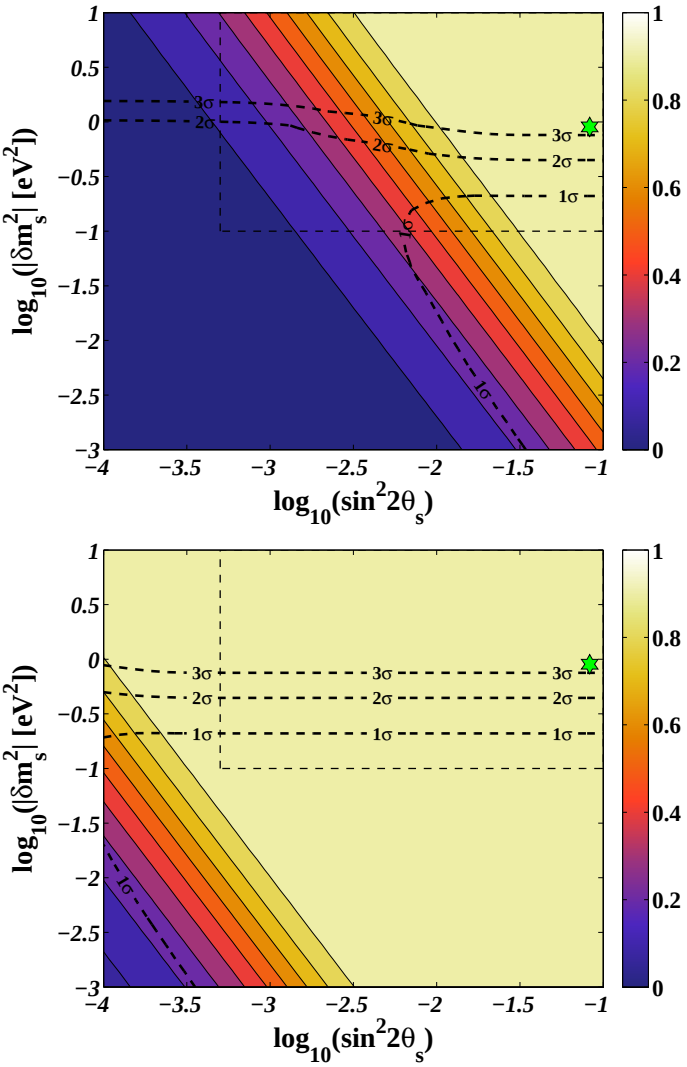


Figure 9.1: Iso-$\delta N_{\rm eff}$ contours in the $\sin^2 2\theta_s - \delta m_s^2$ plane for $L^{(\mu)} = 0$ and $\delta m_s^2 > 0$ (top panel) and $\delta m_s^2 < 0$ (bottom panel). The green hexagon denotes the ν_s best-fit mixing parameters as in the $3+1$ global fit in [106]: $(\delta m_s^2, \sin^2 2\theta_s) = (0.9\ \text{eV}^2, 0.089)$. The $1-2-3\sigma$ contours denote the CMB+LSS allowed regions for ν_s with sub-eV mass as in [84]. In order to facilitate the comparison with the results presented in Sec 9.3, a dashed rectangle denotes the parameter-space described by (9.31b).

In the bottom panel the change in $T_{\rm max}$ as δm_s^2 varies is evident, and provided that $T_{\rm max}$ is higher than the active neutrino decoupling temperature the vacuum mixing in this case is large enough that complete thermalisation always occurs. For the non-resonant case the end result is that isocontours of $\delta N_{\rm eff}$ always lie at constant values of $\delta m_s^2 \sin^4 2\theta_s$, as can be seen in the top panel of Fig. 9.1.

In the inverted hierarchy the resonance conditions are always satisfied. Therefore, we expect full thermalization for a larger region of the mass-mixing parameters than in NH, as confirmed in Fig. 9.1. In this case, thermalisation

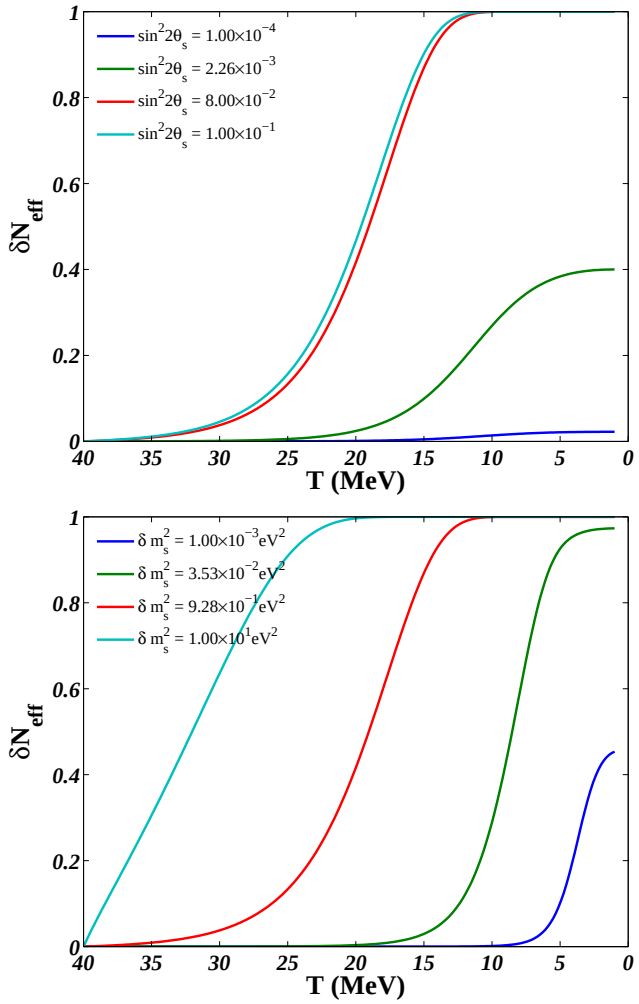


Figure 9.2: Top panel: $\delta N_{\rm eff}$ as a function of the temperature for four different mixing angles ($\sin^2\ 2\theta_s = 10^{-4}, 2\times10^{-3}, 5\times10^{-2}, 10^{-1}$) and fixed mass difference ($\delta m_s^2 = 0.93$ eV^2). Bottom panel: $\delta N_{\rm eff}$ as a function of the temperature for four different mass differences ($\delta m_s^2 = 10^{-3}, 3.5\times10^{-2}, 9.3\times10^{-1}, 10$ eV^2) and fixed mixing angle ($\sin^2 2\theta_s = 0.051$). Thermalisation begins earlier and is more effective for larger mass differences and for larger mixing angles.

may proceed through resonant conversions alone. For illustration, we choose the point of Fig. 9.1 with $(\delta m_s^2, \sin^2\theta_s) = (-3.3\ \text{eV}^2, 6\times10^{-4})$ for which $\delta N_{\rm eff} = 0.55$ and we show the percentage of active (N_a) and sterile (N_s) neutrinos as a function of x for different T in Fig. 9.3. The thermalisation is not complete and it is nearly instantaneous as the resonance moves through the momentum spectrum and the resulting dip in the active sector is quickly repopulated from the background.

We have presented results for $L^{(\mu)} = 0$ only, but the case of $L^{(e)} = 0$ shows exactly the same trend as in Fig. 9.1. However, the region with $\delta N_{\rm eff} = 1$ is slightly smaller than the one shown in Fig. 9.1. This is due to the fact that ν_e's have a larger potential than $\nu_{\mu,\tau}$ (because of the charged current interaction contribution) and therefore resonances occur at slight lower temperatures.

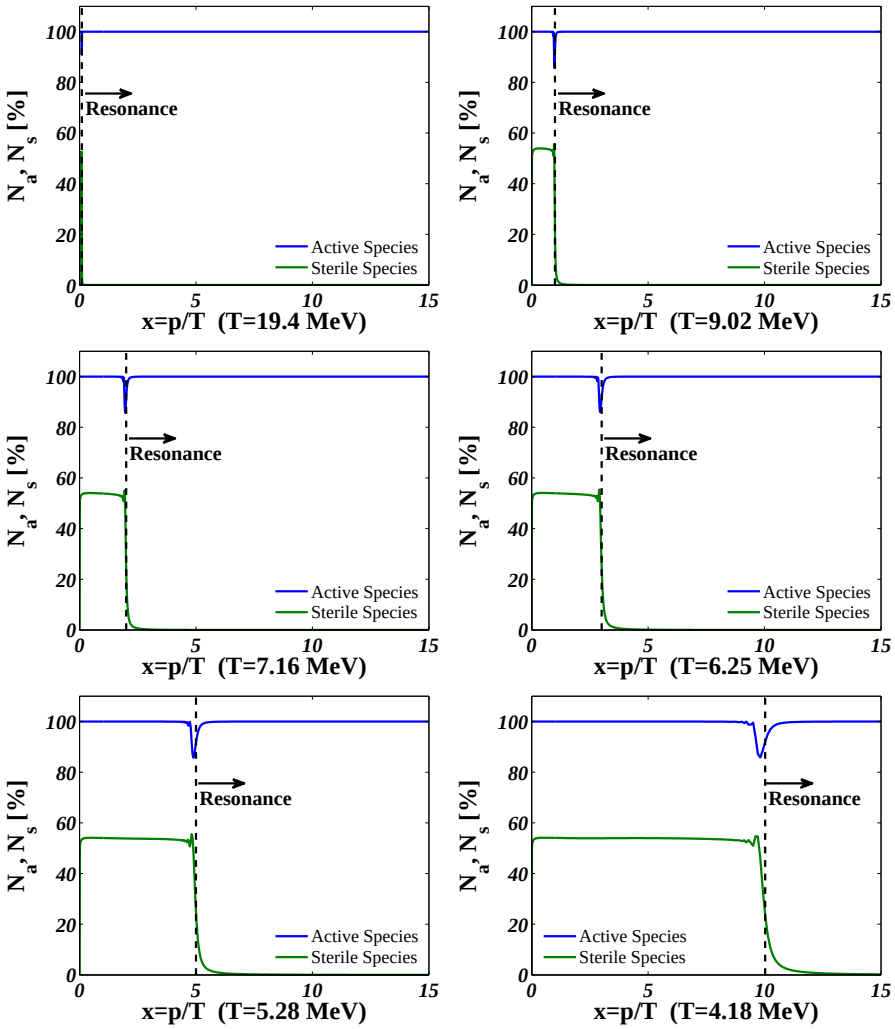


Figure 9.3: Temperature evolution of active and sterile neutrino distributions for the resonant case $(\delta m_s^2, \sin^2\theta_s) = (-3.3\ \mathrm{eV}^2, 6\times10^{-4})$ and $L^{(\mu)} = 0$.

The case of large initial lepton asymmetry

We now discuss the thermalisation degree for initial large lepton asymmetry. In principle, one would expect a lepton asymmetry of the same order of magnitude as the baryon asymmetry ($\eta \simeq 10^{-10}$). However, since neutrinos are neutral particles, $L^{(a)} = 10^{-2} - 10^{-1}$ is not presently excluded [108, 117, 126] by the requirement of charge neutrality. A large lepton asymmetry is responsible for blocking the active-sterile flavor conversions by an in-medium suppression of the mixing angle; therefore it has been invoked as a means of significantly reducing the sterile abundance [96]. A large lepton number can be generated by e.g. an Affleck-Dine mechanism [125] or other models that are able to produce large lepton asymmetries and small baryonic ones [127, 128]. Another interesting possibility is to grow the lepton asymmetry from some initial $L^{(a)} \sim \mathcal{O}(10^{-10})$ using active-sterile oscillations [69, 124, 132]. Solving the QKE's in IH and with an initially small but non-zero lepton number, our preliminary results point toward a final lepton number varying between 10^{-5}

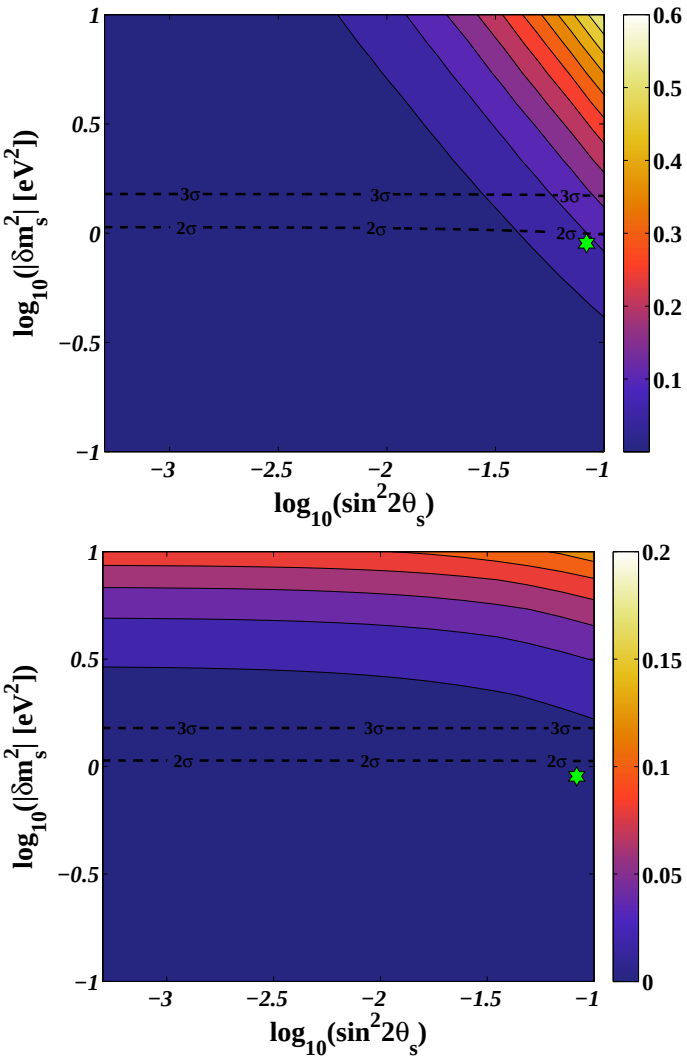


Figure 9.4: Iso-$\delta N_{\rm eff}$ contours in the $\sin^2 2\theta_s - \delta m_s^2$ plane for $L^{(\mu)} = 10^{-2}$ and $\delta m_s^2 > 0$ (top panel) and $\delta m_s^2 < 0$ (bottom panel), as in Fig. 9.1.

and 10^{-2} depending on the mixing parameters. For illustrative purposes, we choose to adopt $L^{(a)} = 10^{-2}$.

Figure 9.4 shows the $\delta N_{\rm eff}$ contour plot for $L^{(\mu)} = 10^{-2}$ and $\delta m_s^2 > 0$ (top panel) and $\delta m_s^2 < 0$ (bottom panel). The region with full thermalisation is now much smaller than in Fig. 9.1.

As we discuss in detail in Appendix A, a large value of $L^{(a)}$ confines the resonances to very small or large values of x, far away from the maximum of the active neutrino momentum distribution (see also [101]). Only at relatively low temperature does the resonance begin to move through the momentum distribution. What happens next is qualitatively very different for normal and inverted hierarchy. For NH the lepton asymmetry decreases as the resonance moves. This causes a run-away effect because as $L^{(a)}$ decreases the resonance moves faster, causing a faster decrease in $L^{(a)}$. When $L^{(a)}$ becomes less than approximately 10^{-5} (see Eq. (9.40)), the resonance disappears and the remaining evolution after this point is equivalent to the $L^{(a)} = 0$ NH case.

For sufficiently large δm_s^2 and $\sin^2 2\theta_s$ the non-resonant production after the resonance disappears can be significant. However, the required mass difference and mixing to obtain the same degree of thermalisation are much larger than in the $L^{(a)} = 0$ case.

The rapid depletion of $L^{(a)}$ in NH causes the numerical solution to continue after this point with a very small time step. No further resonant production will occur after this point, but as we discuss above, some non-resonant thermalisation has yet to happen at this stage. To circumvent this problem, we stop the code when $L^{(a)}$ becomes very close to zero and restart it again with $L^{(a)} = 0$ using the static approximation discussed in Appendix B.

For IH the lepton asymmetry increases when the resonance moves, causing it to move slower and effectively blocking population of the sterile state until very late. For the range of mixing parameters studied here production of sterile neutrinos is effectively blocked until after the active species decouples, leading to a very small $\delta N_{\rm eff}$. In Appendix A we give equations for the position of the resonances for finite $L^{(a)}$ along with useful approximations valid in different limits.

For ν_s mixing parameters as in [106], $\delta N_{\rm eff} \sim 0$ in IH and $\delta N_{\rm eff} = 0.05$ in NH. Constraints from BBN, CMB, and LSS have usually assumed a fully thermalised sterile state, but as also mentioned in [84] a finite lepton asymmetry can effectively block thermalisation and make this assumption invalid. In that case an eV sterile neutrino will not be in conflict with the cosmological neutrino mass bound, but of course the extra energy density preferred by CMB and LSS will then not be associable with the light sterile neutrino.

We finally note that since we have solved the quantum kinetic equations using the 1 sterile + 1 active approximation, only one lepton asymmetry is relevant in our equations (either e or μ). However, in the real 3+1 scenario there will be 3 separate flavour asymmetries and active-active oscillations will lead to some degree of equilibration between these asymmetries. While there will be some quantitative differences between our 1+1 treatment and the full 3+1 scenario we do expect the same qualitative behaviour, i.e. a blocking of thermalisation due to confinement of the active-sterile resonances.

9.4 Conclusions

Recent cosmological data seem to favor an excess of radiation beyond three neutrino families and photons, and light sterile neutrinos are possible candidates. The upcoming measurement of $\delta N_{\rm eff}$ by Planck will confirm or rule out the existence of such extra radiation with high precision [104, 105].

Light sterile neutrinos could thermalise prior to neutrino decoupling, contributing to the relativistic energy density in the early universe. Present data coming from CMB+LSS, and BBN allow the existence of one sub-eV mass sterile family but do not prefer extra fully thermalised sterile neutrinos in the eV-mass range since they violate the hot dark matter limit on the neutrino mass. However, the assumption of full thermalisation is not necessarily justified. In this paper, we have studied the evolution of active and sterile neutrinos in the early universe in order to calculate the effective number of thermalized species after $T \sim 1$ MeV when active neutrinos have decoupled

and slightly before BBN commences. We have studied the amount of thermalisation for initial null and large ($L^{(a)} = 10^{-2}$) lepton asymmetry, for a range of mass-mixing parameters and for both normal and inverted mass hierarchies.

Assuming null initial lepton asymmetry, we find that the assumption of full thermalisation is justified for eV-mass sterile neutrinos with relatively large mixing (as suggested by short-baseline oscillation data). This inevitably leads to tension between CMB+LSS data which prefers very light sterile neutrinos and Solar, reactor and short-baseline data which prefers a mass around 1 eV or higher.

On the other hand, for large initial lepton asymmetries light sterile neutrinos are not (or only partially) thermalised for almost all the scanned parameter space. This provides a loophole for eV sterile neutrinos to be compatible with CMB+LSS constraints. For lepton asymmetries around 10^{-2} almost no thermalisation occurs for the parameters preferred by Solar, reactor and short-baseline data, and the sterile neutrinos would contribute very little to the current dark matter density.

One remaining open question neglected in this work is related to the impact of sterile neutrinos on BBN. The ν_e and $\bar{\nu}_e$ flux distributions are affected by active-sterile conversions and they enter the weak rates regulating the neutron-proton equilibrium (see [129] for a review on the topic). Therefore the ^{4}He abundance is sensitive to the presence of sterile families. In particular $\delta N_{\rm eff} > 0$ and a less populated ν_e spectrum are both responsible for increasing the freeze-out temperature of the ratio n/p and therefore for a larger ^{4}He abundance.

For small $L^{(a)}$ and the mixing parameters discussed here the active-sterile oscillations occur well before BBN commences, while for large $L^{(a)}$ the active-sterile oscillations are no longer decoupled from the active ones and can occur close to the BBN temperature. We refer the reader to [130] for a discussion of BBN constraints on the sterile sector, but also stress that for large values of $L^{(a)}$ any quantitative exclusion limits in mixing parameter space would require solving the full QKEs including all three active species. This is clearly beyond the scope of the present paper, but remains an interesting and important calculation.

Note added: After the initial version of this paper was finalised, a semi-analytic estimate of the BBN effect in the 3+1 scenario using the quantum rate equations has appeared [131].

Acknowledgments

The authors are grateful to Georg G. Raffelt for valuable discussions. This work was partly supported by the Deutsche Forschungsgemeinschaft under the grant EXC-153 and by the European Union FP7 ITN INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442). I.T. thanks the Alexander von Humboldt Foundation for support.

9.5 Appendix

Location of the resonances

Imposing the resonance condition for neutrinos ($V_z = 0$) and for antineutrinos ($\overline{V}_z = 0$), one finds the locations of the resonances [94]. In order to make explicit the x-dependence, we define

$$V_0 = \frac{\widetilde{V_0}}{x} \text{ and } V_1 = \widetilde{V_1} x \ . \tag{9.32}$$

Introducing

$$\ell = \begin{cases} \text{sign}[L^{(a)}] & \text{for particles} \\ -\text{sign}[L^{(a)}] & \text{for anti-particles} \end{cases} \tag{9.33}$$

the resonance conditions ($V_z = 0$ and $\overline{V}_z = 0$) can be written

$$\widetilde{V_1} x^2 + \ell \, |V_L| \, x + \widetilde{V_0} = 0 \ . \tag{9.34}$$

We define $m \equiv \text{sign}[\delta m_s^2]$ and write the solution in the following way

$$x_{\text{res}} = x_0 \left[A\ell \pm \sqrt{A^2 - m} \right] \equiv x_0 F^{\pm}_{\ell m}(A) \ , \tag{9.35}$$

where we have defined

$$x_0 = \sqrt{\frac{m\widetilde{V_0}}{\widetilde{V_1}}} \ , \tag{9.36}$$

$$A = \frac{|V_L|}{2\sqrt{m\widetilde{V_0}\widetilde{V_1}}} \ , \tag{9.37}$$

$$F^{\pm}_{\ell m}(A) = \left[A\ell \pm \sqrt{A^2 - m} \right] \ . \tag{9.38}$$

Note that x_0 is always real and positive. In order to have a physical solution, $F^{\pm}_{\ell m}$ has to be real and positive. This condition is satisfied for $F^{+}_{\pm 1,-1}(A)$ for any A and $F^{\pm}_{+1,+1}(A)$ for $A \geq 1$. Thus, we always have two physical solutions when $m = -1$, one for particles and one for anti-particles. On the other hand, when $m = +1$ and $A \geq 1$, we have two resonances: when $\ell > 0$, they occur for particles and when $\ell < 0$, they occur for anti-particles being in both cases responsible for destroying the lepton number. These equations reproduce the ones reported in [94] when $m = -1$.

We can expand the solutions for small and large $L^{(a)}$:

$$F^{+}_{-1,-1} = -A + \sqrt{1 + A^2} \simeq \begin{cases} 1 - A + \frac{A^2}{2} - \cdots & A \to 0^+ \\ \frac{1}{2A} - \cdots & A \to \infty \end{cases} \tag{9.39a}$$

$$F^{+}_{+1,-1} = \ A + \sqrt{1 + A^2} \simeq \begin{cases} 1 + A + \frac{A^2}{2} - \cdots & A \to 0^+ \\ 2A + \frac{1}{2A} - \cdots & A \to \infty \end{cases} \tag{9.39b}$$

$$F^{+}_{+1,+1} = \ A + \sqrt{A^2 - 1} \simeq \begin{cases} 1 + \sqrt{2}\sqrt{A-1} + \cdots & A \to 1^+ \\ 2A - \frac{1}{2A} - \cdots & A \to \infty \end{cases} \tag{9.39c}$$

$$F^{-}_{+1,+1} = \ A - \sqrt{A^2 - 1} \simeq \begin{cases} 1 - \sqrt{2}\sqrt{A-1} + \cdots & A \to 1^+ \\ \frac{1}{2A} - \cdots & A \to \infty \end{cases} \tag{9.39d}$$

with A and x_0 assuming the following expressions

$$A = \frac{6\zeta(3)}{\pi^3}\sqrt{\frac{10}{7\sqrt{2}}}\frac{T\left|L^{(a)}\right| M_z\sqrt{G_F}}{\sqrt{\cos 2\theta_s\,|\delta m_s^2|\,(n_\nu + n_{\bar\nu})\,g}} \simeq 7.28\times 10^4\; T_{\rm MeV}\frac{\left|L^{(a)}\right|}{\sqrt{\cos 2\theta_s\,|\delta m_s^2|_{\rm eV^2}\,(n_\nu + n_{\bar\nu})\,g}} \tag{9.40}$$

$$x_0 = \frac{3}{\pi}\sqrt{\frac{5}{7\sqrt{2}}}\frac{M_z}{\sqrt{G_F}T^3}\sqrt{\frac{\cos 2\theta_s\,|\delta m_s^2|}{(n_\nu + n_{\bar\nu})\,g}} \simeq 1.81\times 10^4\; T_{\rm MeV}^{-3}\sqrt{\frac{\cos 2\theta_s\,|\delta m_s^2|_{\rm eV^2}}{(n_\nu + n_{\bar\nu})\,g}}\;.$$

For $L^{(a)} = 10^{-2}$ we have $A \gg 1$ and therefore the lowest resonance will be

$$x_{\rm res,low} \simeq \frac{x_0}{2A} = \frac{\pi^2\cos 2\theta_s\,|\delta m_s^2|}{4\sqrt{2}\zeta(3)\,T^4\left|L^{(a)}\right| G_F} \simeq 0.12\frac{\cos 2\theta_s\,|\delta m_s^2|_{\rm eV^2}}{T_{\rm MeV}^4\left|L^{(a)}\right|}\;. \tag{9.41}$$

Note that $x_{\rm res,low}$ is independent on the sign of the mass hierarchy and the total neutrino density. Moreover, from the previous equation, we can extract the temperature at which the lowest resonance starts sweeping the bulk of the Dermi-Dirac distribution. This provides a good estimate of when resonant thermalisation sets in. For example, in the limit of large lepton number, assuming $x_{\rm res,low} \simeq 0.1$, we find $T_{\rm res,low} \simeq 3$ MeV for $(\delta m_s^2, \sin^2 2\theta_s) = (1\ {\rm eV}^2, 10^{-2})$. On the other hand, the higher resonance has no effect at all. In fact

$$x_{\rm res,high} \simeq x_0 \times 2A = \frac{180\zeta(3)\left|L^{(a)}\right| M_z^2}{7\pi^2 T_{\rm MeV}^2\,(n_\nu + n_{\bar\nu})\,g} \simeq 2.6\times 10^{10}\frac{\left|L^{(a)}\right|}{(n_\nu + n_{\bar\nu})\,g T_{\rm MeV}^2}\;. \tag{9.42}$$

Therefore, for $L^{(a)} = 0.01$, $x_{\rm res,high}$ will pass through the peak of the Fermi-Dirac distribution at $T \simeq 1$ GeV. At that temperature, the damping term is so strong that no oscillations occur and thermalisation is inhibited.

9.6 Adiabatic approximation

The so-called "adiabatic" approximation was first introduced in [102] and, under certain conditions, it allows one to derive an approximate analytic solution of the QKE's. In this section, we closely follow the derivation from first principles of [70]. Such derivation assumes that the rate of repopulation ($\dot P_0$) vanishes, however the more careful analysis of [103], including a non-zero repopulation rate, turns out to give the same final formula for P_y. Therefore we choose to adopt the simpler derivation.

Assuming $\dot P_0 = 0$, Eq. (9.4) can be written as a homogeneous matrix equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} -D & -V_z & 0 \\ V_z & -D & -V_x \\ 0 & V_x & 0 \end{bmatrix} P_z\;, \tag{9.43}$$

or using a vectorial notation

$$\frac{\mathrm{d}\boldsymbol{P}}{\mathrm{d}t} = \mathcal{K}\boldsymbol{P}\;. \tag{9.44}$$

The matrix $\mathcal{K}$ can be diagonalised by a time-dependent matrix $\mathcal{U}$, such that $\mathcal{U}\mathcal{K}\mathcal{U}^{-1} = \mathcal{D}$. The matrix $\mathcal{U}$ defines an instantaneous diagonal basis through $\boldsymbol{Q} \equiv \mathcal{U}\boldsymbol{P}$ and, in principle, the evolution equation for $\boldsymbol{Q}$ is non-trivial:

$$\frac{\mathrm{d}\boldsymbol{Q}}{\mathrm{d}t} = \mathcal{K}\boldsymbol{Q} - \mathcal{U}\frac{\mathrm{d}\mathcal{U}^{-1}}{\mathrm{d}t}\boldsymbol{Q} \ . \tag{9.45}$$

However, if we assume that Eq. (9.45) is dominated by the first term, the differential equation can be easily solved. This is the so-called "adiabatic" approximation and its applicability has been analysed thoroughly in [70]. Quoting [70], it is applicable when

$$\frac{V_x}{\sqrt{D^2 + V_z^2}} \ll 1 \ , \tag{9.46a}$$

$$T \ll 3\mathrm{MeV} \ , \tag{9.46b}$$

$$\left|\frac{\mathrm{d}L^{(a)}}{\mathrm{d}T_{\mathrm{MeV}}}\right| \ll 5 \times 10^{-11} T_{\mathrm{MeV}}^4 \ . \tag{9.46c}$$

Equation (9.46a) is not easily stated as just a limit on temperature. If we are not close to the resonance and V_z is dominated by V_0, we find $\tan 2\theta_s \ll 1$ which is true for our parameter space. If we are close to the resonance, the criterion depends on $L^{(a)}$ through x_{res} (see Appendix A for a discussion of the position of the resonances). Using Eqs. (9.35,9.39), we find

$$L^{(a)} \gg 10^{-5} : \frac{|V_x|}{D} \gtrsim \frac{|\delta m_s^2| \sin 2\theta_s}{C_a G_F^2 x_{\mathrm{res,low}}^2 T^6} \sim 5 \times 10^{11} \frac{T_{\mathrm{MeV}}^2}{|\delta m_s^2|} \frac{\sin 2\theta_s}{\cos^2 2\theta_s} L^{(a)2} \ , \tag{9.47}$$

$$L^{(a)} \ll 10^{-5} : \frac{|V_x|}{D} \gtrsim \frac{|\delta m_s^2| \sin 2\theta_s}{C_a G_F^2 x_0^2 T^6} \sim 50 \tan 2\theta_s \ . \tag{9.48}$$

For large $L^{(a)}$, we almost always break the approximation at the lowest resonance. But, since the resonance occurs at a very low momentum, it would have no effect on the physics anyway. In principle, it could still affect numerics but we did not encounter problems on this particular front. For small $L^{(a)}$, we are safe for most of the parameter space and, as for large $L^{(a)}$, if the resonance is not sitting in a populated part of the Fermi-Dirac distribution, there should not be any impact on the physics from breaking this approximation slightly. This also applies to the third condition: If the resonance is not in the middle of a populated part of the distribution, we do not have a fast evolution of $L^{(a)}$ and the approximation is valid.

Equation (9.45) can be formally solved by

$$Q_i(t) = \exp\left(\int_{t_0}^{t} k_i(t')\mathrm{d}t'\right) Q_i(t_0), \tag{9.49}$$

where the k_i's are the eigenvalues of $\mathcal{K}$. Expanding those to lowest order in V_x, we have

$$k_1 = -D + iV_z, \qquad k_2 = -D - iV_z, \qquad k_3 = -\frac{V_x^2 D}{D^2 + V_z^2}. \tag{9.50}$$

Assuming that D is large and V_x satisfies Eq. (9.46a), we find

$$Q_1(t) = Q_2(t) = 0, \quad Q_3(t) = Q_3(t_0). \tag{9.51}$$

The adiabatic approximation allows us to relate P_x, P_y and P_z through

$$\begin{bmatrix} P_x(t) \\ P_y(t) \\ P_z(t) \end{bmatrix} = \mathcal{U}^{-1}(t) \begin{bmatrix} Q_1(t) \\ Q_2(t) \\ Q_3(t) \end{bmatrix} = \mathcal{U}^{-1}(t) \begin{bmatrix} 0 \\ 0 \\ Q_3(t_0) \end{bmatrix} = Q_3(t_0)\boldsymbol{s}_3(t), \tag{9.52}$$

where $\boldsymbol{s}_3(t)$ is the third column in $\mathcal{U}^{-1}(t)$ which is also the normalised eigenvector corresponding to k_3. We have

$$\boldsymbol{s}_3(t) = N \begin{bmatrix} 1 \\ -(D+k_3)/V_z \\ -V_x(D+k_3)/(V_z k_3) \end{bmatrix}, \tag{9.53}$$

with N a normalisation constant. We can now relate P_x and P_y to P_z to lowest order in V_x:

$$P_x(t) = \frac{V_x V_z}{D^2 + V_z^2} P_z(t) \ , \tag{9.54}$$

$$P_y(t) = -\frac{V_x D}{D^2 + V_z^2} P_z(t) \ . \tag{9.55}$$

Substituting V_z by $\overline{V}_z$ gives the corresponding relations for anti-particles.

Part IV

Reflections

Chapter 10

Current and upcoming experiments

10.1 Planck and N_{eff}

The data from the Planck satellite is currently being analysed by the Planck collaboration, and the results should appear in December 2012. The projected sensitivity on N_{eff} should be around 0.2, so within 6 months we will know if Dark Radiation exists or not. If Planck finds $N_{\text{eff}} = 3.0 \pm 0.2$, the sterile neutrino hypothesis as an explanation for the experimental anomalies will be disfavoured. However, if the anomalies persists in future ground based neutrino experiments, we might need to think about ways to suppress sterile neutrino production in the early Universe. As we have demonstrated, this can be accomplished by a large lepton asymmetry. It is also possible that Planck confirms the hints of $N_{\text{eff}} > 3$, and that would of course boost interest in Dark Radiation considerably.

10.2 LHC and Dark Matter

The LHC will be running until the end of 2012 in the current configuration, but so far there are no signs of physics beyond the Standard Model. (I am counting the Higgs boson as part of the Standard Model). Since the dark matter particle is at most weakly interacting, it is hard to detect in the LHC, so two different strategies are used. They look for hints of specific models of new physics (SUSY models, etc), and given these parameters they might conclude something about the dark matter particle. Or, they look for missing energy events.

Chapter 11

Final conclusions

In this chapter I will give an update on the subjects covered in the papers in part III.

11.1 Sommerfeld enhancement

The cosmic ray excesses are still there, but the dark matter explanation does not look very good any more. One reason is the fact that dark matter annihilating with large boost factors are severely constrained by now, both from the energy spectrum as we showed, but even more so from the impact on reionisation which affects the CMB anisotropy spectrum. It turned out to be hard to make dark matter annihilate primarily into leptons, and even in models where this was realised at tree-level, hadronic states would often emerge through electroweak bremsstrahlung. It was also realised, that the spectrum could be explained quite easily with astrophysics, more specifically by nearby pulsars. This is a likely explanation, although it has not been proved. The anisotropy bounds are so strong that they can in fact be used to set bounds on ordinary, thermal WIMP dark matter depending on the dark matter mass, so this will be an interesting observable for Planck. The forecast for Planck [144] suggests that a thermal WIMP can be ruled out for $m_\xi \lesssim 100\text{GeV}$.

11.2 Non-cold relics

Taking advantage of the freedom in selecting the quadrature nodes is quite important for good performance in Boltzmann codes, and similar ideas were subsequently implemented in the widely used Boltzmann code `CAMB`. There might still be room for improvement in `CLASS` and in general. One idea would be to use the Fermi-Dirac distribution as a weight function. Since the integrals over the Laguerre polynomials multiplied by the Fermi-Dirac distribution can be found analytically

$$\int_0^\infty \mathrm{d}x \frac{L_0(x)}{e^x+1} = \log(2), \tag{11.1}$$

$$\int_0^\infty \mathrm{d}x \frac{L_0(x)}{e^x+1} = \log(2) - \frac{\pi^2}{12}, \tag{11.2}$$

$$\vdots = \vdots, \tag{11.3}$$

we could use Wheeler's algorithm [139] to compute the nodes and weights. It is hard to say if it will be a significant improvement over Laguerre quadrature, but it could be worth trying. Regarding the fluid approximation for massive neutrinos, it is not obvious how to improve it.

11.3 Thermalisation of sterile neutrinos

This paper was an important first step for my code `LASAGNA`[1]. I have tried to make it relatively user friendly, and I intent to make it public within a year. Our original goal for `LASAGNA`, was to study the chaotic amplification of the lepton number which occurs in the inverted hierarchy. With the upcoming Planck satellite, the thermalisation project seemed more timely, so I did not have time to complete the chaos project in time for the dissertation.

I am currently also involved in a project to calculate the production of hidden photons using a modified version of `LASAGNA`. In this case we will interchange the effective interaction with the actual collision-terms for Coulomb-scattering, bremsstrahlung and pair-production. The advantage of solving the full QKE-system is that we get the spectral distortions to the CMB black body spectrum directly, and that could potentially constrain the parameter space of such models.

If Planck confirms the presence of Dark Radiation, it would natural to study the thermalisation in even more detail by extending `LASAGNA` to 3 species and/or including a more realistic scattering kernel.

[1] It could be an acronym for Lepton Asymmetric Sterile-Active momentum-Grid Neutrino Analyser...

Bibliography

[5] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, "A Theory of Dark Matter," *Phys. Rev.* **D79** (2009) 015014, arXiv:0810.0713 [hep-ph].

[6] D. J. Fixsen *et al.*, "The Cosmic Microwave Background Spectrum from the Full COBE/FIRAS Data Set," *Astrophys. J.* **473** (1996) 576, arXiv:astro-ph/9605054.

[7] J. B. Dent, S. Dutta, and R. J. Scherrer, "Thermal Relic Abundances of Particles with Velocity- Dependent Interactions," arXiv:0909.4128 [astro-ph.CO].

[8] J. Zavala, M. Vogelsberger, and S. D. M. White, "Relic density and CMB constraints on dark matter annihilation with Sommerfeld enhancement," arXiv:0910.5221 [astro-ph.CO].

[9] T. Bringmann, "Particle Models and the Small-Scale Structure of Dark Matter," *New J. Phys.* **11** (2009) 105027, arXiv:0903.0189 [astro-ph.CO].

[10] P. Gondolo *et al.*, "DarkSUSY: Computing supersymmetric dark matter properties numerically," *JCAP* **0407** (2004) 008, arXiv:astro-ph/0406204.

[11] P. Gondolo, J. Edsjö, P. Ullio, L. Bergström, M. Schelke, E. Baltz, T. Bringmann, and G. Duda, "DarkSUSY homepage." Online.

[12] E. W. Kolb and M. S. Turner, "The Early universe," *Front. Phys.* **69** (1990) 1–547.

[13] **PAMELA** Collaboration, O. Adriani *et al.*, "An anomalous positron abundance in cosmic rays with energies 1.5.100 GeV," *Nature* **458** (2009) 607–609, arXiv:0810.4995 [astro-ph].

[14] **HEAT** Collaboration, S. W. Barwick *et al.*, "Measurements of the cosmic-ray positron fraction from 1- GeV to 50-GeV," *Astrophys. J.* **482** (1997) L191–L194, arXiv:astro-ph/9703192.

[15] J. J. Beatty *et al.*, "New measurement of the cosmic-ray positron fraction from 5-GeV to 15-GeV," *Phys. Rev. Lett.* **93** (2004) 241102, arXiv:astro-ph/0412230.

[16] **AMS-01** Collaboration, M. Aguilar *et al.*, "Cosmic-ray positron fraction measurement from 1-GeV to 30- GeV with AMS-01," *Phys. Lett.* **B646** (2007) 145–154, arXiv:astro-ph/0703154.

[17] G. L. Kane, L.-T. Wang, and T. T. Wang, "Supersymmetry and the cosmic ray positron excess," *Phys. Lett.* **B536** (2002) 263–269, arXiv:hep-ph/0202156.

[18] D. Hooper, J. E. Taylor, and J. Silk, "Can supersymmetry naturally explain the positron excess?," *Phys. Rev.* **D69** (2004) 103509, arXiv:hep-ph/0312076.

[19] I. Cholis, D. P. Finkbeiner, L. Goodenough, and N. Weiner, "The PAMELA Positron Excess from Annihilations into a Light Boson," *JCAP* **0912** (2009) 007, arXiv:0810.5344 [astro-ph].

[20] E. A. Baltz, J. Edsjo, K. Freese, and P. Gondolo, "The cosmic ray positron excess and neutralino dark matter," *Phys. Rev.* **D65** (2002) 063511, arXiv:astro-ph/0109318.

[21] J. Chang *et al.*, "An excess of cosmic ray electrons at energies of 300.800 GeV," *Nature* **456** (2008) 362–365.

[22] D. P. Finkbeiner, "Microwave ISM Emission Observed by WMAP," *Astrophys. J.* **614** (2004) 186–193, arXiv:astro-ph/0311547.

[23] G. Dobler and D. P. Finkbeiner, "Extended Anomalous Foreground Emission in the WMAP 3-Year Data," *Astrophys. J.* **680** (2008) 1222–1234, arXiv:0712.1038 [astro-ph].

[24] D. Hooper, D. P. Finkbeiner, and G. Dobler, "Evidence Of Dark Matter Annihilations In The WMAP Haze," *Phys. Rev.* **D76** (2007) 083012, arXiv:0705.3655 [astro-ph].

[25] A. W. Strong *et al.*, "Gamma-ray continuum emission from the inner Galactic region as observed with INTEGRAL/SPI," *Astron. Astrophys.* **444** (2005) 495, arXiv:astro-ph/0509290.

[26] D. J. Thompson, D. L. Bertsch, and R. H. O'Neal, Jr., "The highest-energy photons seen by the Energetic Gamma Ray Experiment Telescope (EGRET) on the Compton Gamma Ray Observatory," arXiv:astro-ph/0412376.

[27] P. McDonald, R. J. Scherrer, and T. P. Walker, "Cosmic microwave background constraint on residual annihilations of relic particles," *Phys. Rev.* **D63** (2001) 023001, arXiv:astro-ph/0008134.

[28] J. Bernstein and S. Dodelson, "ASPECTS OF THE ZELDOVICH-SUNYAEV MECHANISM," *Phys. Rev.* **D41** (1990) 354.

[29] J. L. Feng, M. Kaplinghat, and H.-B. Yu, "Sommerfeld Enhancements for Thermal Relic Dark Matter," arXiv:1005.4678 [hep-ph].

[30] J. L. Feng, M. Kaplinghat, and H.-B. Yu, "Halo Shape and Relic Density Exclusions of Sommerfeld- Enhanced Dark Matter Explanations of Cosmic Ray Excesses," *Phys. Rev. Lett.* **104** (2010) 151301, arXiv:0911.0422 [hep-ph].

[31] M. Cirelli, M. Kadastik, M. Raidal, and A. Strumia, "Model-independent implications of the e+, e-, anti-proton cosmic ray spectra on properties of Dark Matter," *Nucl. Phys.* **B813** (2009) 1–21, arXiv:0809.2409 [hep-ph].

[32] J. Hisano, S. Matsumoto, and M. M. Nojiri, "Explosive dark matter annihilation," *Phys. Rev. Lett.* **92** (2004) 031303, arXiv:hep-ph/0307216.

[33] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, "Non-perturbative effect on dark matter annihilation and gamma ray signature from galactic center," *Phys. Rev.* **D71** (2005) 063528, arXiv:hep-ph/0412403.

[34] J. March-Russell, S. M. West, D. Cumberbatch, and D. Hooper, "Heavy Dark Matter Through the Higgs Portal," *JHEP* **07** (2008) 058, arXiv:0801.3440 [hep-ph].

[35] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, "CMB Constraints on WIMP Annihilation: Energy Absorption During the Recombination Epoch," *Phys. Rev.* **D80** (2009) 043526, arXiv:0906.1197 [astro-ph.CO].

[36] X.-L. Chen and M. Kamionkowski, "Particle decays during the cosmic dark ages," *Phys. Rev.* **D70** (2004) 043502, arXiv:astro-ph/0310473.

[37] E. Pierpaoli, "Decaying particles and the reionization history of the universe," *Phys. Rev. Lett.* **92** (2004) 031301, arXiv:astro-ph/0310375.

[38] S. R. Furlanetto, S. P. Oh, and E. Pierpaoli, "The Effects of Dark Matter Decay and Annihilation on the High-Redshift 21 cm Background," *Phys. Rev.* **D74** (2006) 103502, arXiv:astro-ph/0608385.

[39] M. Mapelli, A. Ferrara, and E. Pierpaoli, "Impact of dark matter decays and annihilations on reionization," *Mon. Not. Roy. Astron. Soc.* **369** (2006) 1719–1724, arXiv:astro-ph/0603237.

[40] L. Zhang, X. Chen, M. Kamionkowski, Z.-g. Si, and Z. Zheng, "Contraints on radiative dark-matter decay from the cosmic microwave background," *Phys. Rev.* **D76** (2007) 061301, arXiv:0704.2444 [astro-ph].

[41] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, "CMB constraints on Dark Matter models with large annihilation cross-section," *Phys. Rev.* **D80** (2009) 023505, arXiv:0905.0003 [astro-ph.CO].

[42] M. Cirelli, F. Iocco, and P. Panci, "Constraints on Dark Matter annihilations from reionization and heating of the intergalactic gas," *JCAP* **0910** (2009) 009, arXiv:0907.0719 [astro-ph.CO].

[43] S. Cassel, "Sommerfeld factor for arbitrary partial wave processes," *J. Phys.* **G37** (2010) 105009, arXiv:0903.5307 [hep-ph].

[44] T. R. Slatyer, "The Sommerfeld enhancement for dark matter with an excited state," *JCAP* **1002** (2010) 028, arXiv:0910.5713 [hep-ph].

[45] M. Cirelli, A. Strumia, and M. Tamburini, "Cosmology and Astrophysics of Minimal Dark Matter," *Nucl. Phys.* **B787** (2007) 152–175, arXiv:0706.4071 [hep-ph].

[46] **Particle Data Group** Collaboration, K. Nakamura *et al.*, "Review of particle physics," *J.Phys.G* **G37** (2010) 075021.

[47] A. Lewis and A. Challinor, "Evolution of cosmological dark matter perturbations," *Phys.Rev.* **D66** (2002) 023531, arXiv:astro-ph/0203507 [astro-ph].

[48] D. Boyanovsky and J. Wu, "Small scale aspects of warm dark matter : power spectra and acoustic oscillations," *Phys.Rev.* **D83** (2011) 043524, arXiv:1008.0992 [astro-ph.CO].

[49] M. Laine and M. Shaposhnikov, "Sterile neutrino dark matter as a consequence of nuMSM-induced lepton asymmetry," *JCAP* **0806** (2008) 031, arXiv:0804.4543 [hep-ph].

[50] 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, arXiv:0812.3256 [hep-ph].

[51] J. R. Pritchard and E. Pierpaoli, "Neutrino mass from cosmological 21 cm observations," *Nucl.Phys.Proc.Suppl.* **188** (2009) 31–33.

[52] J. Lesgourgues and S. Pastor, "Massive neutrinos and cosmology," *Phys.Rept.* **429** (2006) 307–379, arXiv:astro-ph/0603494 [astro-ph].

[53] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti, *et al.*, "Relic neutrino decoupling including flavor oscillations," *Nucl.Phys.* **B729** (2005) 221–234, arXiv:hep-ph/0506164 [hep-ph].

[54] J. Lesgourgues, "The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview," arXiv:1104.2932 [astro-ph.IM].

[55] D. Blas, J. Lesgourgues, and T. Tram, "The Cosmic Linear Anisotropy Solving System (CLASS) II: Approximation schemes," arXiv:1104.2933 [astro-ph.CO].

[56] J. Lesgourgues, "The Cosmic Linear Anisotropy Solving System (CLASS) III: Comparision with CAMB for LambdaCDM," arXiv:1104.2934 [astro-ph.CO].

[57] M. Doran, "Cmbeasy:: an object oriented code for the cosmic microwave background," *JCAP* **0510** (2005) 011, arXiv:astro-ph/0302138 [astro-ph].

[58] A. Lewis, A. Challinor, and A. Lasenby, "Efficient computation of CMB anisotropies in closed FRW models," *Astrophys.J.* **538** (2000) 473–476, arXiv:astro-ph/9911177 [astro-ph].

[59] U. Seljak and M. Zaldarriaga, "A Line of sight integration approach to cosmic microwave background anisotropies," *Astrophys.J.* **469** (1996) 437–444, arXiv:astro-ph/9603033 [astro-ph].

[60] W. Hu, "Structure formation with generalized dark matter," *Astrophys.J.* **506** (1998) 485–494, arXiv:astro-ph/9801234 [astro-ph].

[61] C.-P. Ma and E. Bertschinger, "Cosmological perturbation theory in the synchronous and conformal Newtonian gauges," *Astrophys.J.* **455** (1995) 7–25, arXiv:astro-ph/9506072 [astro-ph].

[62] M. Shoji and E. Komatsu, "Massive Neutrinos in Cosmology: Analytic Solutions and Fluid Approximation," *Phys.Rev.* **D81** (2010) 123516, arXiv:1003.0942 [astro-ph.CO].

[63] A. Cuoco, J. Lesgourgues, G. Mangano, and S. Pastor, "Do observations prove that cosmological neutrinos are thermally distributed?," *Phys.Rev.* **D71** (2005) 123501, arXiv:astro-ph/0502465 [astro-ph].

[64] K. Enqvist, K. Kainulainen, and M. J. Thomson, "Stringent cosmological bounds on inert neutrino mixing," *Nucl.Phys.* **B373** (1992) 498–528.

[65] L. Stodolsky, "On the Treatment of Neutrino Oscillations in a Thermal Environment," *Phys. Rev.* **D36** (1987) 2273.

[66] B. H. J. McKellar and M. J. Thomson, "Oscillating doublet neutrinos in the early universe," *Phys. Rev.* **D49** (1994) 2710–2728.

[67] G. Sigl and G. G. Raffelt, "General kinetic description of relativistic mixed neutrinos," *Nucl. Phys.* **B406** (1993) 423–451.

[68] D. Boyanovsky and C. Ho, "Production of a sterile species: Quantum kinetics," *Phys.Rev.* **D76** (2007) 085011, arXiv:0705.0703 [hep-ph].

[69] R. Barbieri and A. Dolgov, "Neutrino oscillations in the early universe," *Nucl.Phys.* **B349** (1991) 743–753.

[70] N. F. Bell, R. R. Volkas, and Y. Y. Y. Wong, "Relic neutrino asymmetry evolution from first principles," *Phys. Rev.* **D59** (1999) 113001, arXiv:hep-ph/9809363.

[71] P. Di Bari and R. Foot, "On the sign of the neutrino asymmetry induced by active- sterile neutrino oscillations in the early universe," *Phys. Rev.* **D61** (2000) 105012, arXiv:hep-ph/9912215.

[72] K. Enqvist, K. Kainulainen, and J. Maalampi, "Resonant neutrino transitions and nucleosynthesis," *Phys. Lett.* **B249** (1990) 531–534.

[73] K. Enqvist, K. Kainulainen, and J. Maalampi, "Refraction and oscillations of neutrinos in the early universe," *Nucl. Phys.* **B349** (1991) 754–790.

[74] D. Notzold and G. G. Raffelt, "Neutrino Dispersion at Finite Temperature and Density," *Nucl. Phys.* **B307** (1988) 924.

[75] **LSND** Collaboration, A. Aguilar-Arevalo *et al.*, "Evidence for neutrino oscillations from the observation of anti-nu/e appearance in a anti-nu/mu beam," *Phys. Rev.* **D64** (2001) 112007, arXiv:hep-ex/0104049.

[76] A. Strumia, "Interpreting the LSND anomaly: sterile neutrinos or CPT- violation or...?," *Phys. Lett.* **B539** (2002) 91–101, arXiv:hep-ph/0201134.

[77] M. C. Gonzalez-Garcia and M. Maltoni, "Phenomenology with Massive Neutrinos," *Phys. Rept.* **460** (2008) 1–129, arXiv:0704.1800 [hep-ph].

[78] **MiniBooNE** Collaboration, A. A. Aguilar-Arevalo *et al.*, "Unexplained Excess of Electron-Like Events From a 1-GeV Neutrino Beam," *Phys. Rev. Lett.* **102** (2009) 101802, arXiv:0812.2243 [hep-ex].

[79] **MiniBooNE** Collaboration, A. A. Aguilar-Arevalo *et al.*, "A Search for Electron Antineutrino Appearance at the $\Delta m^2 \sim 1$ eV^2 Scale," *Phys. Rev. Lett.* **103** (2009) 111801, arXiv:0904.1958 [hep-ex].

[80] G. Karagiorgi, Z. Djurcic, J. M. Conrad, M. H. Shaevitz, and M. Sorel, "Viability of $\Delta m^2 \sim 1$ eV^2 sterile neutrino mixing models in light of MiniBooNE electron neutrino and antineutrino data from the Booster and NuMI beamlines," *Phys. Rev.* **D80** (2009) 073001, arXiv:0906.1997 [hep-ph].

[81] S. Razzaque and A. Y. Smirnov, "Searching for sterile neutrinos in ice," *JHEP* **07** (2011) 084, arXiv:1104.1390 [hep-ph].

[82] E. Akhmedov and T. Schwetz, "MiniBooNE and LSND data: non-standard neutrino interactions in a (3+1) scheme versus (3+2) oscillations," *JHEP* **10** (2010) 115, arXiv:1007.4171 [hep-ph].

[83] M. C. Gonzalez-Garcia, M. Maltoni, and J. Salvado, "Robust Cosmological Bounds on Neutrinos and their Combination with Oscillation Results," *JHEP* **08** (2010) 117, arXiv:1006.3795 [hep-ph].

[84] J. Hamann, S. Hannestad, G. G. Raffelt, I. Tamborra, and Y. Y. Y. Wong, "Cosmology Favoring Extra Radiation and Sub-eV Mass Sterile Neutrinos as an Option," *Phys. Rev. Lett.* **105** (2010) 181301, arXiv:1006.5276 [hep-ph].

[85] E. Giusarma *et al.*, "Constraints on massive sterile neutrino species from current and future cosmological data," *Phys. Rev.* **D83** (2011) 115023, arXiv:1102.4774 [astro-ph.CO].

[86] Z. Hou, R. Keisler, L. Knox, M. Millea, and C. Reichardt, "How Additional Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail," arXiv:1104.2333 [astro-ph.CO].

[87] Y. I. Izotov and T. X. Thuan, "The primordial abundance of 4He: evidence for non-standard big bang nucleosynthesis," *Astrophys. J.* **710** (2010) L67–L71, arXiv:1001.4440 [astro-ph.CO].

[88] E. Aver, K. A. Olive, and E. D. Skillman, "A New Approach to Systematic Uncertainties and Self- Consistency in Helium Abundance Determinations," *JCAP* **1005** (2010) 003, arXiv:1001.5218 [astro-ph.CO].

[89] J. Hamann, S. Hannestad, G. G. Raffelt, and Y. Y. Y. Wong, "Sterile neutrinos with eV masses in cosmology – how disfavoured exactly?," *JCAP* **1109** (2011) 034, arXiv:1108.4136 [astro-ph.CO].

[90] G. Mention *et al.*, "The Reactor Antineutrino Anomaly," *Phys. Rev.* **D83** (2011) 073006, arXiv:1101.2755 [hep-ex].

[91] P. Huber, "On the determination of anti-neutrino spectra from nuclear reactors," *Phys. Rev.* **C84** (2011) 024617, arXiv:1106.0687 [hep-ph].

[92] J. Kopp, M. Maltoni, and T. Schwetz, "Are there sterile neutrinos at the eV scale?," *Phys. Rev. Lett.* **107** (2011) 091801, arXiv:1103.4570 [hep-ph].

[93] E. Hairer, S. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations: Stiff and differential-algebraic problems.* Springer series in computational mathematics. Springer-Verlag, 1993.

[94] K. Kainulainen and A. Sorri, "Oscillation induced neutrino asymmetry growth in the early universe," *JHEP* **02** (2002) 020, arXiv:hep-ph/0112158.

[95] K. Abazajian, N. F. Bell, G. M. Fuller, and Y. Y. Y. Wong, "Cosmological lepton asymmetry, primordial nucleosynthesis, and sterile neutrinos," *Phys. Rev.* **D72** (2005) 063004, arXiv:astro-ph/0410175.

[96] Y.-Z. Chu and M. Cirelli, "Sterile neutrinos, lepton asymmetries, primordial elements: How much of each?," *Phys.Rev.* **D74** (2006) 085015, arXiv:astro-ph/0608206 [astro-ph].

[97] A. Melchiorri, O. Mena, S. Palomares-Ruiz, S. Pascoli, A. Slosar, *et al.*, "Sterile Neutrinos in Light of Recent Cosmological and Oscillation Data: A Multi-Flavor Scheme Approach," *JCAP* **0901** (2009) 036, arXiv:0810.5133 [hep-ph].

[98] J. Hamann, S. Hannestad, J. Lesgourgues, C. Rampf, and Y. Y. Y. Wong, "Cosmological parameters from large scale structure - geometric versus shape information," *JCAP* **1007** (2010) 022, arXiv:1003.3999 [astro-ph.CO].

[99] J. Hamann, S. Hannestad, G. G. Raffelt, and Y. Y. Y. Wong, "Observational bounds on the cosmic radiation density," *JCAP* **0708** (2007) 021, arXiv:0705.0440 [astro-ph].

[100] **WMAP** Collaboration, E. Komatsu *et al.*, "Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation," *Astrophys. J. Suppl.* **192** (2011) 18, arXiv:1001.4538 [astro-ph.CO].

[101] X.-D. Shi and G. M. Fuller, "A new dark matter candidate: Non-thermal sterile neutrinos," *Phys. Rev. Lett.* **82** (1999) 2832–2835, arXiv:astro-ph/9810076.

[102] R. Foot and R. Volkas, "Studies of neutrino asymmetries generated by ordinary sterile neutrino oscillations in the early universe and implications for big bang nucleosynthesis bounds," *Phys.Rev.* **D55** (1997) 5147–5176, arXiv:hep-ph/9610229 [hep-ph].

[103] K. S. Lee, R. R. Volkas, and Y. Y. Wong, "Further studies on relic neutrino asymmetry generation. 2. A Rigorous treatment of repopulation in the adiabatic limit," *Phys.Rev.* **D62** (2000) 093025, arXiv:hep-ph/0007186 [hep-ph].

[104] J. Hamann, J. Lesgourgues, and G. Mangano, "Using BBN in cosmological parameter extraction from CMB: A Forecast for PLANCK," *JCAP* **0803** (2008) 004, arXiv:0712.2826 [astro-ph].

[105] L. Perotto, J. Lesgourgues, S. Hannestad, H. Tu, and Y. Y. Wong, "Probing cosmological parameters with the CMB: Forecasts from full Monte Carlo simulations," *JCAP* **0610** (2006) 013, arXiv:astro-ph/0606227 [astro-ph].

[106] C. Giunti and M. Laveder, "Implications of 3+1 Short-Baseline Neutrino Oscillations," *Phys.Lett.* **B706** (2011) 200–207, arXiv:1111.1069 [hep-ph].

[107] S. Dodelson, A. Melchiorri, and A. Slosar, "Is cosmology compatible with sterile neutrinos?," *Phys.Rev.Lett.* **97** (2006) 041301, arXiv:astro-ph/0511500 [astro-ph].

[108] E. Castorina, U. Franca, M. Lattanzi, J. Lesgourgues, G. Mangano, *et al.*, "Cosmological lepton asymmetry with a nonzero mixing angle θ_{13}," arXiv:1204.2510 [astro-ph.CO].

[109] S. Joudaki, "Constraints on Neutrino Mass and Light Degrees of Freedom in Extended Cosmological Parameter Spaces," arXiv:1202.0005 [astro-ph.CO].

[110] E. Giusarma, M. Archidiacono, R. de Putter, A. Melchiorri, and O. Mena, "Sterile neutrino models and nonminimal cosmologies," *Phys.Rev.* **D85** (2012) 083522, arXiv:1112.4661 [astro-ph.CO].

[111] K. M. Nollett and G. P. Holder, "An analysis of constraints on relativistic species from primordial nucleosynthesis and the cosmic microwave background," arXiv:1112.2683 [astro-ph.CO].

[112] A. X. Gonzalez-Morales, R. Poltis, B. D. Sherwin, and L. Verde, "Are priors responsible for cosmology favoring additional neutrino species?," arXiv:1106.5052 [astro-ph.CO].

[113] R. Keisler, C. Reichardt, K. Aird, B. Benson, L. Bleem, *et al.*, "A Measurement of the Damping Tail of the Cosmic Microwave Background Power Spectrum with the South Pole Telescope," *Astrophys.J.* **743** (2011) 28, arXiv:1105.3182 [astro-ph.CO].

[114] J. Dunkley, R. Hlozek, J. Sievers, V. Acquaviva, P. Ade, *et al.*, "The Atacama Cosmology Telescope: Cosmological Parameters from the 2008 Power Spectra," *Astrophys.J.* **739** (2011) 52, arXiv:1009.0866 [astro-ph.CO].

[115] J. Hamann, "Evidence for extra radiation? Profile likelihood versus Bayesian posterior," *JCAP* **1203** (2012) 021, arXiv:1110.4271 [astro-ph.CO].

[116] S. Hannestad and G. G. Raffelt, "Neutrino masses and cosmic radiation density: Combined analysis," *JCAP* **0611** (2006) 016, arXiv:astro-ph/0607101 [astro-ph].

[117] A. Dolgov, S. Hansen, S. Pastor, S. Petcov, G. Raffelt, *et al.*, "Cosmological bounds on neutrino degeneracy improved by flavor oscillations," *Nucl.Phys.* **B632** (2002) 363–382, arXiv:hep-ph/0201287 [hep-ph].

[118] Y. Y. Wong, "Analytical treatment of neutrino asymmetry equilibration from flavor oscillations in the early universe," *Phys.Rev.* **D66** (2002) 025015, arXiv:hep-ph/0203180 [hep-ph].

[119] K. N. Abazajian, J. F. Beacom, and N. F. Bell, "Stringent constraints on cosmological neutrino anti-neutrino asymmetries from synchronized flavor transformation," *Phys.Rev.* **D66** (2002) 013008, arXiv:astro-ph/0203442 [astro-ph].

[120] V. Simha and G. Steigman, "Constraining The Universal Lepton Asymmetry," *JCAP* **0808** (2008) 011, arXiv:0806.0179 [hep-ph].

[121] H.-S. Kang and G. Steigman, "Cosmological constraints on neutrino degeneracy," *Nucl.Phys.* **B372** (1992) 494–520.

[122] L. F. Shampine and M. W. Reichelt, "The MATLAB ODE Suite," *SIAM J. Sci. Comput.* **18** (1997) no. 1, 1–22. http://dx.doi.org/10.1137/S1064827594276424.

[123] T. A. Davis, *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2006.

[124] R. Barbieri and A. Dolgov, "Bounds on Sterile-neutrinos from Nucleosynthesis," *Phys.Lett.* **B237** (1990) 440.

[125] M. Kawasaki, F. Takahashi, and M. Yamaguchi, "Large lepton asymmetry from Q balls," *Phys.Rev.* **D66** (2002) 043516, arXiv:hep-ph/0205101 [hep-ph].

[126] S. Pastor, T. Pinto, and G. G. Raffelt, "Relic density of neutrinos with primordial asymmetries," *Phys.Rev.Lett.* **102** (2009) 241302, arXiv:0808.3137 [astro-ph].

[127] J. A. Harvey and E. W. Kolb, "GRAND UNIFIED THEORIES AND THE LEPTON NUMBER OF THE UNIVERSE," *Phys.Rev.* **D24** (1981) 2090.

[128] A. Dolgov, "Neutrinos in cosmology," *Phys.Rept.* **370** (2002) 333–535, arXiv:hep-ph/0202122 [hep-ph].

[129] G. Steigman, "Primordial Nucleosynthesis in the Precision Cosmology Era," *Ann.Rev.Nucl.Part.Sci.* **57** (2007) 463–491, arXiv:0712.1100 [astro-ph].

[130] A. Dolgov and F. Villante, "BBN bounds on active sterile neutrino mixing," *Nucl.Phys.* **B679** (2004) 261–298, arXiv:hep-ph/0308083 [hep-ph].

[131] A. Mirizzi, N. Saviano, G. Miele, and P. D. Serpico, "Light sterile neutrino production in the early universe with dynamical neutrino asymmetries," arXiv:1206.1046 [hep-ph].

[132] R. Foot, M. J. Thomson, and R. Volkas, "Large neutrino asymmetries from neutrino oscillations," *Phys.Rev.* **D53** (1996) 5349–5353, arXiv:hep-ph/9509327 [hep-ph].

[133] I. Newton, A. Motte, and J. Machin, *The mathematical principles of natural philosophy*. No. vb. 1 in The Mathematical Principles of Natural Philosophy. Printed for B. Motte, 1729. http://books.google.dk/books?id=Tm0FAAAAQAAJ.

[134] I. Newton, *The mathematical principles of natural philosophy*. No. vb. 2 in The Mathematical Principles of Natural Philosophy. printed for Benjamin Motte, 1729. http://books.google.dk/books?id=6EqxPav3vIsC.

[135] A. G. Walker, "On Milne's Theory of World-Structure," *Proceedings of the London Mathematical Society* **s2-42** (1937) no. 1, 90–127, http://plms.oxfordjournals.org/content/s2-42/1/90.full.pdf+html. http://plms.oxfordjournals.org/content/s2-42/1/90.short.

[136] H. P. Robertson, "Kinematics and World-Structure,"*ApJ* **82** (Nov., 1935) 284.

[137] H. P. Robertson, "Kinematics and World-Structure II.,"*ApJ* **83** (Apr., 1936) 187.

[138] H. P. Robertson, "Kinematics and World-Structure III.,"*ApJ* **83** (May, 1936) 257.

[139] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C (2nd ed.): the art of scientific computing*. Cambridge University Press, New York, NY, USA, 1992.

[140] S. P. Martin, "A Supersymmetry primer," arXiv:hep-ph/9709356 [hep-ph].

[141] S. C. Novati, "Microlensing towards the Magellanic Clouds and M31: is the quest for MACHOs still open?," *J.Phys.Conf.Ser.* **354** (2012) 012001, arXiv:1201.5262 [astro-ph.GA].

[142] V. Mukhanov, "Physical foundations of cosmology,".

[143] T. A. Davis, *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2006.

[144] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, "CMB Constraints on WIMP Annihilation: Energy Absorption During the Recombination Epoch," *Phys.Rev.* **D80** (2009) 043526, arXiv:0906.1197 [astro-ph.CO].